Tetsuo Sasao and André B. Fletcher
Introduction to VLBI Systems
Chapter 2
Lecture Notes for KVN Students
Partly Based on Ajou University Lecture Notes
(to be further edited)
Version 1.
Created on December 05, 2004.
Revised on January 14, 2005.
Revised on March 18, 2005.
Revised on April 14, 2005.

Radio Telescope Antennas

Contents

1 Antenna Overview

According to J.D. Kraus, the antenna is 'a region between a guided wave and a free–space wave or vice versa' and 'the antenna interfaces electrons on conductors and photons in space' (see Figure 1). In addition, he says that 'the eye is another such device' (J.D. Kraus, Electromagnetics, 3rd Edition, 1984).

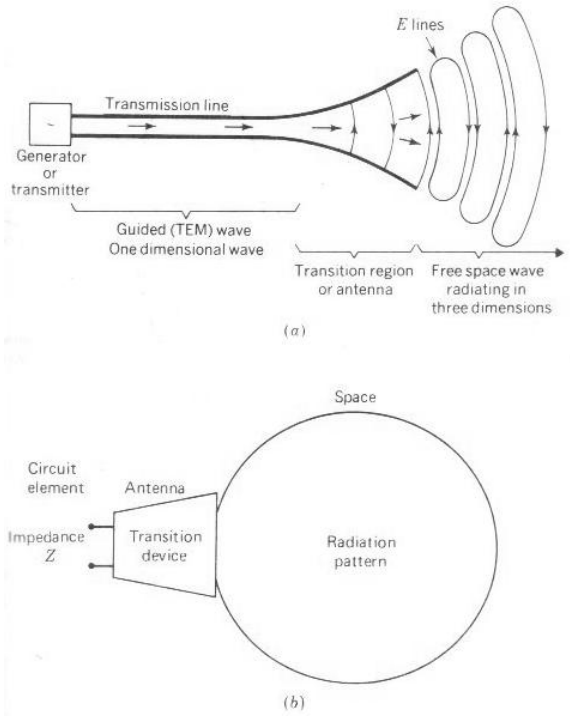

Figure 1: What is an antenna? (From Kraus, 1984).

1.1 Antennas as Radio Telescopes

In the early history of radio astronomy, a variety of antennas, like dipole antennas or horn antennas, were used. However, the overwhelming majority of radio telescopes nowadays adopt the design of the paraboloidal (or, in exceptional cases, spherical) filled aperture antennas.

Existing radio telescope antennas may be classified into subgroups, according to several points of view.

Main reflector design

The rotationally symmetric paraboloidal shape is the most widely used in designs of main reflectors. The 100 m telescope at Effelsberg, Germany, is an example (see Figure 2).

Figure 2: Effelsberg 100 m paraboloidal radio telescope, Germany.

The new Greenbank 100 m telescope (GBT), USA, adopts the offset paraboloidal design in order to achieve maximum efficiency in the reception of radio waves (see Figure 3).

In several huge radio telescopes in the world, partial paraboloidal surfaces are used in a cylindrical paraboloid design. The RATAN–600 telescope in Russia is an example (see Figure 4).

Mounting and tracking design

The most frequently used design for radio telescope mounting is the Alt–Azimuth design, which has a vertical azimuthal axis, and a horizontal altitude axis. The Alt–Azimuth mount (or 'Az–El mount' as frequently referred to by

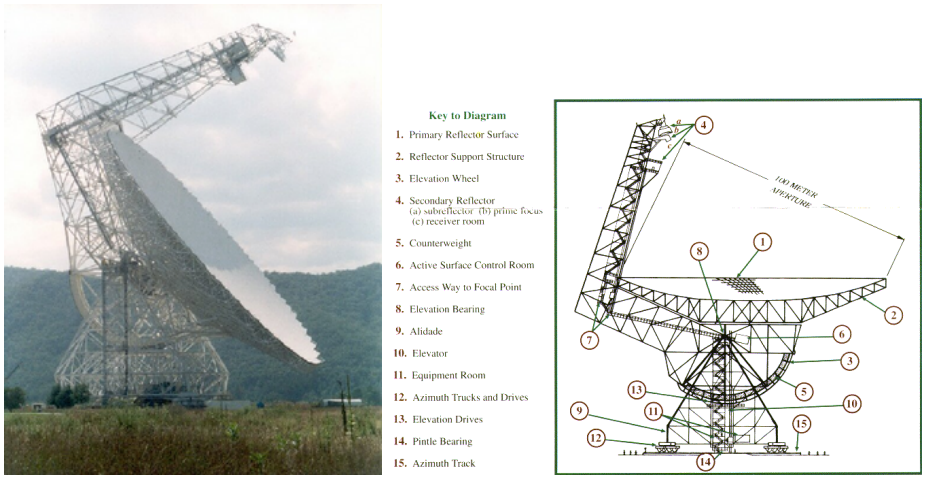

Figure 3: The giant offset paraboloid telescope GBT, USA.

Figure 4: The partial cylindrical paraboloid antenna RATAN–600, Russia.

radio astronomers) is most symmetrical with respect to the direction of gravity, and hence is suited to the stable support of a large and heavy antenna structure. On the other hand, for tracking the diurnal motion of a radio source, real–time coordinate transformation from celestial equatorial (right ascention and declination) coordinates to horizontal (azimuth and elevation) coordinates, as well as simultaneous driving of the antenna around two axes, are always required in this mount system. Moreover, it is in general difficult for an Alt–Azimuth telescope to track sources around the zenith direction. One of the 20 m antennas of the VERA (VLBI Exploration of Radio Astrometry) array at Iriki, Japan, is shown as an example of the Alt–Azimuth radio telescope (left panel in Figure 5).

The most popular mounting design for optical telescopes, the Equatorial mount, has a polar axis parallel to the rotation axis of the Earth, and a declination axis. The Equatorial mount is suited for easy tracking of radio sources in any direction in the sky, and is used for relatively light–weight radio telescopes. The 25 m antennas in the Westerbork Synthesis Radio Telescope (WSRT, the Netherlands) are examples (right panel in Figure 5).

Figure 5: Alt–Azimuth mount antenna at the VERA Iriki station, Japan (left), and Equatorial mount antennas in the WSRT, the Netherlands (right).

A very unique design consisting of a fixed spherical main reflector antenna is used in the giant 305 m radio telescope at Arecibo (Puerto Rico, USA). Instead of driving the main reflector, a cable subsusbspended subreflector with special aberration correction optics is driven to track the radio source (see Figure 6).

Wheel & Track and Yoke & Tower Alt–Azimuth antennas

Figure 7 shows the 20 m VERA antenna, and the 10 m antenna, in NAO Mizusawa, Japan. The entire structure of the 20 m antenna is supported by four wheels, which move on a circular rail fixed on the antenna foundation, in

Figure 6: A spherical fixed main reflector antenna at Arecibo, Puerto Rico, USA, 305 m diameter.

Figure 7: Wheel & Track (left) and Yoke & Tower (right) antennas.

order to change the azimuthal orientation of the antenna. Such an antenna is called a "wheel & track" antenna. In the 10 m antenna, an azimuth gear is attached to the top of the tower–like pedestal, and the antenna structure above the gear is rotated around the azimuth axis by the motor drives. Such an antenna is called "yoke & tower" antenna. It is generally thought that the yoke & tower design is best suited to maintain a fixed intersection point for the Azimuth and Elevation axes – This is an important reference point in geodetic VLBI. However, for mechanical reasons, the wheel & track design is more frequently used for large aperture antennas.

Stiffness

For some large aperture millimeter wave telescopes aimed mainly at spectroscopy and imaging of radio sources, a flexible structure is adopted to realize the so–called "homologous transformation", which is such that the deformed main reflector surface maintains a paraboloid shape, and forms a sharp focus, even when the telescope is tilted under the action of the Earth's gravity. The 45 m Millimeter–Wave Telescope at Nobeyama, Japan, is a successful example of such a design (see left panel in Figure 8).

On the other hand, radio telescope antennas for geodetic and astrometric VLBI observations are usually designed to be very stiff, to eliminate possible errors due to the deformation of the telescope. The 32 m geodetic VLBI antenna at Tsukuba, Japan, is an example of the "stiff" telescope (right panel in Figure 8).

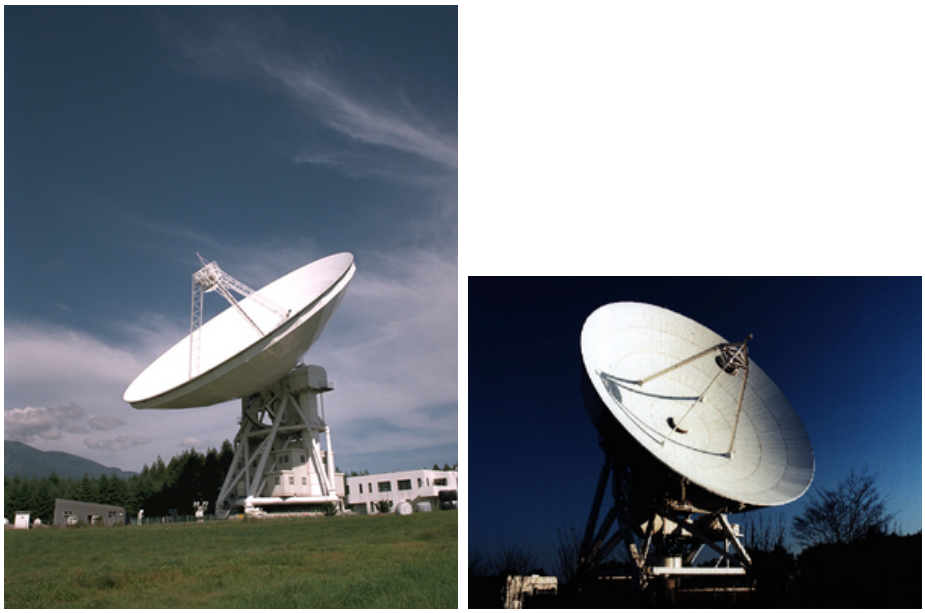

Figure 8: A flexible 45 m Millimeter–Wave Telescope at Nobeyama, Japan (left), and a "stiff" 32 m geodetic VLBI antenna at Tsukuba, Japan (right).

Observing frequency

The most significant factor in antenna design today is the maximum frequency of observation. Antennas are often called as "cm–wave–", "mm–wave–" or "submm-wave" antennas, according to their maximum observing frequency (or shortest observing wavelength). In order to convert wavelength λ to frequency ν, one can use a convenient approximate formula:

$$\nu \text{ (in GHz)} \simeq \frac{30}{\lambda \text{ (in cm)}} . \qquad (1)$$

For cm–wave antennas, the requirements of surface and pointing accuracy are not very severe for present–day antenna manufacturing technology. It is therefore relatively easy to make large antennas in the cm–wave range. At the low–frequency end, the main reflectors of cm–wave antennas can be made of meshed wires. The 25.6 m antenna at Onsala, Sweden, is an example (Figure

Figure 9: A cm–wave antenna at Onsala, Sweden, 25.6 m diameter.

9). For mm–wave antennas, the surface accuracy must be as small as 0.1 mm rms, and the pointing accuracy must be as good as 0.001 degree rms. The large aperture mm–wave antennas like the IRAM 30 m (Spain, EU; left panel in Figure 10), the Nobeyama 45 m (left panel in Figure 8), and the Effelsberg 100 m (Figure 2) are the result of state–of–the–art achievements of modern technology. Some of the mm–wave telescopes, including the TRAO 14 m telescope at Daejeon, Korea (see right panel in Figure 10), are covered by radomes to avoid the effects of strong wind and inhomogeneous solar heating.

The most stringent tolerances in antenna construction are required in the submm–wave range. As a result, all existing submm–wave antennas are relatively small (with diameters around 10 m or smaller), and some of them are located within domes, just like the big optical telescopes. Since the atmosphere is largely opaque in the submm–wave range at low altitude sites near sea level, but fairly transparent at dry high altitude sites, submm–wave telescopes are built on high moutains, with altitudes of 3000–5000 m above sea level (see Figures 11 and 12).

Figure 10: IRAM mm–wave telescope, Spain, 30 m diameter (left), and radome–covered TRAO 14 m mm–wave telescope at Daejeon, Korea (right).

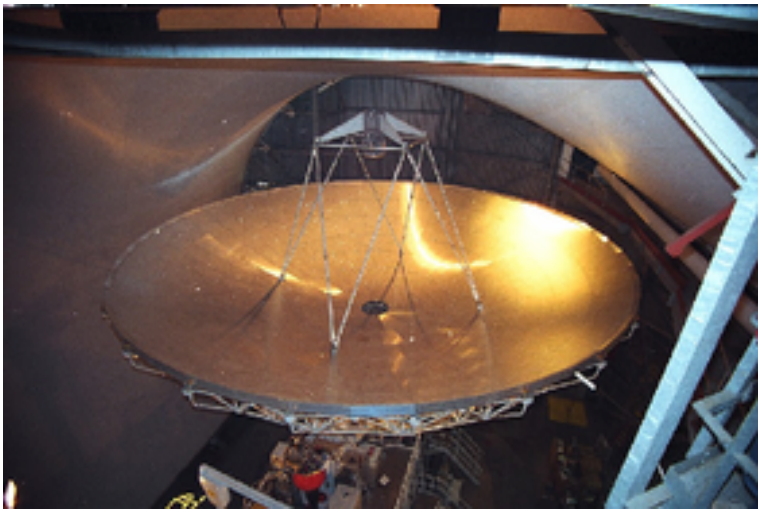

Figure 11: The largest submm–wave antenna JCMT, UK, 15 m diameter, at Mauna Kea, Hawaii.

Figure 12: Main reflector (top) and radome (bottom) of the 1.2 m submm–wave telescope at the top of Mount Fuji, Japan.

1.2 Basic Structure of a Paraboloidal Antenna

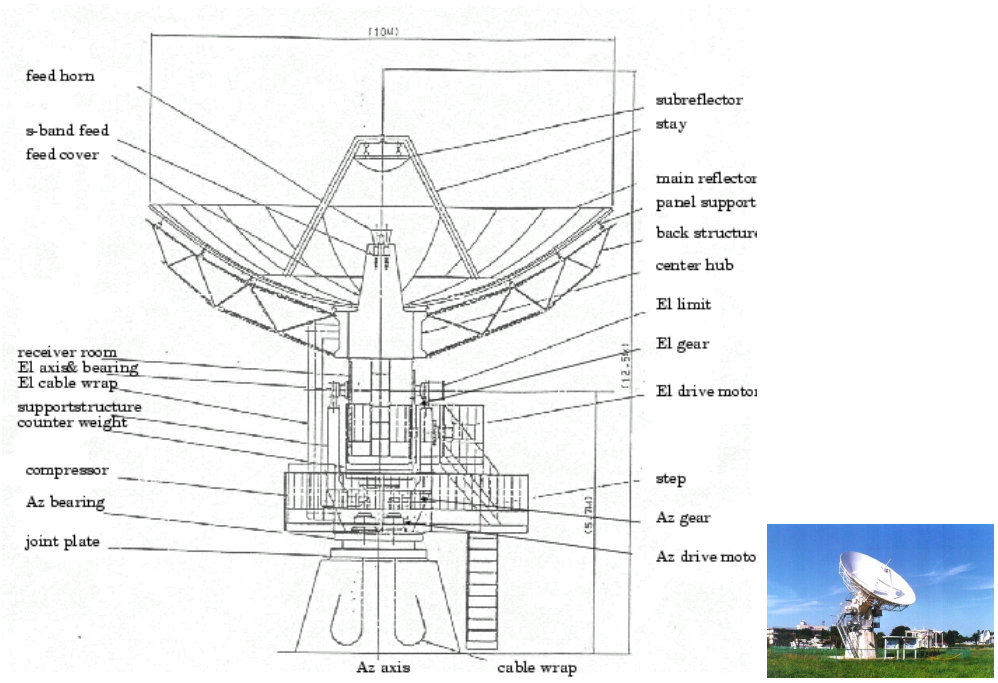

Figure 13: Structure of 10 m VLBI antenna at NAO Mizusawa, Japan.

The basic components of the structure of an Alt–Azimuth paraboloidal antenna are shown in Figure 13. The paraboloidal structure is rotated around the fixed vertical Azimuth axis, and the horizontal Elevation axis, in order to point to any direction in the sky. Usually, the main reflector is composed of a number of panels made of aluminum or carbon fiber, etc, which are fixed to the back structure by adjustable supports. The feed horn is a kind of small antenna, and there is a so–called **waveguide to coaxial–cable (or coaxial to waveguide) converter** at the neck of the feed horn, where the electromagnetic field in space generates the voltage in circuits or vice versa. The waveguide to coaxial–cable converter is the only "real antenna" in strict Kraus' sense, since this is the device that interfaces 'electrons on conductors and photons in space'. The huge structures such as the main reflectors and subreflectors, can be regarded as mere reflecting devices.

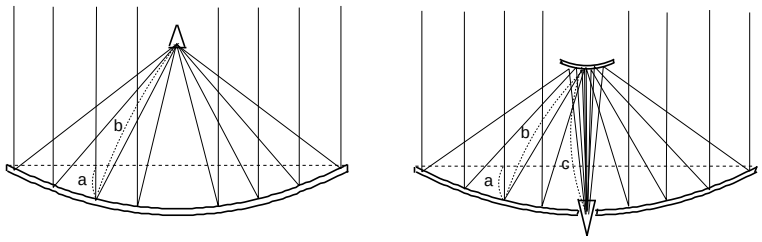

Figure 14: Primary (left) and secondary (right) foci of paraboloidal antennas.

1.3 Why Paraboloid?

Why are paraboloidal antennas the most frequently used for radio telescopes? This is because paraboloids collect plane radio waves coming from astronomical sources towards a focal point where we can place the waveguide to coaxial–cable converter, which transforms the energy of photons in free space to the energy of electrons in conductors in the most efficient way.

In the secondary focus system shown in the right panel of Figure 14, a combination of a paraboloidal main reflector and a hyperboloidal subreflector produces the secondary focus near the main reflector, where the receivers can be conveniently placed. The path length of a ray from the aperture plane — shown by dotted horizontal lines in Figure 14 — to the focus is $a + b$ in the primary focus system, and $a + b + c$ in the secondary focus system; this length is kept constant for any ray. Therefore, the radio waves are collected and summed up with equal phases (i.e. in a "phase–coherent" way) at the foci of the paraboloidal antennas.

There are two possible designs for forming the secondary focus in paraboloidal antennas which use convex (Cassegrain) and concave (Gregorian) hyperboloidal mirrors, respectively, as subreflectors (Figure 15). The Cassegrain design is widely adopted in radio telescopes, because it allows a relatively stable structure against gravitational deformation and wind pressure. On the other hand, the Gregorian design is better suited to the cases where both the primary and secondary foci are used for receiving different frequency bands. In fact, in the Gregorian design, the primary focus feed horn can simply be inserted in front of the subreflector when needed and then removed to allow the secondary focus to be formed.

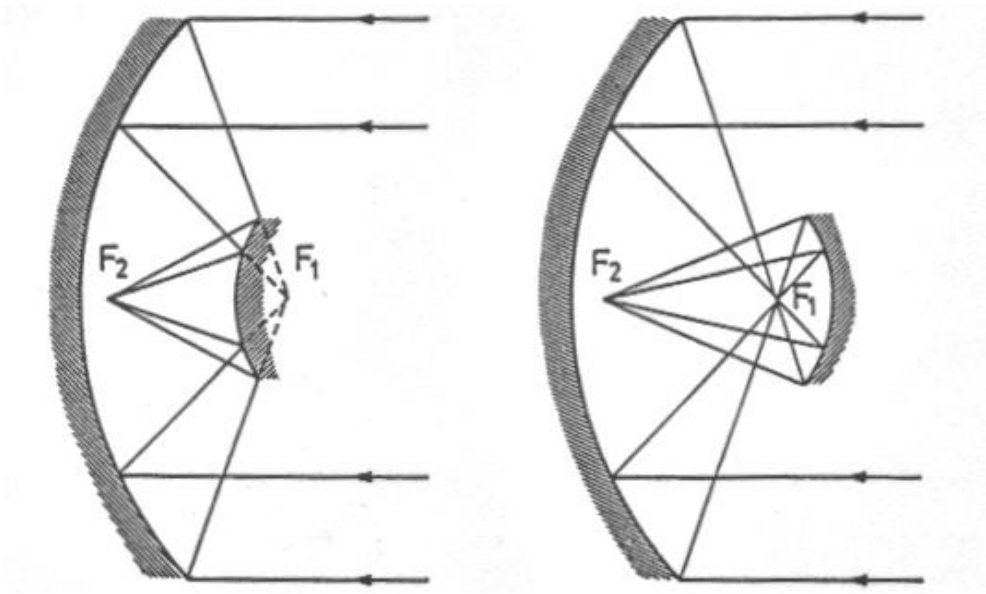

Figure 15: Cassegrain (left) and Gregorian (right) foci of paraboloidal antennas.

2 Antenna Beams

2.1 Some Elements of Vector Algebra

Many textbooks list formulae of vector algebra such as shown in Figure 17. It is not easy to memorize all these complicated formulae. Moreover, they may often contain typographic errors! Without knowledge of these formulae, however, it is difficult to understand electromagnetic theory of radiotelescope antennas.

Fortunately, it is not necessary at all to memorize all these formulae. Instead, we have to memorize just two symbols and one formula, only. They are the Kronecker symbol δ_{ij}, the Levi–Civita symbol ϵ_{ijk}, and a formula $\epsilon_{ijk}\epsilon_{ilm} = \delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl}$. Here, repeated indecies i implies summation, as we will see below.

Let us consider a rectangular (Cartesian) coordinate system, with basis vectors $\boldsymbol{i}_1$, $\boldsymbol{i}_2$, and $\boldsymbol{i}_3$, in a three–dimensional Euclidean space (Figure 16).

In this coordinate system, we consider a radius vector $\boldsymbol{r}$ to a point with coordinates x_1, x_2, and x_3, as well as a scalar field $\Phi(\boldsymbol{r})$ and a vector field $\boldsymbol{A}(\boldsymbol{r})$ with elements A_1, A_2, and A_3, as functions of $\boldsymbol{r}$.

Now let us introduce following notations and conventions.

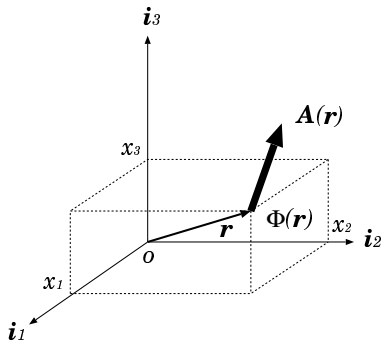

Figure 16: Rectangular coordinate system in a three–dimensional space.

1. **Kronecker's delta symbol** δ_{ij}

   δ_{ij} is 1 when the two indecies take the same value, and 0 when they are different.

$$\delta_{ij} = \begin{cases} 0 & \text{at } i \neq j \\ 1 & \text{at } i = j \end{cases} \qquad (\text{with } \; i, j = 1, 2, 3). \tag{2}$$

   Therefore, $\delta_{11} = \delta_{22} = \delta_{33} = 1$, and $\delta_{12} = \delta_{13} = \cdots = \delta_{32} = 0$.

2. **Levi–Civita's anti–symmetric symbol** ϵ_{ijk}

   ϵ_{123} is defined to be 1, and sign is changed whenever adjacent indecies are substituted.

$$\begin{aligned} &\epsilon_{123} = 1 \\ &\epsilon_{213} = -\epsilon_{123} = -1, \\ &\epsilon_{312} = -\epsilon_{132} = \epsilon_{123} = 1, \\ &\epsilon_{321} = -\epsilon_{231} = \epsilon_{213} = -\epsilon_{123} = -1, \\ &\epsilon_{112} = -\epsilon_{112} = 0, \; \epsilon_{232} = -\epsilon_{232} = 0, \\ &\cdots . \end{aligned} \tag{3}$$

3. **Einstein's summation convention**

   Repeated indecies, *ii* say, imply summation over 11, 22 and

Let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ be arbitrary vector fields assummed to be continous and differentiable everywhere except at a finite number of points, and let ϕ and ψ be arbitrary scalar fields for which the same assumptions are adopted. If $\boldsymbol{A} \cdot \boldsymbol{B}$ is the scalar product and $\boldsymbol{A} \times \boldsymbol{B}$ the vector product then the following algebraic relations are true:

$$\begin{aligned}
\boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C}) &= (\boldsymbol{A} \times \boldsymbol{B}) \cdot \boldsymbol{C} = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}) = (\boldsymbol{B}, \boldsymbol{C}, \boldsymbol{A}) \\
&= (\boldsymbol{C}, \boldsymbol{A}, \boldsymbol{B}) = -(\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{B}) = -(\boldsymbol{C}, \boldsymbol{B}, \boldsymbol{A}) \\
&= -(\boldsymbol{B}, \boldsymbol{A}, \boldsymbol{C}) \,, \qquad \text{(A.1)} \\
\boldsymbol{A} \times (\boldsymbol{B} \times \boldsymbol{C}) &= (\boldsymbol{A} \cdot \boldsymbol{C})\, \boldsymbol{B} - (\boldsymbol{A} \cdot \boldsymbol{B})\, \boldsymbol{C} \,, \qquad \text{(A.2)} \\
\boldsymbol{A} \times (\boldsymbol{B} \times \boldsymbol{C}) + \boldsymbol{B} \times (\boldsymbol{C} \times \boldsymbol{A}) + \boldsymbol{C} \times (\boldsymbol{A} \times \boldsymbol{B}) &= 0 \,, \qquad \text{(A.3)} \\
(\boldsymbol{A} \times \boldsymbol{B}) \cdot (\boldsymbol{C} \times \boldsymbol{B}) &= \boldsymbol{A} \cdot [\boldsymbol{B} \times (\boldsymbol{C} \times \boldsymbol{D})] \\
&= (\boldsymbol{A} \cdot \boldsymbol{C})(\boldsymbol{B} \cdot \boldsymbol{D}) - (\boldsymbol{A} \cdot \boldsymbol{D})(\boldsymbol{B} \cdot \boldsymbol{C}) \,, \qquad \text{(A.4)} \\
(\boldsymbol{A} \times \boldsymbol{B}) \times (\boldsymbol{C} \times \boldsymbol{D}) &= [(\boldsymbol{A} \times \boldsymbol{B}) \cdot \boldsymbol{D}]\, \boldsymbol{C} - [(\boldsymbol{A} \times \boldsymbol{B}) \cdot \boldsymbol{C}]\, \boldsymbol{D} \,. \qquad \text{(A.5)}
\end{aligned}$$

Introducing the gradient of a scalar as $\nabla\phi, \nabla$ considered as a differential operator obeys the following identities

$$\begin{aligned}
\operatorname{grad}(\phi\,\psi) &= \nabla(\phi\,\psi) = \phi\,\nabla\psi + \psi\,\nabla\phi \,, \qquad \text{(A.6)} \\
\operatorname{div}(\phi\,\boldsymbol{A}) &= \nabla \cdot (\phi\,\boldsymbol{A}) = \boldsymbol{A} \cdot \nabla\phi + \phi\,\nabla \cdot \boldsymbol{A} \,, \qquad \text{(A.7)} \\
\operatorname{curl}(\phi\,\boldsymbol{A}) &= \operatorname{rot}(\phi\,\boldsymbol{A}) = \nabla \times (\phi\,\boldsymbol{A}) = \phi\,\nabla \times \boldsymbol{A} - \boldsymbol{A} \times \nabla\phi \,, \qquad \text{(A.8)} \\
\operatorname{div}(\boldsymbol{A} \times \boldsymbol{B}) &= \nabla \cdot (\boldsymbol{A} \times \boldsymbol{B}) = \boldsymbol{B} \cdot (\nabla \times \boldsymbol{A}) - \boldsymbol{A} \cdot (\nabla \times \boldsymbol{B}) \,, \qquad \text{(A.9)} \\
\operatorname{curl}(\boldsymbol{A} \times \boldsymbol{B}) &= \operatorname{rot}(\boldsymbol{A} \times \boldsymbol{B}) = \nabla \times (\boldsymbol{A} \times \boldsymbol{B}) \,, \\
&= \boldsymbol{A}\,(\nabla \cdot \boldsymbol{B}) - \boldsymbol{B}\,(\nabla \cdot \boldsymbol{A}) + (\boldsymbol{B} \cdot \nabla)\,\boldsymbol{A} - (\boldsymbol{A} \cdot \nabla)\,\boldsymbol{B} \,, \qquad \text{(A.10)} \\
\operatorname{grad}(\boldsymbol{A} \cdot \boldsymbol{B}) &= \nabla(\boldsymbol{A} \cdot \boldsymbol{B}) = \boldsymbol{A} \times (\nabla \times \boldsymbol{B}) + \boldsymbol{B} \times (\nabla \times \boldsymbol{A}) \\
&\quad + (\boldsymbol{B} \cdot \nabla)\,\boldsymbol{A} + (\boldsymbol{A} \cdot \nabla)\,\boldsymbol{B} \,. \qquad \text{(A.11)}
\end{aligned}$$

Figure 17: Typical formulae of vector algebra.

33. For example,

$$A_i B_i \equiv \sum_{i=1}^{3} A_i B_i = A_1 B_1 + A_2 B_2 + A_3 B_3,$$
$$\text{and} \tag{4}$$
$$C_{jj} \equiv \sum_{i=1}^{3} C_{ii} = C_{11} + C_{22} + C_{33}.$$

4. **Nabla symbol of spatial derivative ∇**

∇ is a vector like differential operator with three 'elements'

$$\frac{\partial}{\partial x_1}, \quad \frac{\partial}{\partial x_2}, \quad \text{and } \frac{\partial}{\partial x_3}, \tag{5}$$

where $\frac{\partial}{\partial x_i}$ stands for a partial derivative with respect to x_i.

If we apply the above symbols and notations to vector fields $\boldsymbol{A}(\boldsymbol{r})$ and $\boldsymbol{B}(\boldsymbol{r})$, as well as to a scalar field $\Phi(\boldsymbol{r})$, we can express:

- a scalar (or inner) product $\boldsymbol{A} \cdot \boldsymbol{B}$ as

$$\boldsymbol{A} \cdot \boldsymbol{B} = A_i B_i, \tag{6}$$

- i–th element of a vector (or outer) product $\boldsymbol{A} \times \boldsymbol{B}$ as

$$(\boldsymbol{A} \times \boldsymbol{B})_i = \epsilon_{ijk} A_j B_k, \tag{7}$$

  in fact, $\epsilon_{1jk} A_j B_k = A_2 B_3 - A_3 B_2$, $\epsilon_{2jk} A_j B_k = A_3 B_1 - A_1 B_3$, and $\epsilon_{3jk} A_j B_k = A_1 B_2 - A_2 B_1$, in agreement with the definition of the vector product,

- i–th element of a gradient $grad\Phi$ as

$$(grad\Phi)_i = (\nabla\Phi)_i = \frac{\partial}{\partial x_i}\Phi = \frac{\partial \Phi}{\partial x_i}, \tag{8}$$

- a divergence $div\boldsymbol{A}$ as

$$div\boldsymbol{A} = \nabla \cdot \boldsymbol{A} = \frac{\partial}{\partial x_i} A_i = \frac{\partial A_i}{\partial x_i}, \tag{9}$$

  and

- i–th element of a rotation (or curl) $rot\boldsymbol{A}$ as

$$(rot\boldsymbol{A})_i = (\nabla \times \boldsymbol{A})_i = \epsilon_{ijk}\frac{\partial}{\partial x_j}A_k = \epsilon_{ijk}\frac{\partial A_k}{\partial x_j}. \tag{10}$$

A very useful relationship is known between the Levi–Civita's and Kronecker's symbols, which is described by the formula mentioned earlier:

$$\epsilon_{ijk}\epsilon_{ilm} = \delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl}. \tag{11}$$

We can verify this formula by examining all possible combinations of indecies.

In view of Einstein's summation convention, equation (11) is equivalent to

$$\epsilon_{1jk}\epsilon_{1lm} + \epsilon_{2jk}\epsilon_{2lm} + \epsilon_{3jk}\epsilon_{3lm} = \delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl}. \tag{12}$$

Since j, k, l and m may be 1, 2 or 3, equation (11) or (12) has, in general, $3^4 = 81$ components. However, symmetry conditions greately reduce number of components which we really have to individually consider.

First, let us consider what happens if we substitute indecies j and k, or l and m. Because of the antisymmetric property of Levi–Civita symbol, the LHS of equation (11) is antisymmetric (only sign is changed) with respect to such a substitution, and the RHS is also antisymmetric, since if we substitute j and k, for example, the RHS changes its sign as:

$$\delta_{kl}\delta_{jm} - \delta_{km}\delta_{jl} = -(\delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl}).$$

This says that, if we prove equation (11) for jk or lm, then the equation is automatically proven for kj or ml. Moreover, if $j = k$ or $l = m$, the both sides must be equal to 0 (therefore, equal to each other), since for any number A, if $A = -A$, then $A = 0$. Consequently, among 9 components of jk and 9 components of lm, we have to individually consider only 3 and 3 components, which are '12', '13' and '23', for example (see equation (13)).

$$jk = \begin{pmatrix} 11 & \mathbf{12} & \mathbf{13} \\ 21 & 22 & \mathbf{23} \\ 31 & 32 & 33 \end{pmatrix}, \quad \text{and} \quad lm = \begin{pmatrix} 11 & \mathbf{12} & \mathbf{13} \\ 21 & 22 & \mathbf{23} \\ 31 & 32 & 33 \end{pmatrix}. \tag{13}$$

Hence, in total, $3 \times 3 = 9$ components were left to be individually considered.

Furthermore, it can be easily seen that if l or m in equation (11) is not equal to neither j nor k, then both sides must be equal to 0. In fact, if jk is '12', for example, equation (12) becomes

$$\epsilon_{312}\epsilon_{3lm} = \delta_{1l}\delta_{2m} - \delta_{1m}\delta_{2l}.$$

Both sides of this equation are equal to 0, if l or m is 3.

So, we have to consider individually, only three cases when both jk and lm = '12', both jk and lm = '13', and both jk and lm = '23'. They are

$$\begin{aligned}
&\text{for } '12', \quad LHS = \epsilon_{312}\epsilon_{312} = 1, \qquad &&\text{and} \quad RHS = \delta_{11}\delta_{22} = 1,\\
&\text{for } '13', \quad LHS = \epsilon_{213}\epsilon_{213} = 1, \qquad &&\text{and} \quad RHS = \delta_{11}\delta_{33} = 1,\\
&\text{for } '23', \quad LHS = \epsilon_{123}\epsilon_{123} = 1, \qquad &&\text{and} \quad RHS = \delta_{22}\delta_{33} = 1.
\end{aligned}$$

Thus, we proved equation (11).

If we use the above symbols and equation (11), all the complicated formulae of vector algebra are derived in straight forward ways. For example,

$$\begin{aligned}
&[\boldsymbol{A}\times(\boldsymbol{B}\times\boldsymbol{C})]_i = \epsilon_{ijk}A_j\epsilon_{klm}B_lC_m\\
&= \epsilon_{kij}\epsilon_{klm}A_jB_lC_m = (\delta_{il}\delta_{jm}-\delta_{im}\delta_{jl})A_jB_lC_m\\
&= A_jB_iC_j - A_jB_jC_i = B_iA_jC_j - C_iA_jB_j\\
&= [\boldsymbol{B}(\boldsymbol{A}\cdot\boldsymbol{C}) - \boldsymbol{C}(\boldsymbol{A}\cdot\boldsymbol{B})]_i,
\end{aligned}$$

and, therefore,

$$\boldsymbol{A}\times(\boldsymbol{B}\times\boldsymbol{C}) = \boldsymbol{B}(\boldsymbol{A}\cdot\boldsymbol{C}) - \boldsymbol{C}(\boldsymbol{A}\cdot\boldsymbol{B}). \tag{14}$$

We used here a property of Kronecker's symbol: $\delta_{ij}A_j = A_i$.

Also,

$$(\boldsymbol{A}\times\boldsymbol{B})\cdot(\boldsymbol{C}\times\boldsymbol{D}) = \epsilon_{ijk}\epsilon_{ilm}A_jB_kC_lD_m = A_jB_kC_jD_k - A_jB_kC_kD_j,$$

and, hence

$$(\boldsymbol{A}\times\boldsymbol{B})\cdot(\boldsymbol{C}\times\boldsymbol{D}) = (\boldsymbol{A}\cdot\boldsymbol{C})(\boldsymbol{B}\cdot\boldsymbol{D}) - (\boldsymbol{A}\cdot\boldsymbol{D})(\boldsymbol{B}\cdot\boldsymbol{C}). \tag{15}$$

Furthermore,

$$\begin{aligned}
&[\nabla\times(\nabla\times\boldsymbol{A})]_i = \epsilon_{ijk}\tfrac{\partial}{\partial x_j}\epsilon_{klm}\tfrac{\partial}{\partial x_l}A_m\\
&= (\delta_{il}\delta_{jm}-\delta_{im}\delta_{jl})\tfrac{\partial}{\partial x_j}\tfrac{\partial}{\partial x_l}A_m = \tfrac{\partial}{\partial x_j}\tfrac{\partial}{\partial x_i}A_j - \tfrac{\partial}{\partial x_j}\tfrac{\partial}{\partial x_j}A_i\\
&= [\nabla(\nabla\cdot\boldsymbol{A}) - \nabla^2\boldsymbol{A}]_i,
\end{aligned}$$

where

$$\nabla^2 \equiv \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_1^3},$$

and, therefore,

$$\nabla \times (\nabla \times \boldsymbol{A}) = \nabla(\nabla \cdot \boldsymbol{A}) - \nabla^2 \boldsymbol{A}. \tag{16}$$

Other useful examples are:

$$\nabla \cdot (\boldsymbol{A} \times \boldsymbol{B}) = \frac{\partial}{\partial x_i}\epsilon_{ijk}(A_j B_k) = B_k \epsilon_{kij}\frac{\partial}{\partial x_i}A_j - A_j \epsilon_{jik}\frac{\partial}{\partial x_i}B_k, \quad \Rightarrow$$

$$\nabla \cdot (\boldsymbol{A} \times \boldsymbol{B}) = \boldsymbol{B} \cdot (\nabla \times \boldsymbol{A}) - \boldsymbol{A} \cdot (\nabla \times \boldsymbol{B}), \tag{17}$$

$$[\nabla \times (\nabla \Phi)]_i = \epsilon_{ijk}\frac{\partial}{\partial x_j}\frac{\partial}{\partial x_k}\Phi = \epsilon_{ikj}\frac{\partial}{\partial x_k}\frac{\partial}{\partial x_j}\Phi = -\epsilon_{ijk}\frac{\partial}{\partial x_j}\frac{\partial}{\partial x_k}\Phi = 0, \quad \Rightarrow$$

$$rot(grad\Phi) = \nabla \times (\nabla \Phi) = 0, \tag{18}$$

$$\nabla \cdot (\nabla \times \boldsymbol{A}) = \frac{\partial}{\partial x_i}\epsilon_{ijk}\frac{\partial}{\partial x_j}A_k = \epsilon_{ijk}\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}A_k = 0, \quad \Rightarrow$$

$$div(rot\boldsymbol{A}) = \nabla \cdot (\nabla \times \boldsymbol{A}) = 0, \tag{19}$$

and,

$$\nabla \cdot (\Phi \boldsymbol{A}) = \frac{\partial}{\partial x_i}(\Phi A_i) = \frac{\partial \Phi}{\partial x_i}A_i + \Phi\frac{\partial A_i}{\partial x_i} = \nabla\Phi \cdot \boldsymbol{A} + \Phi\,\nabla \cdot \boldsymbol{A}, \quad \Rightarrow$$

$$div(\Phi \boldsymbol{A}) = grad\Phi \cdot \boldsymbol{A} + \Phi\, div\boldsymbol{A}. \tag{20}$$

2.2 Electromagnetic Waves in a Free Space

Prior to considering receptions and transmissions of radio waves (or, more generally speaking, electromagnetic waves) by antennas , we will briefly discuss general behaviors of the electromagnetic waves in a free space, based on the classical theory of electromagnetic fields.

2.2.1 Maxwell Equations

Let us consider following vector and scalar field quantities:

symbol	*name*	*unit in* SI *system*
$\boldsymbol{E}$	electric field intensity	V m^{-1}
$\boldsymbol{D}$	electric flux density	A s m^{-2}
$\boldsymbol{H}$	magnetic field intensity	A m^{-1}
$\boldsymbol{B}$	magnetic flux density	V s m^{-2}
$\boldsymbol{J}$	current density	A m^{-2}
ρ	charge density	A s m^{-3}
σ	specific conductivity	A V^{-1} m^{-1}
ϵ	permittivity	A s V^{-1} m^{-1}
μ	permeability	V s A^{-1} m^{-1}

Note that V A = W (watt) = J s^{-1}, V A^{-1} = Ω (ohm), and V s m^{-2} = T (tesla).

These quantities obey the basic equations of the electromagnetics, the Maxwell equations:

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t}, \tag{21}$$

$$\nabla \times \boldsymbol{H} = \boldsymbol{J} + \frac{\partial \boldsymbol{D}}{\partial t}, \tag{22}$$

$$\nabla \cdot \boldsymbol{D} = \rho, \tag{23}$$

$$\nabla \cdot \boldsymbol{B} = 0, \tag{24}$$

$$\boldsymbol{J} = \sigma \boldsymbol{E}, \tag{25}$$

$$\boldsymbol{D} = \epsilon \boldsymbol{E}, \tag{26}$$

$$\boldsymbol{B} = \mu \boldsymbol{H}. \tag{27}$$

In the vacuum, coefficients in the last three equations, which are sometimes

called as 'equations of state', take following values:

$$\begin{aligned}
&\sigma_0 = 0, \\
&\epsilon_0 = 8.854 \times 10^{-12} \ \text{A s V}^{-1}\ \text{m}^{-1}, \\
&\mu_0 = 1.257 \times 10^{-6} \ \text{V s A}^{-1}\ \text{m}^{-1}, \\
&\text{which satisfy} \\
&\epsilon_0 \mu_0 = c_0^{-2},
\end{aligned} \tag{28}$$

where suffix 0 implies a vacuum value, and $c_0 = 2.998 \times 10^8$ is the light velocity in the vacuum.

2.2.2 Equation of Continuity

From Maxwell equations (22) and (23), and a formula of vector algebra given in equation (19), we obtain

$$\frac{\partial \boldsymbol{\rho}}{\partial t} + \nabla \cdot \boldsymbol{J} = 0. \tag{29}$$

This equation implies that time variation of the electric charge, contained within a finite region, is eqaul to the electric current flowing into (or flowing out of) the region within a unit time. Therefore, this equation is called 'equation of continuity of electric charge' or 'equation of conservation of electric charge'.

2.2.3 Conservation of Energy and Poynting Vector

If we take scalar products of equations (21) and (22) with -$\boldsymbol{H}$ and $\boldsymbol{E}$, respectively, and sum them up, we obtain

$$\begin{aligned}
\boldsymbol{E} \cdot \frac{\partial \boldsymbol{D}}{\partial t} + \boldsymbol{H} \cdot \frac{\partial \boldsymbol{B}}{\partial t} &= -\boldsymbol{E} \cdot \boldsymbol{J} + \boldsymbol{E} \cdot (\nabla \times \boldsymbol{H}) - \boldsymbol{H} \cdot (\nabla \times \boldsymbol{E}) \\
&= -\boldsymbol{E} \cdot \boldsymbol{J} - \nabla \cdot (\boldsymbol{E} \times \boldsymbol{H}),
\end{aligned} \tag{30}$$

where we used a formula of vector algebra given in equation (17).

Considering equations (25), (26), and (27), and assuming that permittivity ϵ and permeability μ are constant in time, we can express equation (30) in a form:

$$\frac{1}{2} \frac{\partial (\boldsymbol{E} \cdot \boldsymbol{D} + \boldsymbol{H} \cdot \boldsymbol{B})}{\partial t} + \nabla \cdot (\boldsymbol{E} \times \boldsymbol{H}) = -\sigma E^2. \tag{31}$$

Taking into account that

1. $u \equiv \frac{1}{2}(\boldsymbol{E} \cdot \boldsymbol{D} + \boldsymbol{H} \cdot \boldsymbol{B})$ is the energy density of the electromagnetic field [J m^{-3}],

2. $\boldsymbol{S} \equiv \boldsymbol{E} \times \boldsymbol{H}$ is the energy flux of the electromagnetic field, called 'Poynting vector' [W m^{-2}],

3. $-\sigma E^2$ is the Joule heat generated per unit time and unit volume due to the Ohmic dissipation [W m^{-3}],

we can interpret equation (31) as an equation of conservation of eletromagnetic energy (Figure 18):

$$\frac{\partial u}{\partial t} + \nabla \cdot \boldsymbol{S} = -\sigma E^2. \tag{32}$$

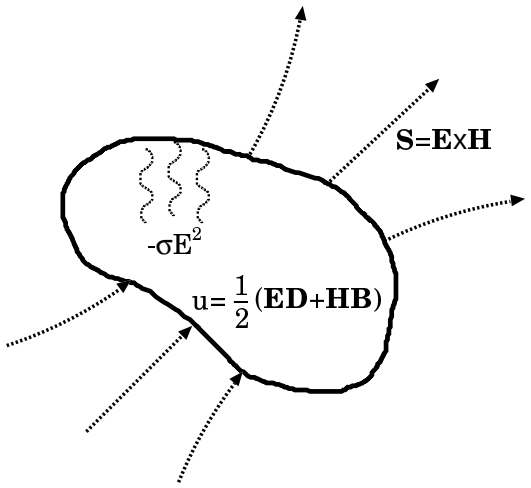

Figure 18: Energy conservation of electromagnetic energy.

Among the quantities characterizing radiation from an astronomical source, which were introduced in the previous chapter "Basic Knowledge of Radio Astronomy', the 'power flux density' (or 'energy/radiation flux density') $\mathcal{S}$:

$$\mathcal{S} = \int_{-\infty}^{\infty} \mathcal{S}_\nu \, d\nu, \tag{33}$$

where $\mathcal{S}_\nu$ is the spectral flux density, has the same unit [W m^{-2}] as the Poynting vector $\boldsymbol{S}$. According to a standard definition adopted by the IEEE

(1977), the power flux density is equal to the time average of the Poynting vector of the electromagnetic wave. This is an important definition which interrelates the radio astronomy with the electromagnetics, as we will see later.

2.2.4 Wave Equations in a Stationary Homogeneous Medium

Let us consider a medium, where the specific conductivity σ, permittivity ϵ and permeability μ are real and constant both in time and space, and the charge density is zero ($\rho = 0$, i.e., the medium is electrically neutral).

In this case, the first four of the Maxwell equations (21) – (24) take simpler forms:

$$\begin{aligned} \nabla \times \boldsymbol{E} &= -\mu \frac{\partial \boldsymbol{H}}{\partial t}, \\ \nabla \times \boldsymbol{H} &= \sigma \boldsymbol{E} + \epsilon \frac{\partial \boldsymbol{E}}{\partial t}, \\ \nabla \cdot \boldsymbol{E} &= 0, \\ \nabla \cdot \boldsymbol{H} &= 0. \end{aligned}$$

Taking time derivatives of the first two equations, and using relations:

$$\begin{aligned} \nabla \times (\nabla \times \boldsymbol{E}) &= \nabla(\nabla \cdot \boldsymbol{E}) - \nabla^2 \boldsymbol{E} = -\nabla^2 \boldsymbol{E}, \\ \nabla \times (\nabla \times \boldsymbol{H}) &= \nabla(\nabla \cdot \boldsymbol{H}) - \nabla^2 \boldsymbol{H} = -\nabla^2 \boldsymbol{H}, \end{aligned}$$

which are derived from a formula of vector algebra in equation (16), and the last two of the above equations, we obtain

$$\begin{aligned} \epsilon\mu \frac{\partial^2 \boldsymbol{E}}{\partial t^2} + \sigma\mu \frac{\partial \boldsymbol{E}}{\partial t} - \nabla^2 \boldsymbol{E} = 0, \\ \epsilon\mu \frac{\partial^2 \boldsymbol{H}}{\partial t^2} + \sigma\mu \frac{\partial \boldsymbol{H}}{\partial t} - \nabla^2 \boldsymbol{H} = 0, \end{aligned}$$

which are equations of the electromagnetic wave in a dissipative medium. If we further assume a dissipationless medium, where $\sigma = 0$, and, therefore, $\boldsymbol{J} = 0$ (no current), the wave equations are described in a familiar form:

$$\begin{aligned} \frac{\partial^2 \boldsymbol{E}}{\partial t^2} - c^2 \nabla^2 \boldsymbol{E} = 0, \\ \frac{\partial^2 \boldsymbol{H}}{\partial t^2} - c^2 \nabla^2 \boldsymbol{H} = 0, \end{aligned} \tag{34}$$

where

$$c \equiv \frac{1}{\sqrt{\epsilon\mu}}, \tag{35}$$

is the velocity of the electromagnetic wave (or light), which is equal to $c_0 = 2.998 \times 10^8$ m s^{-1} in the vacuum.

2.2.5 Monochromatic Plane Waves

The simplest way to describe a solution of the wave equations (34) in the stationary, homogeneous, neutral ($\rho = 0$), and dissipationless ($\sigma = 0$) medium is to represent it as a superposition of monochromatic plane waves:

$$\begin{aligned} \boldsymbol{E} &= \boldsymbol{E}_{0r}\cos(\boldsymbol{k}\cdot\boldsymbol{r} - \omega t) - \boldsymbol{E}_{0i}\sin(\boldsymbol{k}\cdot\boldsymbol{r} - \omega t), \\ \boldsymbol{H} &= \boldsymbol{H}_{0r}\cos(\boldsymbol{k}\cdot\boldsymbol{r} - \omega t) - \boldsymbol{H}_{0i}\sin(\boldsymbol{k}\cdot\boldsymbol{r} - \omega t), \end{aligned} \tag{36}$$

where $\omega \equiv 2\pi\nu$ is an angular frequency of a particular monochromatic plane wave component (for simplicity, we will call the component as a 'plane wave'), $\boldsymbol{r}$ is a radius vector with components x_1, x_2, and x_3 at a certain point in the medium, $\boldsymbol{k}$ is a wave number vector of the plane wave, which satisfies

$$\omega^2 = c^2 k^2, \quad \text{with} \;\; k = \mid \boldsymbol{k} \mid, \tag{37}$$

while $\boldsymbol{E}_{0r}$, $\boldsymbol{E}_{0i}$, $\boldsymbol{H}_{0r}$, and $\boldsymbol{H}_{0i}$ are constant vector coefficients. For further convenience, we introduce complex vector coefficients:

$$\boldsymbol{E}_0 = \boldsymbol{E}_{0r} + i\boldsymbol{E}_{0i}, \quad \text{and} \;\; \boldsymbol{H}_0 = \boldsymbol{H}_{0r} + i\boldsymbol{H}_{0i}, \tag{38}$$

where i is the imaginary unit, and express equation (36) for a plane wave in an equivalent form:

$$\begin{aligned} \boldsymbol{E} &= \Re\,[\boldsymbol{E}_0 e^{i(\boldsymbol{k}\cdot\boldsymbol{r} - \omega t)}], \\ \boldsymbol{H} &= \Re\,[\boldsymbol{H}_0 e^{i(\boldsymbol{k}\cdot\boldsymbol{r} - \omega t)}], \end{aligned} \tag{39}$$

where $\Re$ implies a real part, using the well known Euler formulae:

$$e^{i\theta} = \cos\theta + i\sin\theta, \quad \text{and} \quad e^{-i\theta} = \cos\theta - i\sin\theta, \quad \text{for any } \theta.$$

Furthermore, we will omit this real part symbol $\Re$ in most of following discussions, since, as long as we perform linear operations, including differentiation and integration, to quantities described in the form as equation (39), there is no difference, if we take the real part before or after the linear operations. Therefore, we can perform all the calculations for complex quantities, and, only after we obtain final solutions, recall that actual physical quantities correspond to real parts of the complex ones. This will greatly simplify our mathematical manipulations. Thus, we will use a complex version of equation (39):

$$\begin{aligned} \boldsymbol{E} &= \boldsymbol{E}_0 e^{i(\boldsymbol{k}\cdot\boldsymbol{r} - \omega t)}, \\ \boldsymbol{H} &= \boldsymbol{H}_0 e^{i(\boldsymbol{k}\cdot\boldsymbol{r} - \omega t)}, \end{aligned} \tag{40}$$

where $\boldsymbol{E}$ and $\boldsymbol{H}$ are now regarded as complex quantities, remembering that actual electric and magnetic field intensities are $\Re\,\boldsymbol{E}$ and $\Re\,\boldsymbol{H}$, respectively, which we hereafter denote as

$$\boldsymbol{E}_r = \Re\boldsymbol{E}, \quad \text{and} \quad \boldsymbol{H}_r = \Re\boldsymbol{H}. \tag{41}$$

Now we can easily verify that the plane waves as described in equation (40) really satisfy wave equations (34). In fact, if we take the electric field $\boldsymbol{E}$ in equation (40), for example, we have

$$\frac{\partial^2 \boldsymbol{E}}{\partial t^2} = \boldsymbol{E}_0 \frac{\partial^2}{\partial t^2} e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)} = -\omega^2 \boldsymbol{E}_0 e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)} = -\omega^2 \boldsymbol{E}, \tag{42}$$

and also

$$\nabla^2 \boldsymbol{E} = \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_i}\boldsymbol{E} = \boldsymbol{E}_0 \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_i} e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)},$$

where

$$\begin{aligned}\frac{\partial}{\partial x_i} e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)} &= i\frac{\partial}{\partial x_i}(\boldsymbol{k}\cdot\boldsymbol{r}) e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)} = i\frac{\partial}{\partial x_i}(k_j x_j) e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)} \\ &= ik_j \frac{\partial x_j}{\partial x_i} e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)} = ik_j \delta_{ij} e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)} = ik_i e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)},\end{aligned}$$

and, therefore,

$$\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_i} e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)} = -k^2 e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)},$$

which yields

$$\nabla^2 \boldsymbol{E} = -k^2 \boldsymbol{E}_0 e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)} = -k^2 \boldsymbol{E}. \tag{43}$$

From equations (42) and (43), it is clear that the plane wave $\boldsymbol{E}$ satisfies the wave equation:

$$\frac{\partial^2 \boldsymbol{E}}{\partial t^2} - c^2 \nabla^2 \boldsymbol{E} = 0,$$

in view of the relation $\omega^2 = c^2 k^2$, given in equation (37), between the angular frequency ω and the wave number k. The same discussion holds for the magnetic field $\boldsymbol{H}$, too.

The reason, why the wave in equation (40) is called 'plane wave', is evident from Figure 19. In fact, a surface of the constant phase ($\boldsymbol{k}\cdot\boldsymbol{r}-\omega t = \text{const}$), or a 'wave front' forms a flat plane, which is perpendicular to the wave number vector $\boldsymbol{k}$, and moves along $\boldsymbol{k}$ with the speed $\omega/k = c$ as time t increases.

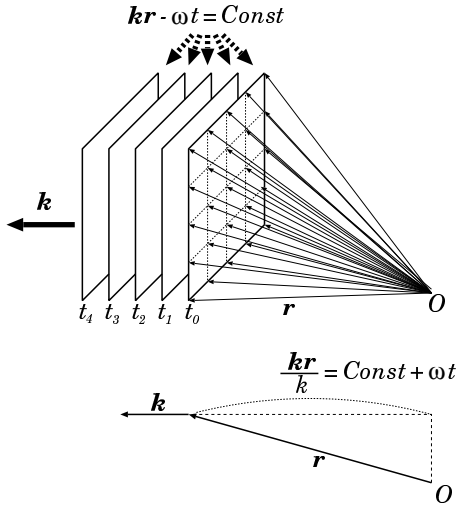

Figure 19: Wave front of a plane wave forms a flat plane moving towards $\boldsymbol{k}$ with the speed $\omega/k = c$.

2.2.6 Electric and Magnetic Fields in a Plane Wave

Although any plane wave, with arbitrary constant $\boldsymbol{E}_0$ or $\boldsymbol{H}_0$ in equation (40), satisfies the wave equation (34), as we have just seen above, actual electric and magnetic fields in an electromagnetic wave must be mutually related to each other, and must exhibit certain characteristic properties.

These additional properties, which will be shown below, come from the fact that the electric $\boldsymbol{E}$ and magnetic $\boldsymbol{H}$ fields must satisfy, not only the wave equations (40), where they are completely separated, but also the original Maxwell equations, where they are related to each other.

1. **The waves are transversal.**
   In our homogeneous medium case, Maxwell equations (23) and (24) are reduced to

$$\nabla \cdot \boldsymbol{E} = 0,$$
$$\nabla \cdot \boldsymbol{H} = 0.$$

   Inserting the plane wave form of the electric field $\boldsymbol{E}$, given in equation (40), into the upper one of the above equations, we have

$$\nabla \cdot \boldsymbol{E} = E_{0i} \frac{\partial}{\partial x_i} e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)} = ik_i E_{0i} e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)} = i\boldsymbol{k}\cdot\boldsymbol{E}_0 e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)} = 0.$$

   Since this relation holds for arbitrary $e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)}$, we have $\boldsymbol{k}\cdot\boldsymbol{E}_0 = 0$ and, therefore, $\boldsymbol{k}\cdot\boldsymbol{E} = 0$. The same argument holds for the magnetic field

as well in the lower equation. Thus, we obtain

$$\begin{aligned} \boldsymbol{k} \cdot \boldsymbol{E} &= 0, \\ \boldsymbol{k} \cdot \boldsymbol{H} &= 0. \end{aligned} \tag{44}$$

Taking real parts of these complex equations, we easily see that the actual physical fields $\boldsymbol{E}_r$ and $\boldsymbol{H_r}$, which are the real parts $\boldsymbol{E}_r = \Re\boldsymbol{E}$ and $\boldsymbol{H}_r = \Re\boldsymbol{H}$ as defined in equation (41), also satisfy

$$\begin{aligned} \boldsymbol{k} \cdot \boldsymbol{E}_r &= 0, \\ \boldsymbol{k} \cdot \boldsymbol{H}_r &= 0, \end{aligned} \tag{45}$$

which show that the vector fields are perpendicular to the direction $\boldsymbol{n}$ of the wave propagation:

$$\boldsymbol{n} \equiv \frac{\boldsymbol{k}}{k}. \tag{46}$$

This means that the plane electromagnetic waves are transversal.

2. **Orthogonality of electric and magnetic fields.**
In our stationary, homogeneous, dissipationless ($\sigma = 0,\ \boldsymbol{J} = 0$), and neutral ($\rho = 0$) medium, Maxwell equations (21) and (22) are reduced to

$$\begin{aligned} \nabla \times \boldsymbol{E} &= -\mu \frac{\partial \boldsymbol{H}}{\partial t}, \\ \nabla \times \boldsymbol{H} &= \epsilon \frac{\partial \boldsymbol{E}}{\partial t}. \end{aligned}$$

Let us insert the plane wave forms of the electric ($\boldsymbol{E}$) and magnetic ($\boldsymbol{H}$) field intensities, given in equation (40), into the upper one of the above equations.

Since

$$(\nabla \times \boldsymbol{E})_i = \epsilon_{ijk} E_{0k} \frac{\partial}{\partial x_j} e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)} = i\,\epsilon_{ijk} k_j E_{0k} e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)},$$

we obtain

$$\nabla \times \boldsymbol{E} = i\,\boldsymbol{k} \times \boldsymbol{E}.$$

Also,

$$\frac{\partial \boldsymbol{H}}{\partial t} = \boldsymbol{H}_0 \frac{\partial}{\partial t} e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)} = -i\,\omega \boldsymbol{H}_0 e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)} = -i\,\omega \boldsymbol{H}.$$

Similar relations:

$$\nabla \times \boldsymbol{H} = i\,\boldsymbol{k} \times \boldsymbol{H}, \ \text{ and } \ \frac{\partial \boldsymbol{E}}{\partial t} = -i\,\omega \boldsymbol{E},$$

hold for the lower equation. Therefore, we have

$$\begin{aligned} \boldsymbol{k} \times \boldsymbol{E} &= \mu\, \omega \boldsymbol{H}, \\ \boldsymbol{k} \times \boldsymbol{H} &= -\epsilon\, \omega \boldsymbol{E}. \end{aligned} \tag{47}$$

Taking real parts of these complex equations, we obtain the same relations:

$$\begin{aligned} \boldsymbol{k} \times \boldsymbol{E}_r &= \mu\, \omega \boldsymbol{H}_r, \\ \boldsymbol{k} \times \boldsymbol{H}_r &= -\epsilon\, \omega \boldsymbol{E}_r, \end{aligned} \tag{48}$$

for the actual physical field quantities $\boldsymbol{E}_r$ and $\boldsymbol{H}_r$ (equation (41)), as well.

Therefore, we have

$$\boldsymbol{E}_r \cdot \boldsymbol{H}_r = 0, \tag{49}$$

which means that the electric and magnetic fields are perpendicular (orthogonal) to each other (Figure 20).

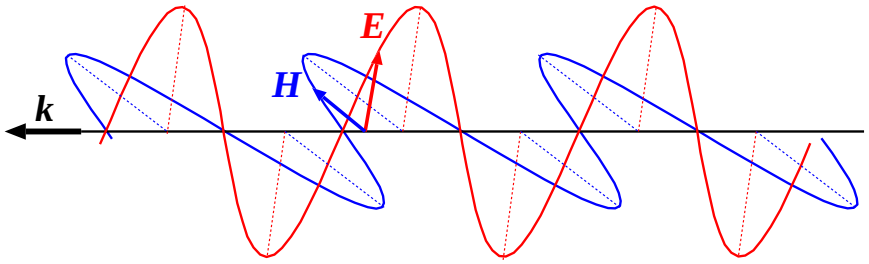

Figure 20: Electric and magnetic fields in a monochromatic plane wave are perpendicular to a direction of propagation, and to each other.

3. **Amplitudes of electric and magnetic fields. Intrinsic impedance.** From equations (45) and (48), and also from a formula of vector algebra given in equation (15), we have

$$\mid \boldsymbol{E}_r \mid^2 = \boldsymbol{E}_r \cdot \boldsymbol{E}_r = \frac{1}{\omega^2 \epsilon^2} (\boldsymbol{k} \times \boldsymbol{H}_r) \cdot (\boldsymbol{k} \times \boldsymbol{H}_r) = \frac{1}{\omega^2 \epsilon^2} k^2 \mid \boldsymbol{H}_r \mid^2,$$

and

$$\mid \boldsymbol{H}_r \mid^2 = \frac{1}{\omega^2 \mu^2} k^2 \mid \boldsymbol{E}_r \mid^2 .$$

Since we have

$$\frac{k^2}{\omega^2} = \frac{1}{c^2} = \epsilon\, \mu,$$

from equaions (37) and (35), the above two equations are equivalent, and express the same relation, describing the amplitude ratio of the electric and magnetic fields in a monochromatic plane wave:

$$\mid \boldsymbol{E}_r \mid = \sqrt{\frac{\mu}{\epsilon}} \mid \boldsymbol{H}_r \mid, \quad \text{or} \quad \mid \boldsymbol{H}_r \mid = \sqrt{\frac{\epsilon}{\mu}} \mid \boldsymbol{E}_r \mid . \tag{50}$$

The coefficient:

$$Z \equiv \sqrt{\frac{\mu}{\epsilon}}, \tag{51}$$

which is equal to

$$Z = \frac{\mid \boldsymbol{E}_r \mid}{\mid \boldsymbol{H}_r \mid},$$

is called 'intrinsic impedance' of the medium, and is measured by a unit: [V A^{-1} = Ω (Ohm)], which is the same with the unit of the resistance in an electrical circuit. The value of the intrinsic impedance in the vacuum is equal to

$$Z_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} = 376.3 \ \ \Omega.$$

4. **Poynting vector.**
From equations (45) and (48), and also from equation (14) of vector algebra, the Poynting vetor $\boldsymbol{S}$ of the monochromatic plane wave is expressed as

$$\begin{aligned} \boldsymbol{S} &= \boldsymbol{E}_r \times \boldsymbol{H}_r = \frac{1}{\omega\,\mu} \boldsymbol{E}_r \times (\boldsymbol{k} \times \boldsymbol{E}_r) = \frac{k}{\omega\,\mu}\, \boldsymbol{n} \mid \boldsymbol{E}_r \mid^2 \\ &= \sqrt{\frac{\epsilon}{\mu}}\, \boldsymbol{n} \mid \boldsymbol{E}_r \mid^2 = \frac{1}{Z}\, \boldsymbol{n} \mid \boldsymbol{E}_r \mid^2, \end{aligned}$$

where

$$\boldsymbol{n} = \frac{\boldsymbol{k}}{k},$$

as defined in equation (46), and, therefore,

$$\boldsymbol{S} = \frac{1}{Z}\, \boldsymbol{n} \mid \boldsymbol{E}_r \mid^2 = Z\, \boldsymbol{n} \mid \boldsymbol{H}_r \mid^2 . \tag{52}$$

2.3 Generation of Electromagnetic Waves

Now, let us consider generation of electromagnetic waves from a matter (from an antenna, in particular), where we must take into account existence of the electric current and the charge.

If we try to describe this new problem, still using the electric and magnetic field intensities $\boldsymbol{E}$ and $\boldsymbol{H}$ as basic quantities, as we did in the previous section for the free space case, equations involved get fairly complicated. A much more transparent treatment is achieved, when we use so–called electromagnetic potentials as basic quantities.

2.3.1 Electromagnetic Potentials

As we saw in equations (19) and (18), for any vector field $\boldsymbol{F}(\boldsymbol{r})$ and any scalar field $f(\boldsymbol{r})$, we have identity relations:

$$\begin{aligned} \nabla \cdot (\nabla \times \boldsymbol{F}) &= 0, \\ \nabla \times (\nabla f) &= 0. \end{aligned}$$

It is known that following inverse relations generally hold (see, for example, "Classical Electricity and Magnetism" by Panofsky and Phillips, Dover Pubns):

> If a relation $\nabla \cdot \boldsymbol{Q}(\boldsymbol{r}) = 0$ holds everywhere for a vector field $\boldsymbol{Q}(\boldsymbol{r})$, then such a vector field can be always expressed as a rotation of a certain vector field $\boldsymbol{F}(\boldsymbol{r})$, i.e., $\boldsymbol{Q} = \nabla \times \boldsymbol{F}$.
>
> If another relation $\nabla \times \boldsymbol{R}(\boldsymbol{r}) = 0$ holds everywhere for a vector field $\boldsymbol{R}(\boldsymbol{r})$, then such a vector field can be always expressed as a gradient of a certain scalar field $f(\boldsymbol{r})$, i.e., $\boldsymbol{R} = \nabla f$.

From Maxwell equations (24) and (21), we know that relations:

$$\begin{aligned} \nabla \cdot \boldsymbol{B} &= 0, \\ \nabla \times \boldsymbol{E} &= -\frac{\partial \boldsymbol{B}}{\partial t}, \end{aligned}$$

hold everywhere.

Therefore, in view of the upper one of the above two equations, we can express the magnetic flux density $\boldsymbol{B}$ as

$$\boldsymbol{B} = \nabla \times \boldsymbol{A}, \tag{53}$$

through an appropriate vector field $\boldsymbol{A}(\boldsymbol{r}, t)$, which we call hereafter 'vector potential'.

Then, the lower one of the above two equations now becomes

$$\nabla \times (\boldsymbol{E} + \frac{\partial \boldsymbol{A}}{\partial t}) = 0.$$

Therefore, we can express the vector $\boldsymbol{E} + \partial \boldsymbol{A}/\partial t$ through an appropriate scalar field $\Phi(\boldsymbol{r}, t)$, which we call 'scalar potential', as

$$\boldsymbol{E} + \frac{\partial \boldsymbol{A}}{\partial t} = -\nabla \Phi,$$

and, hence, the electric field intensity $\boldsymbol{E}$ is expressed as

$$\boldsymbol{E} = -\nabla \Phi - \frac{\partial \boldsymbol{A}}{\partial t}. \tag{54}$$

The vector potential $\boldsymbol{A}(\boldsymbol{r}, t)$ and scalar potential $\Phi(\boldsymbol{r}, t)$, brought together, are called 'electromagnetic potentials'. They are regarded as auxiliary quantities, which are not directly measurable, but help to effectively express actual, physically measurable, quantities such as $\boldsymbol{E}$ and $\boldsymbol{B}$. Units of the electromagnetic potentials are

$$\begin{aligned} \boldsymbol{A}(\boldsymbol{r}, t): &\quad \mathrm{V\ s\ m^{-1}}, \\ \Phi(\boldsymbol{r}, t): &\quad \mathrm{V}, \end{aligned}$$

respectively.

2.3.2 Lorentz Gauge

Electromagnetic potentials $\boldsymbol{A}(\boldsymbol{r}, t)$ and $\Phi(\boldsymbol{r}, t)$, which satisfy equations (53) and (54), are not unique. In fact, it is easy to confirm that for an arbitrary scalar field $\Lambda(\boldsymbol{r}, t)$, a new set of potentials

$$\begin{aligned} \boldsymbol{A}' &= \boldsymbol{A} + \nabla \Lambda, \\ \Phi' &= \Phi - \frac{\partial \Lambda}{\partial t}, \end{aligned} \tag{55}$$

also satisfy equations (53) and (54), if $\boldsymbol{A}$ and Φ do.

It is a usual practice, in the wave generation problem, to introduce a constraint between $\boldsymbol{A}$ and Φ called 'Lorentz gauge':

$$\nabla \cdot \boldsymbol{A} + \epsilon\, \mu\, \frac{\partial \Phi}{\partial t} = 0, \tag{56}$$

using the above ambiguity, or, better to say, the freedom. The Lorentz gauge, as given in equation (56), allows us to formulate the generation of electromagnetic waves in a very smart way, as we will see below.

It is remarkable that any electromagnetic potentials $\boldsymbol{A}_n$ and Φ_n, which originally do not satisfy the Lorentz gauge:

$$\nabla \cdot \boldsymbol{A_n} + \epsilon\,\mu\,\frac{\partial \Phi_n}{\partial t} \neq 0,$$

can be converted to new ones fulfilling equation (56), by using a scalar field Λ in equation (55), which satisfies an equation:

$$\epsilon\,\mu\,\frac{\partial^2 \Lambda}{\partial t^2} - \nabla^2 \Lambda = \nabla \cdot \boldsymbol{A_n} + \epsilon\,\mu\,\frac{\partial \Phi_n}{\partial t}.$$

Of course, the physically measurable fields, such as $\boldsymbol{E}$ and $\boldsymbol{B}$, are not affected by this conversion.

Therefore, we can always select electromagnetic potentials $\boldsymbol{A}$ and Φ, which correspond to real electromagnetic fields through equations (53) and (54), and, at the same time, satisfy the Lorentz gauge (equation (56)):

$$\begin{aligned} \boldsymbol{B} &= \nabla \times \boldsymbol{A}, \\ \boldsymbol{E} &= -\nabla\Phi - \frac{\partial \boldsymbol{A}}{\partial t}, \\ \nabla \cdot \boldsymbol{A} + \epsilon\,\mu\,\frac{\partial \Phi}{\partial t} &= 0. \end{aligned}$$

Hereafter, we will consider only such electromagnetic potentials.

2.3.3 Wave Equations with Source Terms

We will assume, for simplicity, a homogeneous and stationary medium with respect to the permittivity ϵ and the permeability μ, i.e., $\epsilon = \text{const}$ and $\mu = \text{const}$ both in time and space.

Inserting equations (53) and (54) into a Maxwell equation (22):

$$\nabla \times \boldsymbol{H} = \boldsymbol{J} + \epsilon\,\frac{\partial \boldsymbol{E}}{\partial t},$$

we have

$$\frac{1}{\mu}\nabla \times (\nabla \times \boldsymbol{A}) = \boldsymbol{J} - \epsilon\,(\frac{\partial^2 \boldsymbol{A}}{\partial t^2} + \nabla\frac{\partial \Phi}{\partial t}).$$

In view of equation (16), this equation is reduced to

$$\nabla(\nabla \cdot \boldsymbol{A}) - \nabla^2 \boldsymbol{A} = \mu\,\boldsymbol{J} - \epsilon\,\mu\,\frac{\partial^2 \boldsymbol{A}}{\partial t^2} - \epsilon\,\mu\,\nabla\frac{\partial \Phi}{\partial t},$$

and, then,

$$\nabla^2 \boldsymbol{A} - \epsilon\,\mu\,\frac{\partial^2 \boldsymbol{A}}{\partial t^2} = -\mu\,\boldsymbol{J} + \nabla(\nabla\cdot\boldsymbol{A} + \epsilon\,\mu\,\frac{\partial \Phi}{\partial t}) = -\mu\,\boldsymbol{J},$$

where we used equation (56) of the Lorentz gauge.

Also, another Maxwell equation (23):

$$\epsilon\,\nabla\cdot\boldsymbol{E} = \rho,$$

can be expressed as

$$-\epsilon\,\nabla\cdot(\nabla\Phi + \frac{\partial \boldsymbol{A}}{\partial t}) = \rho,$$

through the electromagnetic potentials, and is reduced to

$$\nabla^2\Phi + \frac{\partial}{\partial t}(\nabla\cdot\boldsymbol{A}) = -\frac{1}{\epsilon}\,\rho,$$

and, then

$$\nabla^2\Phi - \epsilon\,\mu\,\frac{\partial^2\Phi}{\partial t^2} = -\frac{1}{\epsilon}\,\rho,$$

where we again used equation (56) of the Lorentz gauge.

Introducing again the light velocity c ($c^2 = 1\,/\,\epsilon\,\mu$), we obtain equations

$$\nabla^2 \boldsymbol{A} - \frac{1}{c^2}\,\frac{\partial^2 \boldsymbol{A}}{\partial t^2} = -\mu\,\boldsymbol{J}, \tag{57}$$

$$\nabla^2\Phi - \frac{1}{c^2}\,\frac{\partial^2\Phi}{\partial t^2} = -\frac{1}{\epsilon}\,\rho, \tag{58}$$

which have the simple form of the wave equation. The only difference from the free space case in equation (34) is the existence of RHS terms, which are source terms of the wave equations. These equations describe the interaction of the electric current $\boldsymbol{J}$ and charge ρ within a source medium with the electromagnetic waves $\boldsymbol{A}$ and Φ, i.e., the current and charge may generate the waves in a free space (transmission), or the waves in the free space may generate the current and charge in the source medium (reception), as schematically shown in Figure 21.

Equations (57) and (58) are consistent with equation (56) of the Lorentz gauge, as we can easily verify using equation (29) of continuity (or conservation) of electric charge.

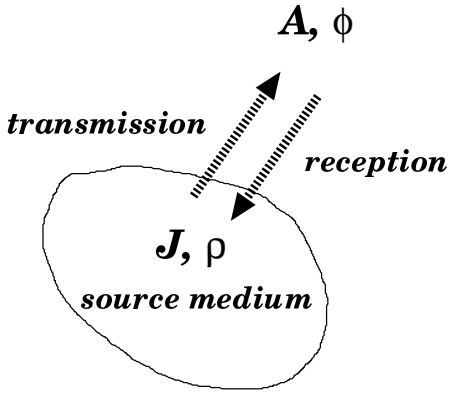

Figure 21: A schematic view of the interaction of the current and charge in a source medium with the electromagnetic waves in a free space.

2.3.4 Solution of the Wave Equation with the Source Term

Let us consider solutions of equations (57) and (58). Since each component of vector equation (57) and scalar equation (58) have the same general form of

$$\nabla^2\Psi(\boldsymbol{r},t) - \frac{1}{c^2}\frac{\partial^2\Psi(\boldsymbol{r},t)}{\partial t^2} = -f(\boldsymbol{r},t), \tag{59}$$

where f and Ψ are functions of the time t and the space $\boldsymbol{r}$, we will confine ourselves to seeking a solution $\Psi(\boldsymbol{r},t)$ of equation (59) with a source term $-f(\boldsymbol{r},t)$. We assume that the source term takes some finite value only within a certain source medium.

In order to express the RHS of equation (59) in an integral form, we use Dirac's delta function $\delta(\boldsymbol{r})$, which has following functional properties:

$$\begin{aligned}
\delta(\boldsymbol{r}) &= 0, \quad \text{when} \quad \boldsymbol{r} \neq 0, \\
\int_V \delta(\boldsymbol{r}-\boldsymbol{r}')\,dV' &= 1, \\
\int_V F(\boldsymbol{r}')\delta(\boldsymbol{r}-\boldsymbol{r}')\,dV' &= F(\boldsymbol{r}), \quad \text{for any function } F(\boldsymbol{r}),
\end{aligned}$$

where $dV' = dx_1'dx_2'dx_3'$ is a volume element, and the integration is taken over some volume V containing a point $\boldsymbol{r}$. Now we can express the RHS of equation (59) as:

$$\nabla^2\Psi(\boldsymbol{r},t) - \frac{1}{c^2}\frac{\partial^2\Psi(\boldsymbol{r},t)}{\partial t^2} = -\int_V f(\boldsymbol{r}',t)\delta(\boldsymbol{r}-\boldsymbol{r}')\,dV', \tag{60}$$

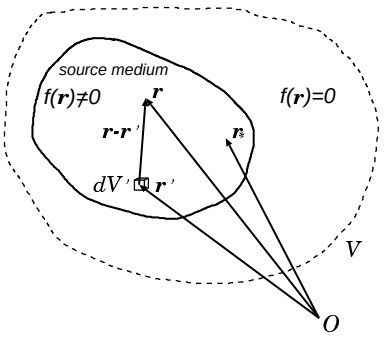

Figure 22: Geometry of the source medium region.

where the integration is taken over an appropriate volume V, containing the source medium and a point $\boldsymbol{r}$ (see Figure 22).

Let us now consider the same wave equation with a time–variable point source at the origin $\boldsymbol{r} = 0$ (O):

$$\nabla^2\psi - \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} = -f(\boldsymbol{r}_*,\, t)\delta(\boldsymbol{r}), \tag{61}$$

where $\boldsymbol{r}_*$ is a radius vector indicating a certain point in the source medium, and $-f(\boldsymbol{r}_*,\, t)$ is the source term of equation (59) at a time t and at the point $\boldsymbol{r} = \boldsymbol{r}_*$. Note that $\boldsymbol{r}_*$ is regarded here as a constant parameter, and therefore, $-f(\boldsymbol{r}_*,\, t)$ in equation (61) is treated as a function of time t, only.

If $\psi(\boldsymbol{r}_*,\, \boldsymbol{r},\, t)$ is a solution of the above equation (61), then $\psi(\boldsymbol{r}_*, \boldsymbol{r}-\boldsymbol{r}',\, t)$ is a solution of equation:

$$\nabla^2\psi - \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} = -f(\boldsymbol{r}_*,\, t)\delta(\boldsymbol{r}-\boldsymbol{r}'),$$

since the LHS of equation (61) is invariant with respect to a parallel shift of the spatial coordinate system: $\boldsymbol{r} \to \boldsymbol{r}-\boldsymbol{r}'$. Consequently, $\psi(\boldsymbol{r}',\, \boldsymbol{r}-\boldsymbol{r}',\, t)$ is a solution of equation:

$$\nabla^2\psi - \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} = -f(\boldsymbol{r}',\, t)\delta(\boldsymbol{r}-\boldsymbol{r}').$$

And, finally, in view of the principle of superposition of solutions in a liner equation,

$$\Psi(\boldsymbol{r}, t) = \int\limits_V \psi(\boldsymbol{r}',\, \boldsymbol{r}-\boldsymbol{r}',\, t)\, dV'. \tag{62}$$

is a solution of our wave equation 60 with a source term:

$$\nabla^2\Psi(\boldsymbol{r},t) - \frac{1}{c^2}\frac{\partial^2\Psi(\boldsymbol{r},t)}{\partial t^2} = -\int_V f(\boldsymbol{r}',\, t)\delta(\boldsymbol{r}-\boldsymbol{r}')\, dV' = -f(\boldsymbol{r},t).$$

Thus, we can use the solution $\psi(\boldsymbol{r}_*,\, \boldsymbol{r},\, t)$ as a Green function, in order to obtain a solution of equation (59), or (60).

Now, the solution of the equation (61) (Green function) is expressed in a form:

$$\psi(\boldsymbol{r}_*, \boldsymbol{r}, t) = \frac{f(\boldsymbol{r}_*,\, t \mp \frac{r}{c})}{4\pi r}, \tag{63}$$

where $r = \mid \boldsymbol{r} \mid = \sqrt{x_i x_i} = \sqrt{x_1^2 + x_2^2 + x_3^2}$, as we can veryfy in the following way.

1. We use following general formulae.
   - For a radius vector $\boldsymbol{r}$,

$$\begin{aligned} \nabla r &= \frac{\boldsymbol{r}}{r}, \text{ and } \nabla\cdot\boldsymbol{r} = 3, \text{ since} \qquad (64)\\ (\nabla r)_i &= \frac{\partial r}{\partial x_i} = \frac{\partial}{\partial x_i}\sqrt{x_j x_j} = \frac{1}{2\sqrt{x_j x_j}}\frac{\partial}{\partial x_i}(x_k x_k) = \frac{x_i}{r}, \\ \text{and} \qquad & \nabla\cdot\boldsymbol{r} = \frac{\partial x_i}{\partial x_i} = \delta_{ii} = 3. \end{aligned}$$

   - For a function $R(r)$ of the radius r,

$$\begin{aligned} \nabla^2 R(r) &= \frac{d^2R(r)}{dr^2} + \frac{2}{r}\frac{dR(r)}{dr}, \text{ since} \qquad (65)\\ \nabla^2 R(r) &= \nabla\cdot\nabla R(r) = \nabla\cdot\left(\frac{dR}{dr}\nabla r\right) = \nabla\cdot\left(\frac{dR}{dr}\frac{\boldsymbol{r}}{r}\right) \\ &= \nabla\left(\frac{1}{r}\frac{dR}{dr}\right)\cdot\boldsymbol{r} + \left(\frac{1}{r}\frac{dR}{dr}\right)\nabla\cdot\boldsymbol{r} \\ &= r\frac{d}{dr}\left(\frac{1}{r}\frac{dR}{dr}\right) + 3\left(\frac{1}{r}\frac{dR}{dr}\right) = \frac{d^2R}{dr^2} + \frac{2}{r}\frac{dR}{dr}. \end{aligned}$$

   - For a particular function $1/r$, which is singular at the origin, we have

$$\nabla^2\left(\frac{1}{r}\right) = -4\pi\,\delta(\boldsymbol{r}), \tag{66}$$

   as we know from the potential theory (electrostatic potential around a point charge, Newtonian gravitaional potential around a point mass,

etc., ...). Equation (66) can be easily confirmed, if we integrate the both sides of the equation through a volume V, containing the origin, and apply to the LHS Gauss's integration theorem:

$$\int_V \nabla \cdot \boldsymbol{A}\, dV = \oint_S \boldsymbol{A} \cdot \boldsymbol{n}\, d\mathcal{S}, \tag{67}$$

where $\boldsymbol{A}$ is an arbitrary vector field, V is a closed volume, covered by a surface $\mathcal{S}$ with a normal vector $\boldsymbol{n}$. In fact, the RHS gives

$$-4\pi \int_V \delta(\boldsymbol{r})\, dV = -4\pi,$$

and the LHS also yields

$$\int_V \nabla\cdot\nabla\left(\frac{1}{r}\right) dV = \oint_S \nabla\left(\frac{1}{r}\right)\cdot\boldsymbol{n}\, d\mathcal{S} = -\oint_S \frac{\boldsymbol{r}}{r^3}\cdot\boldsymbol{n}\, d\mathcal{S} = -\oint d\Omega = -4\pi,$$

where $d\Omega$ is a solid angle element.

2. Let us introduce a notation:

$$u \equiv t \mp \frac{r}{c}.$$

Then, the function ψ in equation (63) is now expressed as

$$\psi(\boldsymbol{r}_*,\, \boldsymbol{r},\, t) = \frac{f(\boldsymbol{r}_*,\, u)}{4\pi r}.$$

Noting that

$$\frac{\partial f}{\partial t} = \frac{df}{du}\frac{\partial u}{\partial t}, \quad \frac{\partial f}{\partial r} = \frac{df}{du}\frac{\partial u}{\partial r}, \quad \text{with} \quad \frac{\partial u}{\partial t} = 1, \quad \text{and} \quad \frac{\partial u}{\partial r} = \mp\frac{1}{c},$$

we obtain

$$\frac{\partial^2 \psi}{\partial t^2} = \frac{1}{4\pi r}\frac{d^2 f}{du^2},$$

and

$$\begin{aligned}
\nabla^2\psi &= \frac{1}{4\pi r}\nabla^2 f + 2\nabla\left(\frac{1}{4\pi r}\right)\cdot\nabla f + f\,\nabla^2\left(\frac{1}{4\pi r}\right) \\
&= \frac{1}{4\pi r}\left(\frac{\partial^2 f}{\partial r^2} + \frac{2}{r}\frac{\partial f}{\partial r}\right) - \frac{2}{4\pi r^2}\frac{\partial f}{\partial r} - f\,\delta(\boldsymbol{r}) \\
&= \frac{1}{4\pi r}\frac{1}{c^2}\frac{d^2 f}{du^2} - f\,\delta(\boldsymbol{r}).
\end{aligned}$$

Therefore, the function $\psi(\boldsymbol{r}_*, \boldsymbol{r}, t)$ in equation (63) really satisfies the wave equation (61) with a point source term at the origin:

$$\nabla^2\psi - \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} = -f(\boldsymbol{r}_*, t \mp \frac{r}{c})\,\delta(\boldsymbol{r}) = -f(\boldsymbol{r}_*, t)\,\delta(\boldsymbol{r}),$$

and, hence, can be used as a Green function for solving equation (59) in the form given in equation (62).

2.3.5 Retarded Potential

Now, from equations (62) and (63), a solution of the wave equation with the source term (59) is expressed as:

$$\Psi(\boldsymbol{r}, t) = \frac{1}{4\pi}\int_V \frac{f(\boldsymbol{r}', t \mp \frac{|\boldsymbol{r}-\boldsymbol{r}'|}{c})}{|\,\boldsymbol{r}-\boldsymbol{r}'\,|}\,dV'. \tag{68}$$

In actuality, there are two choices of solutions corresponding to '+' and '−' signs of the $\mp$ term in the numerator of the integrand. One solution with '−' sign is called 'retarded potential', and another with '+' sign is called 'advanced potential'.

Meanings of the retarded and advanced potentials can be better understood in the simplest case of the point source term at the origin ($\boldsymbol{r} = 0$), when the solution takes the same form as the one in equation (63):

$$\Psi(\boldsymbol{r}, t) = \frac{g(t \mp \frac{r}{c})}{4\pi r},$$

where $g(t)$ is a function describing the time variation of the source term.

The retarded potential with '−' sign describes that some pattern of time variation, originated in the source region (a point in this case), **propagates outward** with decreasing amplitude in proportion to $1/r$. The pattern at a distant point (with larger r) is **retarded** compared with the one at a near point (with smaller r). This represents a typical example of the transmission of the electromagnetic wave from the source region (Figure 23).

On the other hand, the advanced potential with '+' sign describes that some pattern of time variation **propagates inward**, and induces the same sort of time variation in the source region. The pattern at a distant point (with larger r) is **advanced** compared with the one at a near point (with smaller r). This obviously represents the reception of the electromagnetic wave, coming from outside, at the source region (Figure 23). However, this is not a 'typical' example of the wave reception, because the proportionality

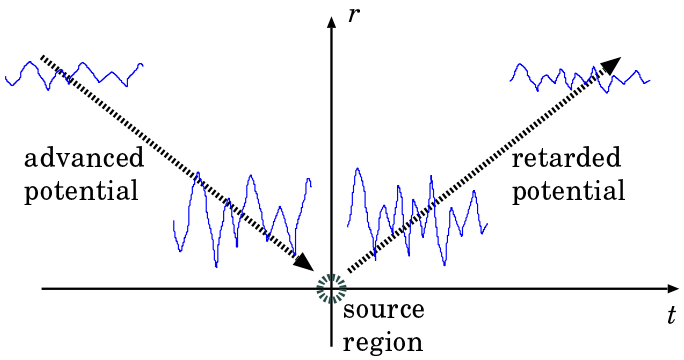

Figure 23: Pattern propagations in retarded and advanced potentials.

to $1/r$ implies that the incoming wave must converge at the source region. This occurs, for example, in the converging wave reflected by a paraboloidal mirror of a radio telescope antenna, but in a limited space–time range.

We will consider, hereafter, the problem of the transmission of electromagnetic waves from a source region (or an antenna). Therefore, we will use the retarded potential only.

Coming back to the electromagnetics, we obtain retarded potential solutions for wave equations (57) and (58) with source terms as follows:

$$\boldsymbol{A}(\boldsymbol{r},\, t) \;=\; \frac{\mu}{4\pi}\int \frac{\boldsymbol{J}(\boldsymbol{r}',\, t-\frac{|\boldsymbol{r}-\boldsymbol{r}'|}{c})}{\mid \boldsymbol{r}-\boldsymbol{r}' \mid}\, dV', \tag{69}$$

$$\Phi(\boldsymbol{r},\, t) \;=\; \frac{1}{4\pi\epsilon}\int \frac{\rho(\boldsymbol{r}',\, t-\frac{|\boldsymbol{r}-\boldsymbol{r}'|}{c})}{\mid \boldsymbol{r}-\boldsymbol{r}' \mid}\, dV'. \tag{70}$$

Since the vector potential $\boldsymbol{A}$ and the scalar potential Φ derived here are not independent to each other, but related by Lorentz gauge in equation (56):

$$\nabla \boldsymbol{A} + \epsilon\mu\,\frac{\partial \Phi}{\partial t} = 0,$$

we will consider hereafter the vector potential $\boldsymbol{A}$ only.

2.3.6 Transmission of Radio Wave from a Harmonically Oscillating Source

Let us consider a case when the current density $\boldsymbol{J}(\boldsymbol{r},\, t)$ is harmonically (or, sinusoidally) oscillating everywhere in a source region with the same angular frequency $\omega = 2\pi\nu$:

$$\boldsymbol{J}(\boldsymbol{r},\, t) = \boldsymbol{J}(\boldsymbol{r})e^{-i\omega t}. \tag{71}$$

In this complex representation, we again use the convention, that the actual physical quantity is expressed by the real part of the complex quantity (in the above particular example, the actual quantity is $\Re\,[\boldsymbol{J}(\boldsymbol{r})e^{-i\omega t}]$). The assumption of the same frequency, and the resultant separation of spatial and temporal variables, might seem a little artificial. But this must be valid for a frequency component in a Fourier expansion of the time variable current density.

The vector potential $\boldsymbol{A}(\boldsymbol{r},\,t)$, generated by such a current, is also expressed in a similar form:

$$\boldsymbol{A}(\boldsymbol{r},\,t) = \boldsymbol{A}(\boldsymbol{r})e^{-i\omega t}. \tag{72}$$

In fact, inserting equation (71) to the formula of the retarded potential in equation (69), we obtain

$$\boldsymbol{A}(\boldsymbol{r},\,t) = \left[\frac{\mu}{4\pi}\int \boldsymbol{J}(\boldsymbol{r}')\frac{e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}}{\mid \boldsymbol{r}-\boldsymbol{r}' \mid}\,dV'\right]\,e^{-i\omega t},$$

and, therefore, we know that $\boldsymbol{A}(\boldsymbol{r},\,t)$ is really expressed in the form of equation (72), with

$$\boldsymbol{A}(\boldsymbol{r}) = \frac{\mu}{4\pi}\int \boldsymbol{J}(\boldsymbol{r}')\frac{e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}}{\mid \boldsymbol{r}-\boldsymbol{r}' \mid}\,dV', \tag{73}$$

where $k \equiv \omega/c = 2\pi/\lambda$, with λ being the wavelength corresponding to the angular frequency ω.

Now let us consider the generated electromagnetic wave in a homogeneous medium located outside of the source region, where $\boldsymbol{J} = 0$ and ϵ, μ =const. From the definition of the vector potential and the Maxwell equation, we have:

$$\boldsymbol{B} = \nabla \times \boldsymbol{A}, \quad \text{and, hence,} \quad \boldsymbol{H} = \frac{1}{\mu}\nabla \times \boldsymbol{A}, \tag{74}$$

$$\nabla \times \boldsymbol{H} = \frac{\partial \boldsymbol{D}}{\partial t}, \quad \text{and, hence,} \quad -i\omega\epsilon\boldsymbol{E} = \nabla \times \boldsymbol{H}. \tag{75}$$

Therefore, the magnetic and electric fields of the wave, generated by the harmonically oscillating current $\boldsymbol{J}(\boldsymbol{r},\,t)$ in equation (71), and having the vector potential $\boldsymbol{A}(\boldsymbol{r},\,t)$ of the form of equation (72), are expressed as:

$$\boldsymbol{H}(\boldsymbol{r},\,t) = \boldsymbol{H}(\boldsymbol{r})e^{-i\omega t}, \quad \text{where} \quad \boldsymbol{H}(\boldsymbol{r}) = \frac{1}{\mu}\nabla \times \boldsymbol{A}(\boldsymbol{r}), \tag{76}$$

$$\boldsymbol{E}(\boldsymbol{r},\,t) = \boldsymbol{E}(\boldsymbol{r})e^{-i\omega t}, \quad \text{where} \quad \boldsymbol{E}(\boldsymbol{r}) = i\frac{c}{k}\,\nabla \times (\nabla \times \boldsymbol{A}(\boldsymbol{r})), \tag{77}$$

where we used the relation $c^2 = 1/\epsilon\mu$.

2.3.7 Electromagnetic Fields Far from the Source Region

Let us now calculate the electromagnetic fields generated by the harmonically oscillating current in the source region, taking into account the explicit expression for $\boldsymbol{A}(\boldsymbol{r})$, given in equation (73). For definiteness, we choose the origin of radius vectors $\boldsymbol{r}$ and $\boldsymbol{r}'$ within the source region.

i–th component of the rotation of $\boldsymbol{A}(\boldsymbol{r})$ in equation (76) can be expressed as:

$$(\nabla \times \boldsymbol{A}(\boldsymbol{r}))_i = \epsilon_{ijk} \frac{\partial}{\partial x_j} A_k(\boldsymbol{r}) = \frac{\mu}{4\pi} \epsilon_{ijk} \int J_k(\boldsymbol{r}') \frac{\partial}{\partial x_j} \frac{e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}}{\mid \boldsymbol{r}-\boldsymbol{r}' \mid} \, dV'. \qquad (78)$$

Since

$$\frac{\partial}{\partial x_i} \mid \boldsymbol{r}-\boldsymbol{r}' \mid = \frac{x_i - x_i'}{\mid \boldsymbol{r}-\boldsymbol{r}' \mid},$$

we have, in the integrand of equation (78),

$$\begin{aligned} \frac{\partial}{\partial x_j} \frac{e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}}{\mid \boldsymbol{r}-\boldsymbol{r}' \mid} &= \frac{1}{\mid \boldsymbol{r}-\boldsymbol{r}' \mid} \frac{\partial}{\partial x_j} e^{ik|\boldsymbol{r}-\boldsymbol{r}'|} + e^{ik|\boldsymbol{r}-\boldsymbol{r}'|} \frac{\partial}{\partial x_j} \frac{1}{\mid \boldsymbol{r}-\boldsymbol{r}' \mid} \\ &= \frac{ik(x_j - x_j')}{\mid \boldsymbol{r}-\boldsymbol{r}' \mid^2} e^{ik|\boldsymbol{r}-\boldsymbol{r}'|} - \frac{x_j - x_j'}{\mid \boldsymbol{r}-\boldsymbol{r}' \mid^3} e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}. \qquad (79) \end{aligned}$$

Let us assume that we observe the electromagnetic field at a point, which is far from the source region. If we assume that the distance of the point from the source region is much larger than the wavelength:

$$\mid \boldsymbol{r}-\boldsymbol{r}' \mid \gg \lambda, \qquad (80)$$

(see Figure 24) and, hence,

$$\frac{1}{k \mid \boldsymbol{r}-\boldsymbol{r}' \mid} \ll 1,$$

then we can neglect the second term in the RHS of equation (79), compared with the first term, so that

$$\frac{\partial}{\partial x_j} \frac{e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}}{\mid \boldsymbol{r}-\boldsymbol{r}' \mid} = \frac{ik(x_j - x_j')}{\mid \boldsymbol{r}-\boldsymbol{r}' \mid^2} \, e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}.$$

If we, furthermore, assume that the distance of the observation point from the source region is much larger than the size of the source region (see Figure 24), we have,

$$\mid \boldsymbol{r} \mid \gg \mid \boldsymbol{r}' \mid, \qquad (81)$$

and, therefore, we can finally approximate equation (79) with

$$\frac{\partial}{\partial x_j}\frac{e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}}{|\,\boldsymbol{r}-\boldsymbol{r}'\,|} = \frac{ikx_j}{r^2}\,e^{ik|\boldsymbol{r}-\boldsymbol{r}'|},$$

where $r = |\,\boldsymbol{r}\,|$, as before. Inserting this approximate formula into equation (78), we obtain

$$\nabla\times\boldsymbol{A}(\boldsymbol{r}) = \frac{ik\mu}{4\pi}\frac{\boldsymbol{r}}{r^2}\times\int\boldsymbol{J}(\boldsymbol{r}')\,e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}\,dV' = \frac{ik\mu}{4\pi r}\,\boldsymbol{n}\times\int\boldsymbol{J}(\boldsymbol{r}')\,e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}\,dV', \tag{82}$$

where we introduced a unit vector

$$\boldsymbol{n} = \frac{\boldsymbol{r}}{r}. \tag{83}$$

The same sort of discussions, under the same approximations of equations (80) and (81), lead to

$$\nabla\times(\nabla\times\boldsymbol{A}(\boldsymbol{r})) = -\frac{k^2\mu}{4\pi r}\,\boldsymbol{n}\times(\boldsymbol{n}\times\int\boldsymbol{J}(\boldsymbol{r}')\,e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}\,dV'). \tag{84}$$

Therefore, in view of equations (73), (76) and (77), we have approximate formulae for the vector potential, magnetic field and electric field:

$$\boldsymbol{A}(\boldsymbol{r}) = \frac{\mu}{4\pi r}\int\boldsymbol{J}(\boldsymbol{r}')\,e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}\,dV', \tag{85}$$

$$\boldsymbol{H}(\boldsymbol{r}) = \frac{1}{\mu}\nabla\times\boldsymbol{A}(\boldsymbol{r}) = \frac{ik}{4\pi r}\,\boldsymbol{n}\times\int\boldsymbol{J}(\boldsymbol{r}')\,e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}\,dV', \tag{86}$$

$$\boldsymbol{E}(\boldsymbol{r}) = \frac{c}{k}\nabla\times(\nabla\times\boldsymbol{A}(\boldsymbol{r})) = -\frac{iZk}{4\pi r}\,\boldsymbol{n}\times(\boldsymbol{n}\times\int\boldsymbol{J}(\boldsymbol{r}')\,e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}\,dV'), \tag{87}$$

where $Z = \sqrt{\mu/\epsilon}$ is the intrinsic impedance, as defined in equation (51), and we used relation $c = 1/\sqrt{\epsilon\mu}$, as given in equation (35), in deriving equation (87).

2.3.8 Far Field Solution and Fraunhofer Region

It is worth to note that we did not approximate $|\,\boldsymbol{r}-\boldsymbol{r}'\,|$ by r in the arguments of exponential functions in equations (85), (86) and (87), even though we assumed that $|\,\boldsymbol{r}\,| \gg |\,\boldsymbol{r}'\,|$. Such approximation is not valid, in general, in an argument (or phase) of any sinusoidal function, since only remainder of the argument divided by 2π is meaningful in such a function (for example, $\cos(983517826^\circ + 132^\circ) = \cos(983517826^\circ \times (1 + 1.34212107 \times 10^{-7}))$ is not close to $\cos(983517826^\circ)$ at all).

However, if we find an 'absolutely' small term in the argument, which is much smaller than 1 radian, say, we can safely neglect such a term, and this may allow us to derive a useful approximate formula.

Let a characteristic size of our source region (aperture diameter of an antenna, for example) be D, distance to the observation point from the origin in the source region be r, and the wavelength be $\lambda = 2\pi/k$. We can generally expand the argument in the exponential term $k \mid \boldsymbol{r} - \boldsymbol{r}' \mid$ into a Taylor series:

$$k \mid \boldsymbol{r} - \boldsymbol{r}' \mid = k\sqrt{r^2 - 2\boldsymbol{r} \cdot \boldsymbol{r}' + r'^2} \cong kr\,(1 - \frac{\boldsymbol{r} \cdot \boldsymbol{r}'}{r^2} + \frac{1}{2}\frac{r'^2}{r^2} + \cdots). \tag{88}$$

If a condition, which is called "Fraunhofer condition",

$$\frac{2D^2}{\lambda} \ll r, \tag{89}$$

is satisfied among r, D and λ (see Figure 24), the third term in the expansion

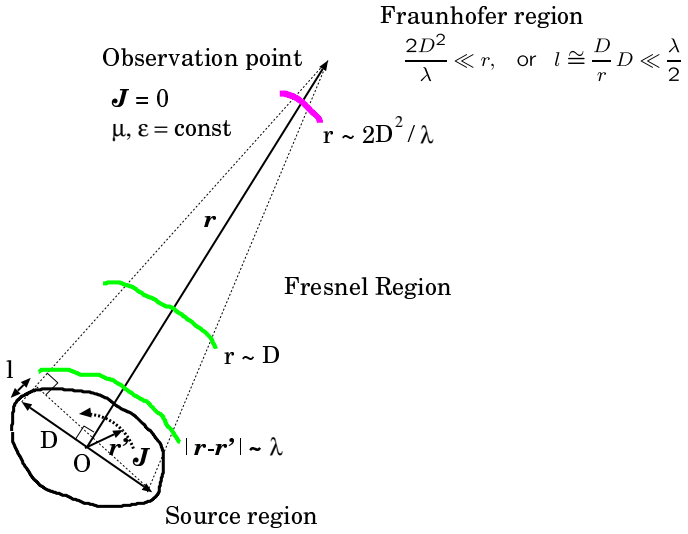

Figure 24: Fraunhofer condition.

of equation (88) fulfills

$$\frac{1}{2}\,k\,\frac{r'^2}{r} \sim \frac{\pi}{\lambda}\frac{D^2}{r} \ll \frac{\pi}{2}.$$

Therefore, we can neglect the third and higher order terms in the expansion of equation (88), in the argument of the exponential function, and obtain

$$e^{ik|\boldsymbol{r}-\boldsymbol{r}'|} \cong e^{ikr - ik\boldsymbol{n}\cdot\boldsymbol{r}'}, \tag{90}$$

where $\boldsymbol{n} = \boldsymbol{r}/r$, as before. Note that, on the contrary, the second term has the order of magnitude of

$$k\boldsymbol{n}\cdot\boldsymbol{r}' \sim \frac{2\pi D}{\lambda},$$

and cannot be neglected at all, in general, in the argument of the exponential (or, sinusoidal) function. In this approximation, equation (85) for the vector potential, for example, is expressed as

$$\boldsymbol{A}(\boldsymbol{r}) = \frac{\mu}{4\pi}\,\frac{e^{ikr}}{r}\int \boldsymbol{J}(\boldsymbol{r}')\,e^{-ik\boldsymbol{n}\cdot\boldsymbol{r}'}\,dV'.$$

Now the integral term is a function of the direction $\boldsymbol{n}$ only, and does not depend on the distance r to the observation point.

The region where $r > 2D^2/\lambda$ is called "Fraunhofer region" or "far field", and the region where $r < 2D^2/\lambda$ is called "Fresnel region" or "near field".

In the Fraunhofer region, we can approximate the exponential term as shown in equation (90). Therefore, introducing a vector $\boldsymbol{T}(\boldsymbol{n})$, characterizing a directional pattern of the radiation in the far field,

$$\boldsymbol{T}(\boldsymbol{n}) = \int \boldsymbol{J}(\boldsymbol{r}')e^{-ik\boldsymbol{n}\cdot\boldsymbol{r}'}\,dV', \tag{91}$$

and taking into account the temporal variation of the electromagnetic fields, we obtain, from equations (72), (76), (77), (85), (86), (87) and (90), so–called far field solutions:

$$\boldsymbol{A}(\boldsymbol{r},\,t) = \frac{\mu}{4\pi}\,\frac{e^{i(kr-\omega t)}}{r}\,\boldsymbol{T}(\boldsymbol{n}), \tag{92}$$

$$\boldsymbol{H}(\boldsymbol{r},\,t) = i\,\frac{1}{2\lambda}\,\frac{e^{i(kr-\omega t)}}{r}\,\boldsymbol{n}\times\boldsymbol{T}(\boldsymbol{n}), \tag{93}$$

$$\boldsymbol{E}(\boldsymbol{r},\,t) = -i\,\frac{Z}{2\lambda}\,\frac{e^{i(kr-\omega t)}}{r}\,\boldsymbol{n}\times[\boldsymbol{n}\times\boldsymbol{T}(\boldsymbol{n})]. \tag{94}$$

These equations describe harmonically oscillating and spherically expanding directional patterns $\boldsymbol{T}(\boldsymbol{n})$, $\boldsymbol{n}\times\boldsymbol{T}(\boldsymbol{n})$, and $\boldsymbol{n}\times[\boldsymbol{n}\times\boldsymbol{T}(\boldsymbol{n})]$, i.e. the **spherical waves**.

It is clear from equations (93) and (94), that we have

$$\boldsymbol{n}\cdot\boldsymbol{H} = 0,\ \ \boldsymbol{n}\cdot\boldsymbol{E} = 0,\ \ \text{and}\ \ \boldsymbol{H}\cdot\boldsymbol{E} = 0, \tag{95}$$

which imply that the magnetic $\boldsymbol{H}$ and electronic $\boldsymbol{E}$ waves are transversal and mutually orthogonal. It is also clear from the same equations that

$$\boldsymbol{E} = -Z\,\boldsymbol{n}\times\boldsymbol{H},\ \ \text{and}\ \ \boldsymbol{H} = \frac{1}{Z}\,\boldsymbol{n}\times\boldsymbol{E}. \tag{96}$$

Taking real parts of these complex equations, we see that the same relations hold for actual electromagnetic fields, which are real parts of the complex expressions: $\boldsymbol{H}_r = \Re\,\boldsymbol{H}$ and $\boldsymbol{E}_r = \Re\,\boldsymbol{E}$, i.e.

$$\boldsymbol{E}_r = -Z\,\boldsymbol{n} \times \boldsymbol{H}_r, \quad \text{and} \quad \boldsymbol{H}_r = \frac{1}{Z}\,\boldsymbol{n} \times \boldsymbol{E}_r.$$

Therefore, the Poynting vector $\boldsymbol{S}$ for the spherical wave is expressed as:

$$\boldsymbol{S} = \boldsymbol{E}_r \times \boldsymbol{H}_r = Z\,\boldsymbol{n} \mid \boldsymbol{H}_r \mid^2 = \frac{1}{Z}\,\boldsymbol{n} \mid \boldsymbol{E}_r \mid^2 . \tag{97}$$

This equation means that the Poynting vector of the spherical wave is directed towards $\boldsymbol{n} = \boldsymbol{r}/r$, which is nothing but the direction of propagation of the wave, and the amplitude ratio of the electric and magnetic fields is

$$\frac{\mid \boldsymbol{E}_r \mid}{\mid \boldsymbol{H}_r \mid} = Z, \tag{98}$$

i.e. equal to the intrinsic impedance of the medium $Z = \sqrt{\mu/\epsilon}$. These properties are the same with those in the plane wave case, which we discussed earlier. Of course, the spherical waves locally approach to the plane waves when $r \to \infty$.

2.3.9 Hertz Dipole

We are now in position to derive some results on transmission of the electromagnetic waves from simple antennas, using the far field solutions in equations (92), (93) and (94). First, we will consider an idealized antenna, composed of an infinitesimal electric dipole called "Hertz dipole", which was discussed by Heinrich Hertz in 1888.

Let us consider an infinitesimally small cylinder with infinitesimal cross section q and infinitesimal length l, which is located at the origin of a rectangular coordinate system and directed towards 3–rd axis (Figure 25). Let us assume a homogeneous electric current I, which is flowing within the cylinder and harmonically oscillating in time $\propto e^{-i\omega t}$.

In this case, the vector $\boldsymbol{T}(\boldsymbol{n})$ characterizing the radiation pattern in the far field in equation (91) takes a particularly simple form. In fact, if we assume that $l \ll \lambda$, where λ is the wavelength, we have

$$k\boldsymbol{n} \cdot \boldsymbol{r}' \sim \frac{2\pi}{\lambda} l \ll 1, \quad \text{and, therefore,} \quad e^{-ik\boldsymbol{n}\cdot\boldsymbol{r}'} \cong 1,$$

in equation (91), and, consequently,

$$\boldsymbol{T}(\boldsymbol{n}) = \int \boldsymbol{J}\,dV' = \boldsymbol{J}\,l\,q = I\,l\,\boldsymbol{i}_3, \tag{99}$$

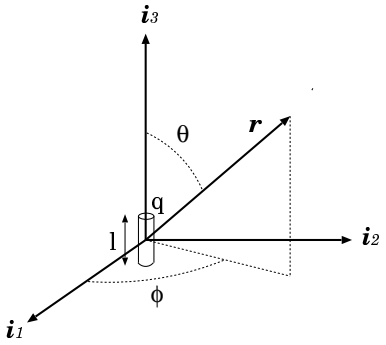

Figure 25: Hertz dipole.

where $\boldsymbol{i}_3$ is the unit vector along the 3–rd axis. Applying equation (99) to equations (92), (93) and (94), we obtain non–zero components of the far field solutions of the transmitted wave:

$$A_3(\boldsymbol{r},\, t) = \frac{\mu\, I\, l}{4\pi}\, \frac{e^{i(kr-\omega t)}}{r}, \tag{100}$$

$$H_\phi(\boldsymbol{r},\, t) = -i\, \frac{I\, l}{2\lambda}\, \frac{e^{i(kr-\omega t)}}{r}\, \sin\theta, \tag{101}$$

$$E_\theta(\boldsymbol{r},\, t) = -i\, \frac{Z\, I\, l}{2\lambda}\, \frac{e^{i(kr-\omega t)}}{r}\, \sin\theta, \tag{102}$$

where θ and ϕ are angular variables shown in Figure 25. Or, for actual physical quantities $\boldsymbol{A}_r$, $\boldsymbol{H}_r$ and $\boldsymbol{E}_r$, which are real parts of the respective complex quantities, we have the non-zero components in the far field:

$$A_{r3}(\boldsymbol{r},\, t) = \Re A_3(\boldsymbol{r},\, t) = \frac{\mu\, I\, l}{4\pi}\, \frac{\cos(kr-\omega t)}{r}, \tag{103}$$

$$H_{r\phi}(\boldsymbol{r},\, t) = \Re H_\phi(\boldsymbol{r},\, t) = \frac{I\, l}{2\lambda}\, \frac{\sin(kr-\omega t)}{r}\, \sin\theta, \tag{104}$$

$$E_{r\theta}(\boldsymbol{r},\, t) = \Re E_\theta(\boldsymbol{r},\, t) = \frac{Z\, I\, l}{2\lambda}\, \frac{\sin(kr-\omega t)}{r}\, \sin\theta. \tag{105}$$

The $\sin\theta$ dependence of the magnetic and electric fields shows the effectiveness of transmission of the electromagnetic wave to a certain direction θ, ϕ, which is called **field pattern**. Since the above equations do not have any dependence on ϕ, the field pattern of Hertz dipole is axisymmetric, and has a torus–like shape shown in Figure 26.

We can now easily calculate Poynting vector of the far field: $\boldsymbol{S} = \boldsymbol{E}_r \times \boldsymbol{H}_r = E_{r\theta} H_{r\phi}\, \boldsymbol{n}$, transmitted from Hertz dipole. If we denote a time–average

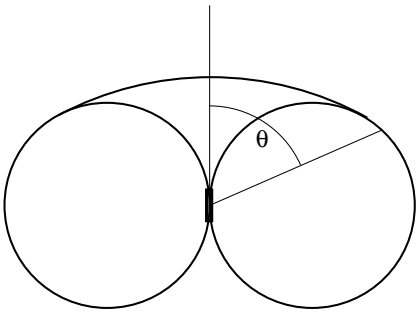

Figure 26: Field pattern of Hertz dipole.

of $\boldsymbol{S}$ as $\langle \boldsymbol{S} \rangle$, then the absolute value of the time–averaged Poynting vector is expressed as:

$$| \langle \boldsymbol{S} \rangle | = \langle E_{r\theta} H_{r\phi} \rangle = \frac{Z}{2} \left(\frac{Il}{2\lambda} \right)^2 \frac{\sin^2 \theta}{r^2}, \tag{106}$$

since, in general,

$$\langle \sin^2(kr - \omega t) \rangle = \frac{1}{2}.$$

The $\sin^2 \theta$ dependence in equation (106) shows the effectiveness of transmission of the electromagnetic power to a certain direction θ, ϕ, which is called **power pattern** (Figure 27).

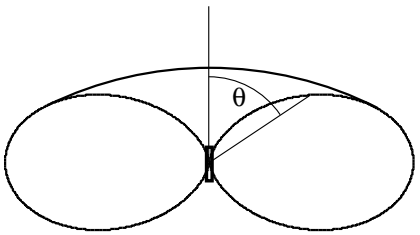

Figure 27: Power pattern of Hertz dipole.

Figure 28 shows time variation of the electromagnetic field in the near field of the oscillating Hertz dipole, which can be calculated using exact equations (73), (76) and (77). This figure is copied from a webpage

http://didaktik.physik.uni-wuerzburg.de/ pkrahmer/home/dipol.html.

One can see an animation movie of the transmission of the electromagnetic wave from the Hertz dipole in this webpage.

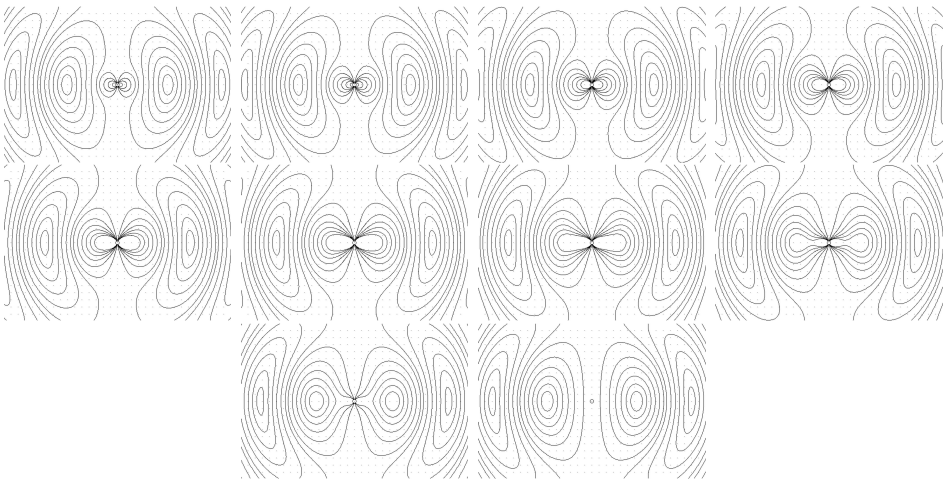

Figure 28: Time variation of the electromagnetic field close to the Hertz dipole (from http://didaktik.physik.uni-wuerzburg.de/ pkrahmer/home/dipol.html).

2.3.10 Linear Dipole Antenna of Finite Length

The Hertz dipole was still an idealized antenna. We now consider a realistic antenna, a linear dipole of finite length, with an oscillating current fed from the center of the linear dipole (Figure 29).

Let us assume that a linear dipole of length L is directed towards $\boldsymbol{i}_3$ axis and its center is located at the origin. Let us also assume that the spatial distribution of the harmonically oscillating electric current ($\propto \exp(-i\omega t)$) is given by an empirical formula:

$$I(z) = I_0 \sin\left[k\left(\frac{L}{2} - \mid z \mid\right)\right], \tag{107}$$

where z is distance from the origin along $\boldsymbol{i}_3$ axis, and $k = 2\pi/\lambda$.

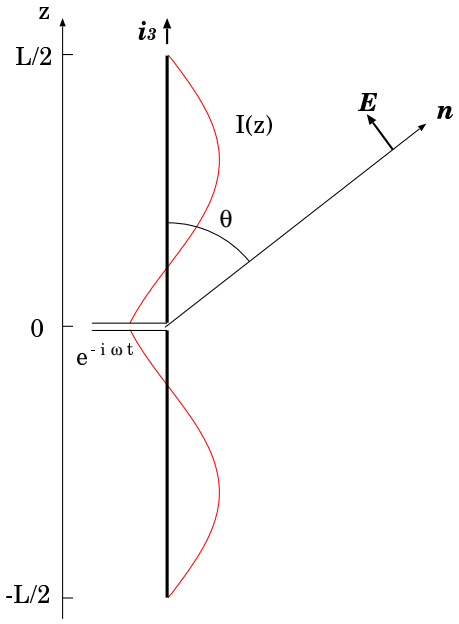

Figure 29: A linear dipole of finite length.

Then, in the far field, we have

$$\begin{aligned}
\boldsymbol{T}(\boldsymbol{n}) &= \int \boldsymbol{J}(\boldsymbol{r}')e^{-ik\boldsymbol{n}\cdot\boldsymbol{r}'}\,dV' = \boldsymbol{i}_3\int_{-L/2}^{L/2} I(z)e^{-ikz\cos\theta}\,dz \\
&= I_0\,\boldsymbol{i}_3\int_{-L/2}^{L/2}\sin\left[k\left(\frac{L}{2}-\mid z\mid\right)\right]e^{-ikz\cos\theta}\,dz \\
&= 2\,I_0\,\boldsymbol{i}_3\int_0^{L/2}\sin\left(\frac{kL}{2}-kz\right)\cos(kz\cos\theta)\,dz \\
&= \frac{2\,I_0}{k\sin^2\theta}\,\boldsymbol{i}_3\left[\cos\left(\frac{kL}{2}\cos\theta\right)-\cos\left(\frac{kL}{2}\right)\right], \qquad (108)
\end{aligned}$$

where θ is an angle of the direction of propagation $\boldsymbol{n}$ from the $\boldsymbol{i}_3$ axis. Therefore, equation (94) gives a non–zero component of the far field solution of the electric field:

$$E_\theta(\boldsymbol{r},\,t) = -i\,\frac{Z\,I_0}{2\pi}\,\frac{e^{i(kr-\omega t)}}{r}\,\frac{\cos\left(\frac{kL}{2}\cos\theta\right)-\cos\left(\frac{kL}{2}\right)}{\sin\theta}. \qquad (109)$$

This result shows that the field pattern of the linear dipole antenna is different with different ratio L/λ between the dipole length and wavelength. Figure

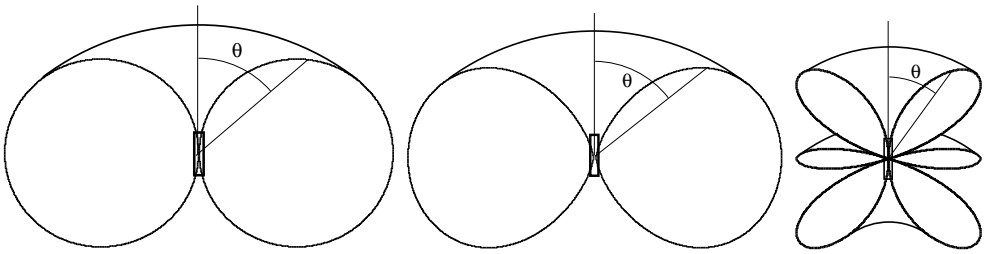

Figure 30: Field patterns of linear dipole antennas with lengths $L = 0.5\lambda$ (left), $L = 1.0\lambda$ (middle), and $L = 1.5\lambda$ (right) in the far field region.

30 shows field patterns of linear dipole antennas with different L/λ ratios in the far field region, calculated on the basis of equation (109).

Thus, we succeeded to transmit the electromagnetic wave from a realistic antenna!

Fifure 31 shows time variation of the near field around the linear dipole (from Kraus, Electromagnetics, Third Edition, Mc Graw-Hill, 1984).

2.4 Transmitting and Receiving Antennas

2.4.1 The Reciprocity Theorem

We derived field patterns of the simplest dipole antennas as dependences of transmitted electromagnetic fields on different directions, which characterize transmission efficiencies towards those directions. It is well known that the same antennas can be used both for transmission and reception of the electromagnetic waves. For receptions, we also call "field patterns" the reception efficiencies from different directions. Then, how the transmission and reception field patterns of the same antenna are related to each other? In order to answer to this question, we will consider Rayleigh's Reciprocity Theorem (1894) as applied to the electromagetics.

The reciprocity theorem states the followings.

> Let the electromagnetic fields, generated by two source currents $\boldsymbol{J}_1$ and $\boldsymbol{J}_2$, be $\boldsymbol{E}_1$, $\boldsymbol{H}_1$ and $\boldsymbol{E}_2$, $\boldsymbol{H}_2$, respectively. If the source currents exist within two finite spatial regions V_1 and V_2 only (Figure 32), and if a volume V includes both the two regions, then we have
>
> $$\int_V (\boldsymbol{E}_1 \cdot \boldsymbol{J}_2 - \boldsymbol{E}_2 \cdot \boldsymbol{J}_1)\, dV = 0. \qquad (110)$$

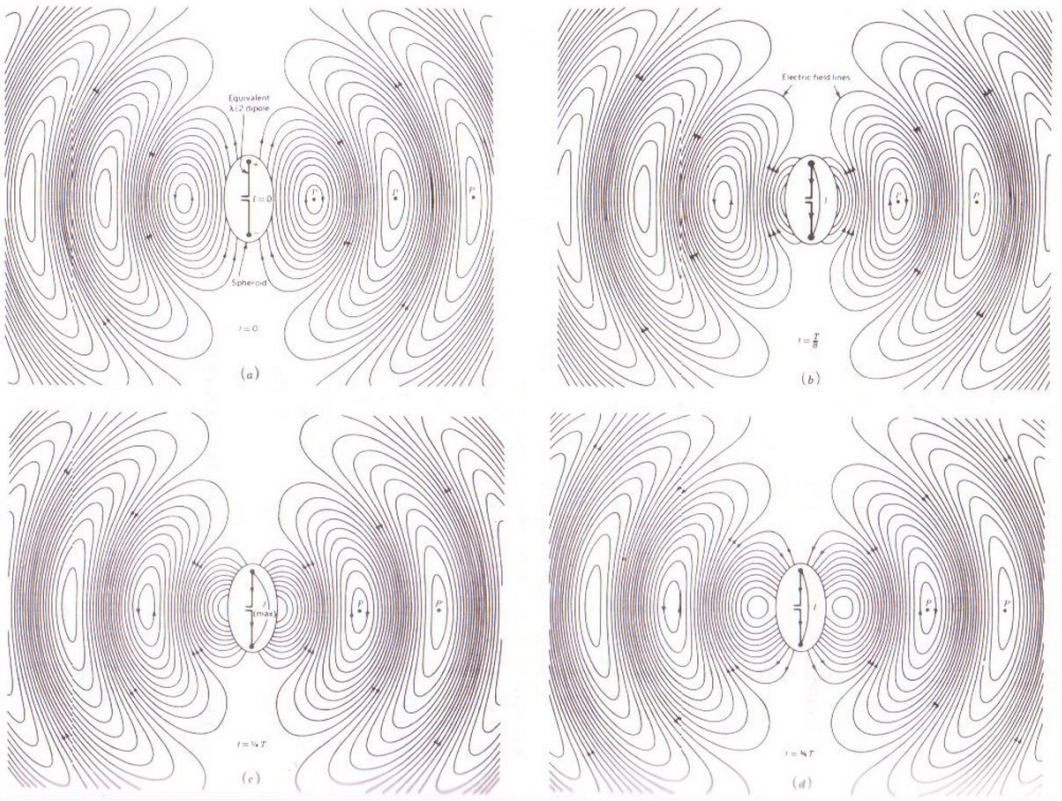

Figure 31: Time variation of the electric field lines in the near field region of a $L = \lambda/2$ linear dipole antenna (from Kraus, 1984).

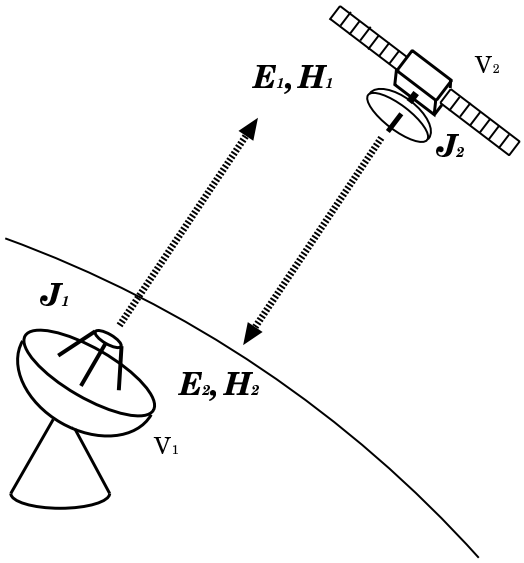

Figure 32: Electromagnetic fields generated by two source currents confined within two finite spatial regions.

Let us prove this theorem in a case of homogeneous medium (ϵ, μ = const), and harmonically oscillating currents ($\propto \exp(-i\omega t)$). In this case, from Maxwell equations (21) and (22), we have:

$$\nabla \times \boldsymbol{H}_1 = \boldsymbol{J}_1 - i\omega\epsilon \boldsymbol{E}_1 \qquad \nabla \times \boldsymbol{H}_2 = \boldsymbol{J}_2 - i\omega\epsilon \boldsymbol{E}_2 \tag{111}$$

$$\nabla \times \boldsymbol{E}_1 = i\omega\mu \boldsymbol{H}_1 \qquad \nabla \times \boldsymbol{E}_2 = i\omega\mu \boldsymbol{H}_2. \tag{112}$$

Using a formula of vector algebra in equation (17):

$$\nabla \cdot (\boldsymbol{A} \times \boldsymbol{B}) = \boldsymbol{B} \cdot (\nabla \times \boldsymbol{A}) - \boldsymbol{A} \cdot (\nabla \times \boldsymbol{B}),$$

and equations (111) and (112), we have

$$\begin{aligned} \nabla \cdot (\boldsymbol{E}_1 \times \boldsymbol{H}_2) &= \boldsymbol{H}_2 \cdot (\nabla \times \boldsymbol{E}_1) - \boldsymbol{E}_1 \cdot (\nabla \times \boldsymbol{H}_2) \\ &= i\omega\mu \boldsymbol{H}_1 \cdot \boldsymbol{H}_2 - \boldsymbol{E}_1 \cdot \boldsymbol{J}_2 + i\omega\epsilon \boldsymbol{E}_1 \cdot \boldsymbol{E}_2, \\ \nabla \cdot (\boldsymbol{E}_2 \times \boldsymbol{H}_1) &= \boldsymbol{H}_1 \cdot (\nabla \times \boldsymbol{E}_2) - \boldsymbol{E}_2 \cdot (\nabla \times \boldsymbol{H}_1) \\ &= i\omega\mu \boldsymbol{H}_1 \cdot \boldsymbol{H}_2 - \boldsymbol{E}_2 \cdot \boldsymbol{J}_1 + i\omega\epsilon \boldsymbol{E}_1 \cdot \boldsymbol{E}_2. \end{aligned}$$

Taking difference of the above equations, we obtain

$$\nabla \cdot (\boldsymbol{E}_1 \times \boldsymbol{H}_2 - \boldsymbol{E}_2 \times \boldsymbol{H}_1) = \boldsymbol{E}_2 \cdot \boldsymbol{J}_1 - \boldsymbol{E}_1 \cdot \boldsymbol{J}_2.$$

Then, integrating the both sides of the above equation through the volume V, which includes source regions of both $\boldsymbol{J}_1$ and $\boldsymbol{J}_2$, and applying Gauss's integration theorem in equation (67) to the LHS, we have

$$\oint (\boldsymbol{E}_1 \times \boldsymbol{H}_2 - \boldsymbol{E}_2 \times \boldsymbol{H}_1) \cdot \boldsymbol{n}_n d\mathcal{S} = \int_V (\boldsymbol{E}_2 \cdot \boldsymbol{J}_1 - \boldsymbol{E}_1 \cdot \boldsymbol{J}_2)\, dV, \qquad (113)$$

where $\boldsymbol{n}_n d\mathcal{S}$ is a vector of the surface element on the volume V with a unit normal vector $\boldsymbol{n}_n$.

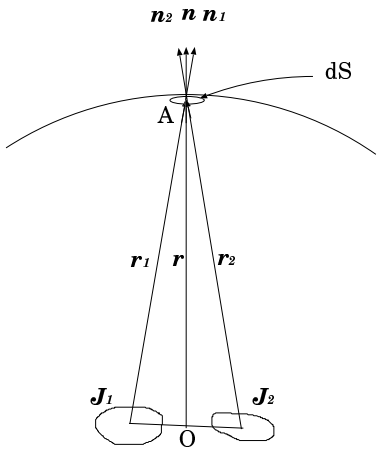

Figure 33: Geometry of the source current regions and radius vectors to a point on the surface of the volume V.

Let us denote radius vectors from centers of the two source current regions to a point on the surface (point A, say) of the volume V as $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$, and let us also denote a radius vector from the origin O, at the middle point of the two current regions, to the point A as $\boldsymbol{r}$, as shown in Figure 33.

If we adopt a sufficiently big volume V, so that its surface is in far field regions from the two current regions, the electromagnetic fields are approximated by the spherical waves, which satisfy

$$\boldsymbol{E}_1 = -Z\, \boldsymbol{n}_1 \times \boldsymbol{H}_1, \qquad \boldsymbol{E}_2 = -Z\, \boldsymbol{n}_2 \times \boldsymbol{H}_2, \qquad (114)$$

where $\boldsymbol{n}_1 = \boldsymbol{r}_1/r_1$ and $\boldsymbol{n}_2 = \boldsymbol{r}_2/r_2$.

Let us further assume, for simplicity, that the surface is a sphere with the center at the origin O. In that case, the unit normal vector $\boldsymbol{n}_n$ to the surface element at the point A is $\boldsymbol{n}_n = \boldsymbol{n} = \boldsymbol{r}/r$. Then, the surface integral in the

LHS of equation (113) becomes

$$\begin{aligned}
&\oint (\boldsymbol{E}_1 \times \boldsymbol{H}_2 - \boldsymbol{E}_2 \times \boldsymbol{H}_1) \cdot \boldsymbol{n}_n \, d\mathcal{S} \\
&\quad = -Z \oint [(\boldsymbol{n}_1 \times \boldsymbol{H}_1) \times \boldsymbol{H}_2 - (\boldsymbol{n}_2 \times \boldsymbol{H}_2) \times \boldsymbol{H}_1] \cdot \boldsymbol{n} \, d\mathcal{S} \\
&\quad = -Z \oint [(\boldsymbol{n} \cdot \boldsymbol{H}_1)(\boldsymbol{n}_1 \cdot \boldsymbol{H}_2) - (\boldsymbol{n} \cdot \boldsymbol{H}_2)(\boldsymbol{n}_2 \cdot \boldsymbol{H}_1) \\
&\qquad - \boldsymbol{n} \cdot (\boldsymbol{n}_1 - \boldsymbol{n}_2)(\boldsymbol{H}_1 \cdot \boldsymbol{H}_2)] \, d\mathcal{S} \\
&\quad = -Z \oint [(\boldsymbol{n} \cdot \boldsymbol{H}_1)(\boldsymbol{n}_1 - \boldsymbol{n}) \cdot \boldsymbol{H}_2 - (\boldsymbol{n} \cdot \boldsymbol{H}_2)(\boldsymbol{n}_2 - \boldsymbol{n}) \cdot \boldsymbol{H}_1 \\
&\qquad - \boldsymbol{n} \cdot (\boldsymbol{n}_1 - \boldsymbol{n}_2)(\boldsymbol{H}_1 \cdot \boldsymbol{H}_2)] \, d\mathcal{S}, \qquad (115)
\end{aligned}$$

where we used a formula of vector algebra

$$\boldsymbol{A} \times (\boldsymbol{B} \times \boldsymbol{C}) = \boldsymbol{B}(\boldsymbol{A} \cdot \boldsymbol{C}) - \boldsymbol{C}(\boldsymbol{A} \cdot \boldsymbol{B}),$$

given in equation (14).

Now, if we tend the radius r of the sphere to the infinity ($r \to \infty$), then

$d\mathcal{S} \propto r^2,$	since $d\mathcal{S} = r^2 \sin\theta \, d\theta \, d\phi,$
$\mid \boldsymbol{H}_1 \mid \mid \boldsymbol{H}_2 \mid \propto r^{-2},$	because they are spherical waves,
$\boldsymbol{n}_1 \to \boldsymbol{n}, \, \boldsymbol{n}_2 \to \boldsymbol{n},$	hence $\boldsymbol{n}_1 - \boldsymbol{n} \to 0, \, \boldsymbol{n}_2 - \boldsymbol{n} \to 0, \, \boldsymbol{n}_1 - \boldsymbol{n}_2 \to 0.$

Consequently, the LHS of equation (115), and therefore LHS of equation (113), too, tends to zero, when $r \to \infty$.

Thus, for a sufficiently large spherical volume V, equation (113) gives us

$$\int_V (\boldsymbol{E}_2 \cdot \boldsymbol{J}_1 - \boldsymbol{E}_1 \cdot \boldsymbol{J}_2) dV = 0. \qquad (116)$$

But, since we assume that the currents $\boldsymbol{J}_1 = 0$ and $\boldsymbol{J}_2 = 0$ outside of the finite regions V_1 and V_2, respectively, equation (116) must hold for an arbitrary volume V, including V_1 and V_2. Thus, we proved the reciprocity theorem.

Equation (116) can be expressed also in a form:

$$\int_{V_1} \boldsymbol{E}_2 \cdot \boldsymbol{J}_1 \, dV = \int_{V_2} \boldsymbol{E}_1 \cdot \boldsymbol{J}_2 \, dV. \qquad (117)$$

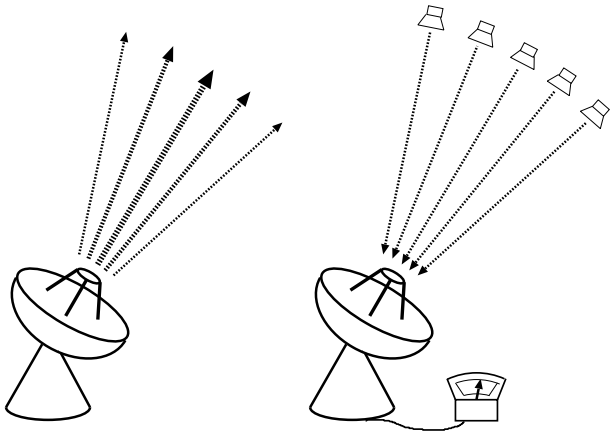

Figure 34: Transmission (left) and reception (right) field patterns of an antenna.

2.4.2 Equivalence of Field Patterns in Transmission and Reception

Suppose an antenna, which is transmitting the radio wave towards the sky. The **transmission field pattern** is defined as the direction dependence of the amplitude of the electromagnetic field transmitted from the antenna (see left panel of Figure 34). Now suppose the same antenna is receiving the radio wave from a hypothetical distant emitter of unit strength located in an arbitrary direction of the sky. The **reception field pattern** is defined as dependence of the amplitude of the received electromagnetic field, or the resultant voltage in a receiver system, on the direction of the emitter (see right panel of Figure 34).

The reciprocity theorem shows us that the transmission field pattern and the reception field pattern of the same antenna are identical.

In order to prove this statement, let us again consider that a current $\boldsymbol{J}_1$ in an antenna 1 generates an electric field $\boldsymbol{E}_1$ around another antenna 2, while a current $\boldsymbol{J}_2$ in the antenna 2 generates an electric field $\boldsymbol{E}_2$ around the antenna 1 (see Figure 32).

Let us assume, for simplicity, that the antennas are small dipoles with length l and cross–section q (Figure 35). This is not necessarily a very unrealistic assumption, since co–axial to waveguide converters in feedhorns of existing antennas are often composed of simple dipoles. Then, we can adopt in equation (117),

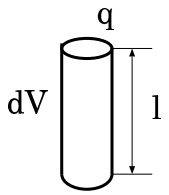

Figure 35: A small dipole with length l and cross-section q.

$$\begin{array}{lcl} \text{volume element} & : & dV = ql \\ \text{current} & : & I = \mid \boldsymbol{J} \mid q \\ \text{voltage} & : & U = El. \end{array}$$

And, therefore, equation (117) can now be expressed as:

$$U_2\, I_1 = U_1\, I_2. \tag{118}$$

Let us now consider that antenna 1 is fixed at some position and antenna 2 moves from position A to position B, and then to position C (Figure 36). Let us denote voltages, generated by the current I_1 in antenna 1 at positions A, B, and C of antenna 2, as U_{1A}, U_{1B}, and U_{1C}. We also denote voltages at antenna 1, generated by the current I_2 of antenna 2 at positions A, B, and C, as U_{2A}, U_{2B}, and U_{2C}. From the reciprocity theorem, as described in equation (118), we have

$$\begin{array}{rcl} U_{2A}\, I_1 & = & U_{1A}\, I_2, \\ U_{2B}\, I_1 & = & U_{1B}\, I_2, \\ U_{2C}\, I_1 & = & U_{1C}\, I_2. \end{array} \tag{119}$$

Therefore, U_1 is always proportional to U_2, irrespective of the direction of antenna 2 viewed from antenna 1. Since U_1 expresses the transmission field pattern, while U_2 expresses the reception pattern, of antenna 1, the proportionality $U_1 \propto U_2$ proves the equivalence of the transmission and reception field patterns of an antenna.

2.5 Transmission from Aperture Plane

2.5.1 Aperture Antennas

Let us consider a transmitting horn antenna, such as shown in Figure 37. Primary wave source (the oscillating current in the co–axial to waveguide

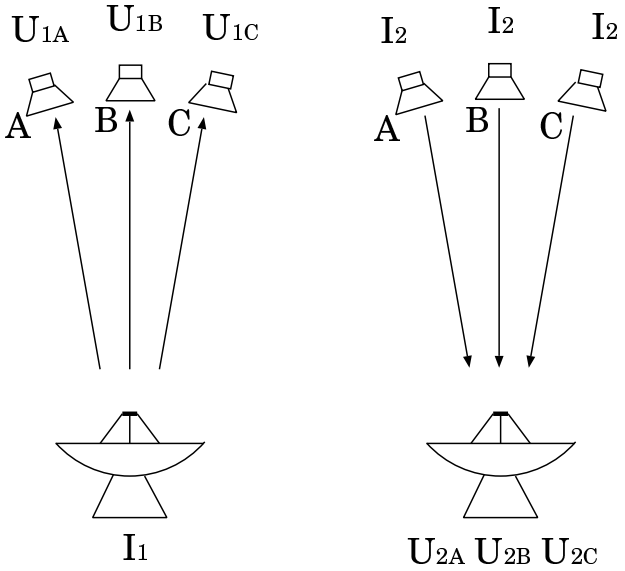

Figure 36: Fixed antenna 1 and moving antenna 2.

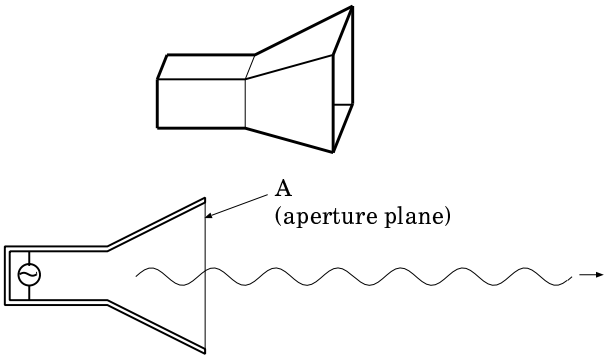

Figure 37: A horn antenna and an aperture plane.

converter) is located in the waveguide tube, and generates the electromagnetic fields in the horn. Then, the electromagnetic fields flow out of the horn towards the external space, through a plane A. Antennas, having such planes, are called "aperture antennas", and the planes, "aperture planes". Paraboloidal antennas are also typical aperture antennas (Figure 38).

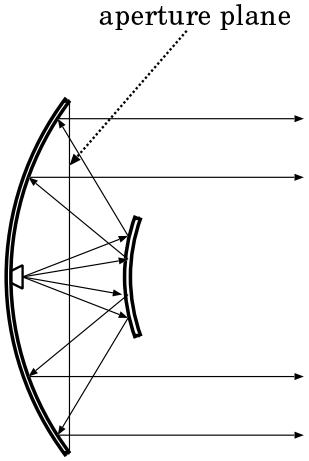

Figure 38: A paraboloidal antenna and an aperture plane.

In order to consider the field transmission pattern of an aperture antenna, it is convenient to regard, as if the electromagnetic waves are generated by some fictitious wave source located on the aperture plane (Figure 39). Such a "secondary wave source" can be uniquely related to the electromagnetic field generated by the primary wave source, the actual oscillating current within the feed horn, and can exactly reproduce the actual field transmission pattern, as we will see below. The secondary source on an aperture plane can be interpreted in terms of Huygens and Fresnel's principle, which states "each point on a primary wavefront can be considered to be a new source of a secondary wave".

2.5.2 Boundary Conditions on the Aperture Plane

If physical properties of a medium are discontinuous at some boundary surface, we solve physical equations on the both sides of the boundary, and then connect the solutions, using a set of boundary conditions. The boundary conditions for electromagnetic fields $\boldsymbol{E}_1$, $\boldsymbol{H}_1$, $\boldsymbol{D}_1$, $\boldsymbol{B}_1$ and $\boldsymbol{E}_2$, $\boldsymbol{H}_2$, $\boldsymbol{D}_2$, $\boldsymbol{B}_2$ in two regions with different ϵ_1, μ_1, σ_1 and ϵ_2, μ_2, σ_2, respectively, are well known: for a boundary surface, which separate the two regions, with a unit

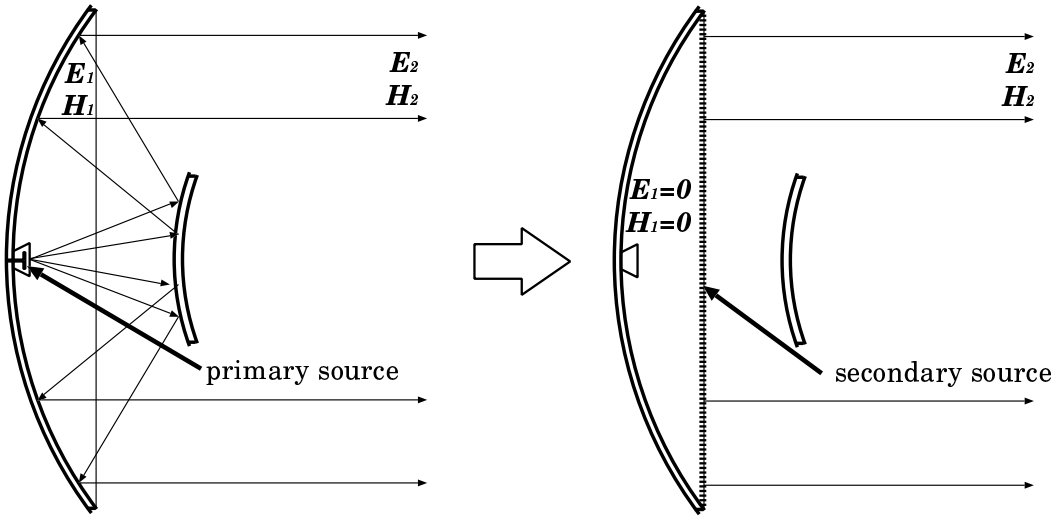

Figure 39: Electromagnetic wave, transmitted from an aperture antenna (left), can be represented as a wave generated by an appropriately chosen secondary wave source on the aperture plane (right).

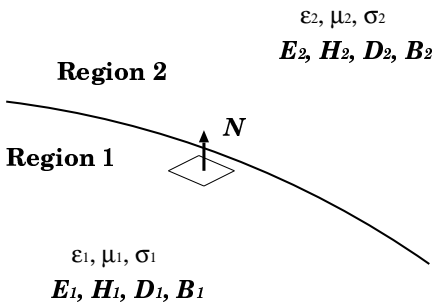

Figure 40: Regions with different electromagnetic properties.

normal vector $\boldsymbol{N}$ (Figure 40), we have

$$\begin{aligned} \boldsymbol{N} \times (\boldsymbol{H}_2 - \boldsymbol{H}_1) &= \boldsymbol{K}, & (120) \\ \boldsymbol{N} \times (\boldsymbol{E}_2 - \boldsymbol{E}_1) &= 0, & (121) \\ \boldsymbol{N} \cdot (\boldsymbol{D}_2 - \boldsymbol{D}_1) &= \Sigma, & (122) \\ \boldsymbol{N} \cdot (\boldsymbol{B}_2 - \boldsymbol{B}_1) &= 0, & (123) \end{aligned}$$

where $\boldsymbol{K}$ [A m^{-1}] is the surface density of electric current (or "surface electric current"), and Σ [A s m^{-2}] is the surface density of electric charge (or "surface electric charge").

However, having these usual boundary conditions alone is not sufficient to fulfil the requirements for the secondary source on the aperture plane, shown in Figure 39, where $\boldsymbol{E}_1 = 0$ and $\boldsymbol{H}_1 = 0$ in the inner side of the aperture plane, while $\boldsymbol{E}_2 \neq 0$ and $\boldsymbol{H}_2 \neq 0$ in the outer side of the aperture plane.

In order to overcome this difficulty, we introduce virtual quantities called "**magnetic current density**" $\boldsymbol{J}_m$, with unit [V m^{-2}], and "**magnetic charge density**" ρ_m, with unit [V s m^{-3}]. They are supposed to satisfy "extended" Maxwell equations:

$$\begin{aligned} \nabla \times \boldsymbol{E} &= -\boldsymbol{J}_m - \frac{\partial \boldsymbol{B}}{\partial t}, & (124) \\ \nabla \cdot \boldsymbol{B} &= \rho_m, & (125) \end{aligned}$$

and to fulfil following boundary conditions:

$$\begin{aligned} \boldsymbol{N} \times (\boldsymbol{H}_2 - \boldsymbol{H}_1) &= 0, & (126) \\ \boldsymbol{N} \times (\boldsymbol{E}_2 - \boldsymbol{E}_1) &= -\boldsymbol{K}_m, & (127) \\ \boldsymbol{N} \cdot (\boldsymbol{D}_2 - \boldsymbol{D}_1) &= 0, & (128) \\ \boldsymbol{N} \cdot (\boldsymbol{B}_2 - \boldsymbol{B}_1) &= \Sigma_m, & (129) \end{aligned}$$

where $\boldsymbol{K}_m$ [V m^{-1}] is the surface density of magnetic current (or "surface magnetic current"), and Σ_m [V s m^{-2}] is the surface density of magnetic charge (or "surface magnetic charge").

Note that an equation of continuity:

$$\frac{\partial \rho_m}{\partial t} + \nabla \cdot \boldsymbol{J}_m = 0, \qquad (130)$$

is derived from equations (124) and (125).

Before discussing the boundary conditions at the aperture plane, let us first consider a closed surface surrounding a primary wave source (Figure 41). If we assume that the medium is continuous across the surface, the

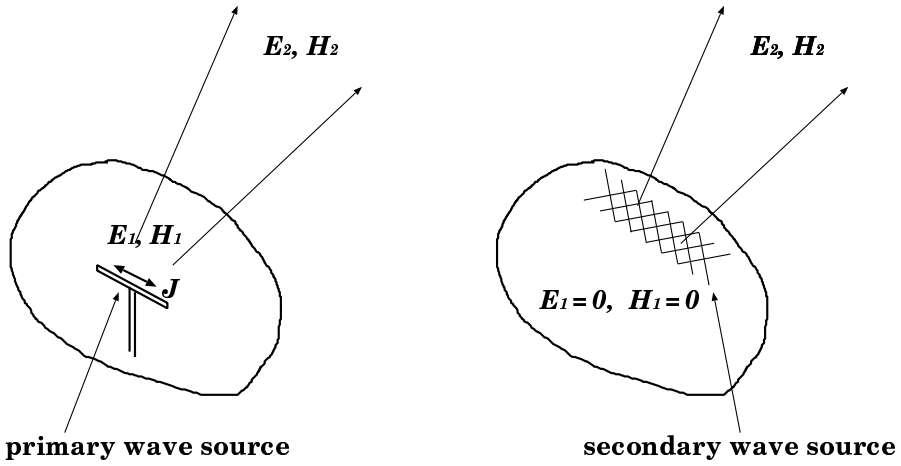

Figure 41: A closed surface surrounding a primary wave source (left), which can be regarded as a secondary wave source with suitable surface currents and surface charges (right).

electromagnetic fields must be also continuous, i.e. conditions $\boldsymbol{E}_1 = \boldsymbol{E}_2$, $\boldsymbol{H}_1 = \boldsymbol{H}_2$, $\boldsymbol{D}_1 = \boldsymbol{D}_2$ and $\boldsymbol{B}_1 = \boldsymbol{B}_2$ must hold on the surface.

Now, suppose we place surface electric current $\boldsymbol{K}$, surface magnetic current $\boldsymbol{K}_m$, surface electric charge Σ and surface magnetic charge Σ_m on the surface, which satisfy

$$\boldsymbol{K} = \boldsymbol{N} \times \boldsymbol{H}_1, \tag{131}$$

$$\boldsymbol{K}_m = -\boldsymbol{N} \times \boldsymbol{E}_1, \tag{132}$$

$$\Sigma = \boldsymbol{N} \cdot \boldsymbol{D}_1, \tag{133}$$

$$\Sigma_m = \boldsymbol{N} \cdot \boldsymbol{B}_1, \tag{134}$$

where $\boldsymbol{N}$ is a unit normal vector, perpendicular to the surface. And then, we remove the primary wave source, and set the electromagnetic fields inside of the closed surface to zero, i.e. $\boldsymbol{E}_1 = 0$, $\boldsymbol{H}_1 = 0$, $\boldsymbol{D}_1 = 0$ and $\boldsymbol{B}_1 = 0$. Of course, the electromagnetic fields become discontinous at the surface, in this case.

Let us compare boundary conditions at the surface in two cases, one with the real continuous fields, and another with the fictitious discontinuous fields, together with the surface currents and the surface charges on the boundary. They are

Continuous fields case	*Discontinuous fields case*
$\boldsymbol{N} \times \boldsymbol{H}_2 = \boldsymbol{N} \times \boldsymbol{H}_1,$	$\boldsymbol{N} \times \boldsymbol{H}_2 = \boldsymbol{K},$
$\boldsymbol{N} \times \boldsymbol{E}_2 = \boldsymbol{N} \times \boldsymbol{E}_1,$	$\boldsymbol{N} \times \boldsymbol{E}_2 = -\boldsymbol{K}_m,$
$\boldsymbol{N} \cdot \boldsymbol{D}_2 = \boldsymbol{N} \cdot \boldsymbol{D}_1,$	$\boldsymbol{N} \cdot \boldsymbol{D}_2 = \Sigma,$
$\boldsymbol{N} \cdot \boldsymbol{B}_2 = \boldsymbol{N} \cdot \boldsymbol{B}_1,$	$\boldsymbol{N} \cdot \boldsymbol{B}_2 = \Sigma_m.$ (135)

As long as we consider these relations as boundary conditions at the closed surface for external solutions of the electromagnetic fields, the left hand side conditions and the right hand side conditions are the same, in view of the eauations from (131) to (134). Therefore, the external fields must be identical for these two sets of boundary conditions. This implies that we can calculate the external fields, provided that the appropriate surface currents and surface charges, including both electric and magnetic, are given on the closed boundary surface.

The same statement must be valid for the aperture plane, if the aperture plane is the only part of a closed surface surrounding the primary wave source through which the electromagnetic fields flow out.

Of course, the magnetic current and the magnetic charge do not exist in the real world. We use them as auxiliary concepts, which help us to fulfil the requirements at the aperture plane, but disappear in our final results, as we will see later.

2.5.3 Wave Equations with Magnetic Current and Magnetic Charge

How can we calculate external electromagnetic fields when the surface currents and the surface charges are given at a surface?

In order to answer to this question, we will first consider the wave generations from the electric current & charge system, which we have seen before, and from the magnetic current & charge system, which are new for us, in parallel.

Electric Current & Charge System	*Magnetic Current & Charge System*
Maxwell equations	
$\nabla \times \boldsymbol{E} = -\dfrac{\partial \boldsymbol{B}}{\partial t},$	$\nabla \times \boldsymbol{E} = -\boldsymbol{J}_m - \dfrac{\partial \boldsymbol{B}}{\partial t}$
$\nabla \times \boldsymbol{H} = \boldsymbol{J} + \dfrac{\partial \boldsymbol{D}}{\partial t},$	$\nabla \times \boldsymbol{H} = \dfrac{\partial \boldsymbol{D}}{\partial t},$
$\nabla \cdot \boldsymbol{D} = \rho,$	$\nabla \cdot \boldsymbol{D} = 0,$
$\nabla \cdot \boldsymbol{B} = 0,$	$\nabla \cdot \boldsymbol{B} = \rho_m.$

Equation of continuity

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \boldsymbol{J} = 0, \qquad \frac{\partial \rho_m}{\partial t} + \nabla \cdot \boldsymbol{J}_m = 0.$$

Electromagnetic potentials

$$\boldsymbol{B} = \nabla \times \boldsymbol{A}, \qquad \boldsymbol{D} = -\nabla \times \boldsymbol{A}_m,$$
$$\boldsymbol{E} = -\nabla\Phi - \frac{\partial \boldsymbol{A}}{\partial t}, \qquad \boldsymbol{H} = -\nabla\Phi_m - \frac{\partial \boldsymbol{A}_m}{\partial t}.$$

Lorentz gauge

$$\nabla \cdot \boldsymbol{A} + \epsilon\,\mu \frac{\partial \Phi}{\partial t} = 0, \qquad \nabla \cdot \boldsymbol{A}_m + \epsilon\,\mu \frac{\partial \Phi_m}{\partial t} = 0.$$

Wave equations with source terms

$$\nabla^2 \boldsymbol{A} - \frac{1}{c^2}\frac{\partial^2 \boldsymbol{A}}{\partial t^2} = -\mu\,\boldsymbol{J}, \qquad \nabla^2 \boldsymbol{A}_m - \frac{1}{c^2}\frac{\partial^2 \boldsymbol{A}_m}{\partial t^2} = -\epsilon\,\boldsymbol{J}_m,$$
$$\nabla^2 \Phi - \frac{1}{c^2}\frac{\partial^2 \Phi}{\partial t^2} = -\frac{1}{\epsilon}\,\rho, \qquad \nabla^2 \Phi_m - \frac{1}{c^2}\frac{\partial^2 \Phi_m}{\partial t^2} = -\frac{1}{\mu}\,\rho.$$

Retarded potentials with harmonically oscillating sources

$$\boldsymbol{A}(\boldsymbol{r},\,t) = \boldsymbol{A}(\boldsymbol{r})e^{-i\omega t}, \quad \boldsymbol{A}_m(\boldsymbol{r},\,t) = \boldsymbol{A}_m(\boldsymbol{r})e^{-i\omega t},$$
$$\boldsymbol{E}(\boldsymbol{r},\,t) = \boldsymbol{E}(\boldsymbol{r})e^{-i\omega t}, \quad \boldsymbol{H}(\boldsymbol{r},\,t) = \boldsymbol{H}(\boldsymbol{r})e^{-i\omega t}.$$
($\boldsymbol{J} = 0$, $\boldsymbol{J}_m = 0$, $\rho = 0$, and $\rho_m = 0$, outside of the source region)

$$\boldsymbol{A}(\boldsymbol{r}) = \frac{\mu}{4\pi}\int \boldsymbol{J}(\boldsymbol{r}')\frac{e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}}{|\,\boldsymbol{r}-\boldsymbol{r}'\,|}\,dV', \quad \boldsymbol{A}_m(\boldsymbol{r}) = \frac{\epsilon}{4\pi}\int \boldsymbol{J}_m(\boldsymbol{r}')\frac{e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}}{|\,\boldsymbol{r}-\boldsymbol{r}'\,|}\,dV',$$
$$\boldsymbol{H}(\boldsymbol{r}) = \frac{1}{\mu}\nabla \times \boldsymbol{A}(\boldsymbol{r}), \qquad \boldsymbol{E}(\boldsymbol{r}) = -\frac{1}{\epsilon}\nabla \times \boldsymbol{A}_m(\boldsymbol{r}),$$
$$\boldsymbol{E}(\boldsymbol{r}) = i\frac{c}{k}\,\nabla \times (\nabla \times \boldsymbol{A}(\boldsymbol{r})), \qquad \boldsymbol{H}(\boldsymbol{r}) = i\frac{c}{k}\,\nabla \times (\nabla \times \boldsymbol{A}_m(\boldsymbol{r})).$$

Far field solutions in Fraunhofer region

$$\boldsymbol{T}(\boldsymbol{n}) = \int \boldsymbol{J}(\boldsymbol{r}')e^{-ik\boldsymbol{n}\cdot\boldsymbol{r}'}\,dV', \qquad \boldsymbol{T}_m(\boldsymbol{n}) = \int \boldsymbol{J}_m(\boldsymbol{r}')e^{-ik\boldsymbol{n}\cdot\boldsymbol{r}'}\,dV',$$
$$\boldsymbol{A}(\boldsymbol{r}) = \frac{\mu}{4\pi}\,\frac{e^{ikr}}{r}\,\boldsymbol{T}(\boldsymbol{n}), \qquad \boldsymbol{A}_m(\boldsymbol{r}) = \frac{\mu}{4\pi}\,\frac{e^{ikr}}{r}\,\boldsymbol{T}_m(\boldsymbol{n}),$$
$$\boldsymbol{H}(\boldsymbol{r}) = i\,\frac{1}{2\lambda}\,\frac{e^{ikr}}{r}\,\boldsymbol{n} \times \boldsymbol{T}(\boldsymbol{n}), \qquad \boldsymbol{E}(\boldsymbol{r}) = -i\,\frac{1}{2\lambda}\frac{e^{ikr}}{r}\,\boldsymbol{n} \times \boldsymbol{T}_m(\boldsymbol{n}).$$
$$\boldsymbol{E}(\boldsymbol{r}) = -i\,\frac{Z}{2\lambda}\frac{e^{ikr}}{r}\,\boldsymbol{n} \times [\boldsymbol{n} \times \boldsymbol{T}(\boldsymbol{n})], \qquad \boldsymbol{H}(\boldsymbol{r}) = -i\,\frac{1}{2\lambda Z}\,\frac{e^{ikr}}{r}\,\boldsymbol{n} \times [\boldsymbol{n} \times \boldsymbol{T}_m(\boldsymbol{n})].$$

Since the vector and scalar potentials are related to each other by Lorentz gauge relations, we considered here wave generations for vector potentials $\boldsymbol{A}(\boldsymbol{r},\, t)$ and $\boldsymbol{A}_m(\boldsymbol{r},\, t)$ only, as we did before.

In a case when both the electric current and the magnetic current co–exist, the electromagnetic fields are expressed by the superposition of two sets of solutions shown above. In the far field region, for example, we have

$$\boldsymbol{H}(\boldsymbol{r}) = i\,\frac{1}{2\lambda}\frac{e^{ikr}}{r}\left\{\boldsymbol{n}\times\boldsymbol{T}(\boldsymbol{n}) - \frac{1}{Z}\,\boldsymbol{n}\times[\boldsymbol{n}\times\boldsymbol{T}_m(\boldsymbol{n})]\right\}, \tag{136}$$

$$\boldsymbol{E}(\boldsymbol{r}) = -i\,\frac{Z}{2\lambda}\frac{e^{ikr}}{r}\left\{\boldsymbol{n}\times[\boldsymbol{n}\times\boldsymbol{T}(\boldsymbol{n})] + \frac{1}{Z}\,\boldsymbol{n}\times\boldsymbol{T}_m(\boldsymbol{n})\right\}. \tag{137}$$

For these spherical waves in the far field region, we again obtain:

$$\boldsymbol{n}\cdot\boldsymbol{H} = 0, \quad \text{and} \quad \boldsymbol{n}\cdot\boldsymbol{E} = 0, \quad \text{(transversal wave)},$$

and

$$\boldsymbol{E} = -Z\,\boldsymbol{n}\times\boldsymbol{H}, \quad \text{and} \quad \boldsymbol{H} = \frac{1}{Z}\,\boldsymbol{n}\times\boldsymbol{E}, \quad (\boldsymbol{E},\, \boldsymbol{H} \text{ orthogonal}).$$

2.5.4 Radio Wave Transmission from a Surface

Now, how to describe the generation of electromagnetic waves from the surface electronic current $\boldsymbol{K}$ and the surface magnetic current $\boldsymbol{K}_m$ on a surface of the secondary wave source?

For this purpose, let us consider "a volume density of a surface density" on a boundary surface. Let a unit normal vector, perpendicular to the boundary

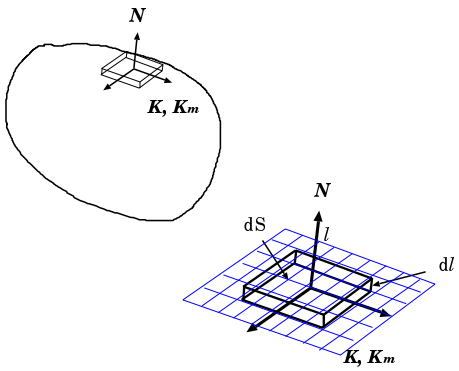

Figure 42: A volume density of a surface density.

sourface, be $\boldsymbol{N}$. Let us consider a local rectangular coordinate system at a

certain point on the surface, which are composed of an axis parallel to $\boldsymbol{N}$ and other two axes in a plane tangential to the boundary surface. If we denote the length along $\boldsymbol{N}$ direction as l, and an infinitesimal surface area element on the boundary as $d\mathcal{S}$, then we can express an infinitesimal volume element at the boundary as

$$dV = dl\, d\mathcal{S}.$$

Then, we can express the "volume densities" $\boldsymbol{J}$ and $\boldsymbol{J}_m$ of the surface densities $\boldsymbol{K}$ and $\boldsymbol{K}_m$ as

$$\begin{aligned} \boldsymbol{J} &= \boldsymbol{K}\,\delta(l - l_0), \\ \boldsymbol{J}_m &= \boldsymbol{K}_m\,\delta(l - l_0), \end{aligned} \tag{138}$$

where $l = l_0$ corresponds to the boundary surface, and $\delta(x)$ is the delta function. Indeed, if we integrate the volume densities through a small volume $\Delta V = \Delta l\, \Delta\mathcal{S}$, containing a part of the boundary surface $l = l_0$, we have

$$\begin{aligned} \int\limits_{\Delta V} \boldsymbol{J}\, dV &= \int\limits_{\Delta V} \boldsymbol{K}\delta(l - l_0)\, dl\, d\mathcal{S} = \int\limits_{\Delta\mathcal{S}} \boldsymbol{K}\, d\mathcal{S}, \\ \int\limits_{\Delta V} \boldsymbol{J}_m\, dV &= \int\limits_{\Delta V} \boldsymbol{K}_m\delta(l - l_0)\, dl\, d\mathcal{S} = \int\limits_{\Delta\mathcal{S}} \boldsymbol{K}_m\, d\mathcal{S}, \end{aligned} \tag{139}$$

as expected.

Therefore, using equations (138), we can express the retarded vector potentials as

$$\begin{aligned} \boldsymbol{A}(\boldsymbol{r},\, t) &= \frac{\mu}{4\pi}\, e^{-i\omega t} \int \boldsymbol{J}(\boldsymbol{r}')\, \frac{e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}}{|\,\boldsymbol{r} - \boldsymbol{r}'\,|}\, dV' = \frac{\mu}{4\pi}\, e^{-i\omega t} \oint \boldsymbol{K}(\boldsymbol{r}')\, \frac{e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}}{|\,\boldsymbol{r} - \boldsymbol{r}'\,|}\, d\mathcal{S}', \\ \boldsymbol{A}_m(\boldsymbol{r},\, t) &= \frac{\mu}{4\pi}\, e^{-i\omega t} \int \boldsymbol{J}_m(\boldsymbol{r}')\, \frac{e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}}{|\,\boldsymbol{r} - \boldsymbol{r}'\,|}\, dV' = \frac{\mu}{4\pi}\, e^{-i\omega t} \oint \boldsymbol{K}_m(\boldsymbol{r}')\, \frac{e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}}{|\,\boldsymbol{r} - \boldsymbol{r}'\,|}\, d\mathcal{S}'. \end{aligned} \tag{140}$$

For the far field solutions, we have

$$\begin{aligned} \boldsymbol{T}(\boldsymbol{n}) &= \oint \boldsymbol{K}(\boldsymbol{r}')e^{-ik\boldsymbol{n}\cdot\boldsymbol{r}'}\, d\mathcal{S}', \\ \boldsymbol{T}_m(\boldsymbol{n}) &= \oint \boldsymbol{K}_m(\boldsymbol{r}')e^{-ik\boldsymbol{n}\cdot\boldsymbol{r}'}\, d\mathcal{S}', \end{aligned} \tag{141}$$

$$\begin{aligned} \boldsymbol{A}(\boldsymbol{r},\, t) &= \frac{\mu}{4\pi}\, \frac{e^{i(kr-\omega t)}}{r}\, \boldsymbol{T}(\boldsymbol{n}), \\ \boldsymbol{A}_m(\boldsymbol{r},\, t) &= \frac{\mu}{4\pi}\, \frac{e^{i(kr-\omega t)}}{r}\, \boldsymbol{T}_m(\boldsymbol{n}), \end{aligned} \tag{142}$$

and

$$\boldsymbol{H}(\boldsymbol{r},\,t) = i\,\frac{1}{2\lambda}\frac{e^{i(kr-\omega t)}}{r}\left\{\boldsymbol{n}\times\boldsymbol{T}(\boldsymbol{n}) - \frac{1}{Z}\,\boldsymbol{n}\times[\boldsymbol{n}\times\boldsymbol{T}_m(\boldsymbol{n})]\right\},$$

$$\boldsymbol{E}(\boldsymbol{r},\,t) = -i\,\frac{Z}{2\lambda}\frac{e^{i(kr-\omega t)}}{r}\left\{\boldsymbol{n}\times[\boldsymbol{n}\times\boldsymbol{T}(\boldsymbol{n})] + \frac{1}{Z}\,\boldsymbol{n}\times\boldsymbol{T}_m(\boldsymbol{n})\right\}, \qquad (143)$$

where $\boldsymbol{n}$ is a unit vector along the direction of the propagation of the spherical wave, as before.

Comparing these equations with equations (136) and (137), we see that their forms are the same, and only difference is in the replacement of the volume integrals in the expressions of $\boldsymbol{T}(\boldsymbol{n})$ and $\boldsymbol{T}_m(\boldsymbol{n})$ by the surface integralas in equations (141).

Remembering here that the surface electric current and the surface magnetic current must equal to

$$\boldsymbol{K} = \boldsymbol{N}\times\boldsymbol{H}_1,$$
$$\boldsymbol{K}_m = -\boldsymbol{N}\times\boldsymbol{E}_1,$$

on the boundary surface with a unit normal vector $\boldsymbol{N}$, in view of equations (131) and (132), we can rewrite $\boldsymbol{T}(\boldsymbol{n})$ and $\boldsymbol{T}_m(\boldsymbol{n})$ in equations (141) as

$$\boldsymbol{T}(\boldsymbol{n}) = \oint \boldsymbol{K}(\boldsymbol{r}')e^{-ik\boldsymbol{n}\cdot\boldsymbol{r}'}\,d\mathcal{S}' = \oint \boldsymbol{N}\times\boldsymbol{H}_1(\boldsymbol{r}')e^{-ik\boldsymbol{n}\cdot\boldsymbol{r}'}\,d\mathcal{S}',$$
$$\boldsymbol{T}_m(\boldsymbol{n}) = \oint \boldsymbol{K}_m(\boldsymbol{r}')e^{-ik\boldsymbol{n}\cdot\boldsymbol{r}'}\,d\mathcal{S}' = -\oint \boldsymbol{N}\times\boldsymbol{E}_1(\boldsymbol{r}')e^{-ik\boldsymbol{n}\cdot\boldsymbol{r}'}\,d\mathcal{S}'. \qquad (144)$$

Now in this last expression, in the right hand side of equation (144), we no longer have the fictituous magnetic current, and the external electromagnetic fields in equations (143) are completely determined by the actual electromagnetic fields on the boundary surface. This is a clear expression of Huygens and Fresnel's principle that the wave front becomes the wave source.

2.5.5 Radio Wave Transmission from an Aperture Antenna

Let us consider an aperture antenna, with an aperture plane much larger in size than the wavelength. It is then possible to control the primary wave source in such a way that harmonically oscillating electromagnetic fields on the aperture plane, $\boldsymbol{H}_a$ and $\boldsymbol{E}_a$, satisfy conditions:

$$\boldsymbol{N}\cdot\boldsymbol{H}_a = 0, \quad \boldsymbol{N}\cdot\boldsymbol{E}_a = 0, \quad \text{(parallel to the aperture plane)},$$
and
$$\boldsymbol{H}_a = \frac{1}{Z}\,\boldsymbol{N}\times\boldsymbol{E}_a, \quad \text{(mutually orthogonal)}, \qquad (145)$$

where $\boldsymbol{N}$ is the unit normal vector, perpendicular to the aperture plane (Figure 43). This corresponds to the case when the electromagnetic fields on

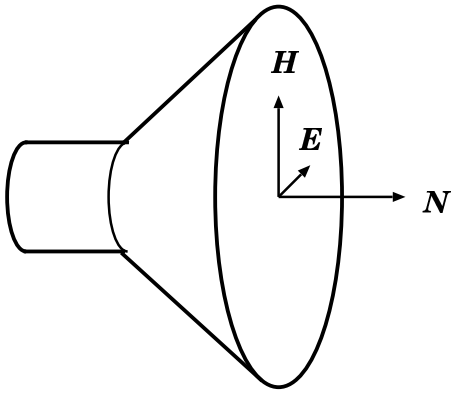

Figure 43: Electromagnetic field on an aperture plane.

the aperture plane are close to plane waves propagating towards the direction perpendicular to the aperture plane.

We now apply these relations to calculate directional pattern vectors $\boldsymbol{T}(\boldsymbol{n})$ and $\boldsymbol{T}_m(\boldsymbol{n})$ in equations (144). Generally speaking, the surface integrals in equations (144) must be taken over a closed surface, surrounding the primary wave source. However, as we saw earlier, the closed surface can be replaced by the aperture plane $\mathcal{A}$, if the aperture plane is the only part of a closed surface through which the electromagnetic fields flow out. This condition is usually fulfilled in aperture antennas with aperture sizes much larger than the wavelengths (Figure 44).

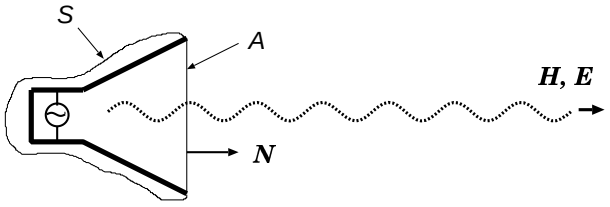

Figure 44: In an aperture antenna with aperture size much larger than the wavelength, most of the generated electromagnetic fields flow out through the aperture plane $\mathcal{A}$. In this respect, the contribution of other portion $\mathcal{S}$ of the closed surface, surrounding the antenna, is negligibly small compared with that of the aperture plane $\mathcal{A}$.

Therefore, we can rewrite equations (144) as

$$\boldsymbol{T}(\boldsymbol{n}) = \int_{\mathcal{A}} \boldsymbol{N} \times \boldsymbol{H}_a \, e^{-ik\boldsymbol{n}\cdot\boldsymbol{r}'} \, d\mathcal{S}' = \frac{1}{Z} \int_{\mathcal{A}} \boldsymbol{N} \times (\boldsymbol{N} \times \boldsymbol{E}_a) \, e^{-ik\boldsymbol{n}\cdot\boldsymbol{r}'} \, d\mathcal{S}',$$
$$\boldsymbol{T}_m(\boldsymbol{n}) = -\int_{\mathcal{A}} \boldsymbol{N} \times \boldsymbol{E}_a \, e^{-ik\boldsymbol{n}\cdot\boldsymbol{r}'} \, d\mathcal{S}', \tag{146}$$

where, in the upper equation, we used an assumed property of the electromagnetic fields on the aperture plane, given in equations (145).

Now, we can express the far field solution of the electric field, transmitted from an aperture antenna, using one of the equations (143)

$$\boldsymbol{E}(\boldsymbol{r},\, t) = -i \, \frac{1}{2\lambda} \frac{e^{i(kr-\omega t)}}{r} \int_{\mathcal{A}} \boldsymbol{n} \times \{\boldsymbol{n} \times [\boldsymbol{N} \times (\boldsymbol{N} \times \boldsymbol{E}_a)] - \boldsymbol{N} \times \boldsymbol{E}_a\} \, e^{-ik\boldsymbol{n}\cdot\boldsymbol{r}'} \, d\mathcal{S}', \tag{147}$$

For large aperture antennas, the electric field in the far field region has finite strength in the direction $\boldsymbol{n}$, which is close to the direction $\boldsymbol{N}$ perpendicular to the aperture plane. Therefore, we approximately take $\boldsymbol{n} = \boldsymbol{N}$ everywhere, except for in the argument of the exponential function. Then, using a formula of the vector algebra in equation (14) and one of the conditions on the aperture plane, given in equations (145), we obtain an equation:

$$\boldsymbol{E}(\boldsymbol{r},\, t) = -i \, \frac{1}{\lambda} \frac{e^{i(kr-\omega t)}}{r} \int_{\mathcal{A}} \boldsymbol{E}_a(\boldsymbol{r}') \, e^{-ik\boldsymbol{n}\cdot\boldsymbol{r}'} \, d\mathcal{S}', \tag{148}$$

which describes a relationship between electric fields on an aperture plane and in a spherical wave in the far field region.

2.5.6 Aperture Illumination and Field Pattern of an Aperture Antenna

In equation (148), the directional characteristics of the aperture antenna are solely determined by the integral:

$$\int_{\mathcal{A}} \boldsymbol{E}_a(\boldsymbol{r}') \, e^{-ik\boldsymbol{n}\cdot\boldsymbol{r}'} \, d\mathcal{S}'.$$

Regarding the distribution of the electric field $\boldsymbol{E}_a(\boldsymbol{r}')$ on the aperture plane, we assume that the electric field is parallel, and keep a constant direction, everywhere (Figure 45). This corresponds to a case of the linearly polarized oscillation. If the actual field distribution is a superposition of different polarization components, for example, a circularly polarized oscillation, which

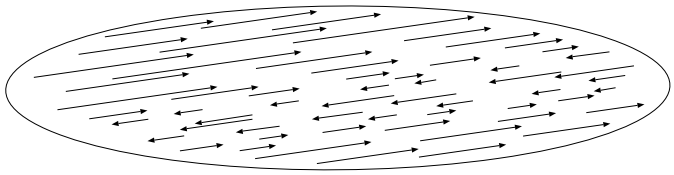

Figure 45: Harmonically oscillating electric field on the aperture plane, keeping a common field direction.

can be decomposed into two linearly polarized oscillations along two mutually perpendicular directions, this discussion is valid for a linear polarization component along a certain direction.

Let us express, in this case, the electric field on the aperture plane as

$$\boldsymbol{E}_a(\boldsymbol{r}') = \boldsymbol{E}_0\, g(\boldsymbol{r}'), \tag{149}$$

where $\boldsymbol{E}_0$ is a constant real vector, and $g(\boldsymbol{r}')$ is a complex (in general) function of the position on the aperture plane. The function $g(\boldsymbol{r}')$ is called "aperture illumination". This shows how the primary wave source (for example, a feed horn) "illuminates" the aperture plane. Using the aperture illumination, we can rewrite equation (148) in a form:

$$\boldsymbol{E}(\boldsymbol{r},\, t) = -i\, \boldsymbol{E}_0\, \lambda\, \frac{e^{i(kr-\omega t)}}{r}\, f(\boldsymbol{n}),$$

where

$$f(\boldsymbol{n}) = \int\limits_{\mathcal{A}} g(\boldsymbol{r}')\, e^{-ik\boldsymbol{n}\cdot\boldsymbol{r}'}\, \frac{d\mathcal{S}'}{\lambda^2}. \tag{150}$$

The non–dimensional function $f(\boldsymbol{n})$ is the field pattern of the aperture antenna in the far field region, which shows the directional characteristics of the generated electric field.

Taking into account that $k = 2\pi/\lambda$, and formally extending the area of the integration to an infinite plane, by just assuming that $g(\boldsymbol{r}') = 0$ outside of the aperture plane $\mathcal{A}$, we obtain, from equation (150),

$$f(\boldsymbol{n}) = \int g(\boldsymbol{r}')\, e^{-i2\pi\boldsymbol{n}\cdot\frac{\boldsymbol{r}'}{\lambda}}\, \frac{d\mathcal{S}'}{\lambda^2}. \tag{151}$$

This equation shows that the aperture illumination $g(\boldsymbol{r}')$ and the field pattern in the far field region $f(\boldsymbol{n})$ of the aperture antenna form **a two dimensional spatial Fourier transform pair**.

2.5.7 Power Pattern of an Aperture Antenna

We can now derive the power pattern of an aperture antenna by calculating time–averaged Poynting vector, as we did for the Hertz dipole.

We saw in equation (97) that the Poynting vector $\boldsymbol{S}$ of the spherical wave is described as

$$\boldsymbol{S} = \boldsymbol{E}_r \times \boldsymbol{H}_r = \frac{1}{Z}\,\boldsymbol{n} \mid \boldsymbol{E}_r \mid^2,$$

where $\boldsymbol{E}_r$ is the real part of the complex form of the electric field, which represents the real physical quantity. Introducing the amplitude $\mid f(\boldsymbol{n}) \mid$ and the phase ϕ of the complex field pattern $f(\boldsymbol{n})$, we have

$$f(\boldsymbol{n}) = \mid f(\boldsymbol{n}) \mid \; e^{i\phi}. \tag{152}$$

Then, from equation (150), we obtain

$$\boldsymbol{E}_r = \Re\boldsymbol{E}(\boldsymbol{r},\, t) = \boldsymbol{E}_0\,\lambda\,\frac{\mid f(\boldsymbol{n}) \mid}{r}\,\sin(kr - \omega t + \phi). \tag{153}$$

Since the time–averaged squared sine function always gives

$$\langle \sin^2(kr - \omega t + \phi) \rangle = \frac{1}{2},$$

where $\langle\ \rangle$ denotes time averaging, we have

$$\langle \mid \boldsymbol{S} \mid \rangle = \frac{1}{2Z} \mid \boldsymbol{E}_0 \mid^2 \frac{\lambda^2}{r^2} \mid f(\boldsymbol{n}) \mid^2. \tag{154}$$

The non–dimensional term $\mid f(\boldsymbol{n}) \mid^2$ in equation (154) characterizes directional pattern of the transmitted power, and could be regarded as power pattern $P(\boldsymbol{n})$ in the far field region of an aperture antenna, though sometimes $\langle \mid \boldsymbol{S} \mid \rangle$ itself is called power pattern. In practice, "**normalized power pattern**" $P_n(\boldsymbol{n})$, which is obtained by dividing the power pattern by its maximum value, is used more frequently than the power pattern. Thus, the normalized power pattern in the far field region of an aperture antenna is given by a formula:

$$P_n(\boldsymbol{n}) = \frac{P(\boldsymbol{n})}{P_{max}} = \frac{\mid f(\boldsymbol{n}) \mid^2}{\mid f_{max} \mid^2}, \tag{155}$$

which is not affected by the difference of the two definitions of the power pattern, mentioned above.

2.5.8 Main Lobe, Sidelobes, HPBW and BWFN

A typical normalized power pattern of an aperture antenna is illustrated in Figure 46. Left panel shows a rectangular plot, with horizontal axis showing angle from the "aperture axis" in radian. The aperture axis is an axis perpendicular to the aperture plane, which is usually parallel to the "beam axis", i.e., the direction of the maximum power pattern. Right panel shows a polar plot, which gives more realistic view of the pattern shape.

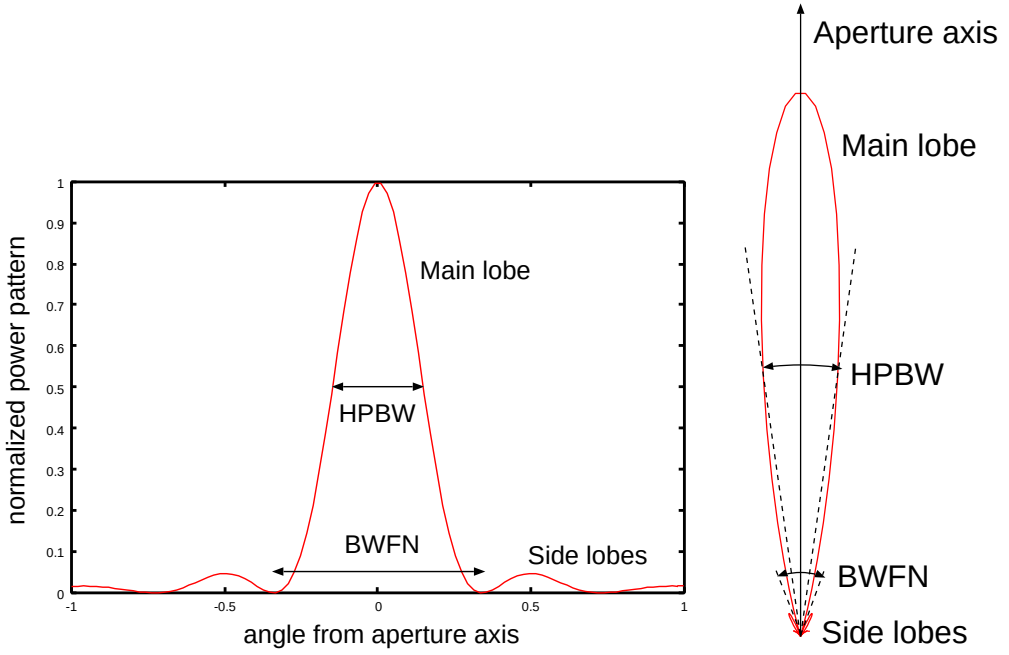

Figure 46: A normalized power pattern of an aperture antenna in a rectangular plot (left) and in a polar plot (right).

In both panels, a large **main lobe** (or **main beam**) is prominent. There are small **sidelobes** in both sides of the main lobe.

Two quantities are frequently used as measures of the width of the main lobe, which is an important characteristic of an aperture antenna. One is **"full half power beam width", or HPBW**, which stands for the twice of the angle of a direction, where the power pattern takes a half of the maximum value, from the aperture axis. Another is **"full beam width between first nulles", or BWFN**, which stands for the twice of the angle of a direction, where the power pattern drops to zero between the main lobe and the first side lobe, from the aperture axis (Figure 46).

2.5.9 Uniformly Illuminated Rectangular Aperture Antenna

Let us consider, as one of the simplest examples of aperture antennas, a rectangular aperture antenna with an aperture plane located in xy–plane, which is uniformely illuminated as:

$$g(x,y) = \begin{cases} 1 & \text{for } \mid x \mid \leq \frac{L_x}{2}, \text{ and } \mid y \mid \leq \frac{L_y}{2}, \\ 0 & \text{else,} \end{cases} \tag{156}$$

(see Figure 47).

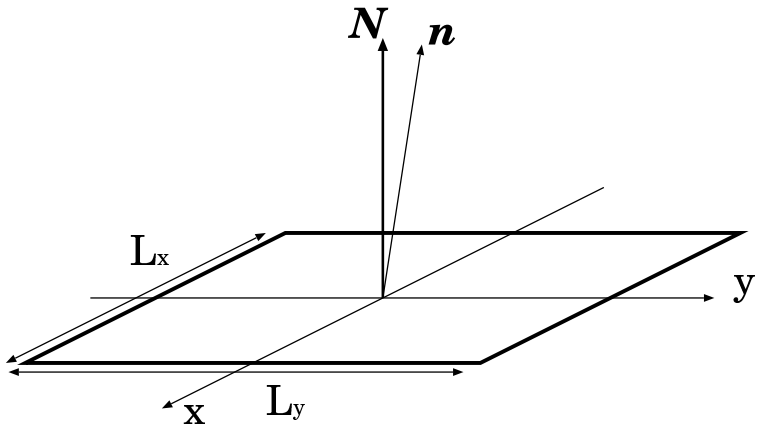

Figure 47: A rectangular aperture antenna.

From equation (151), the field pattern in the far field region $f(\boldsymbol{n})$, in a direction of a unit vector $\boldsymbol{n}$, with components n_x and n_y in the xy–plane, is expressed in this case as

$$f(\boldsymbol{n}) = \iint g(x,y) e^{-i2\pi(n_x x + n_y y)/\lambda} \frac{dx}{\lambda}\frac{dy}{\lambda} = \int_{-\frac{L_y}{2}}^{\frac{L_y}{2}} \int_{-\frac{L_x}{2}}^{\frac{L_x}{2}} e^{-i2\pi(n_x x + n_y y)/\lambda} \frac{dx}{\lambda}\frac{dy}{\lambda}$$

$$= \frac{\sin(\pi n_x L_x/\lambda)}{\pi n_x} \frac{\sin(\pi n_y L_y/\lambda)}{\pi n_y}. \tag{157}$$

Since this function takes the maximum value $(L_x/\lambda)(L_y/\lambda)$ when $\boldsymbol{n} = \boldsymbol{N}$, where $\boldsymbol{N}$ is the unit vector normal to the aperture plane, and, therefore, $n_x = 0,\ n_y = 0$, the normalized power pattern, as given in equation (155), is

$$P_n(\boldsymbol{n}) = \frac{\mid f(\boldsymbol{n}) \mid^2}{\mid f_{max} \mid^2} = \left[\frac{\sin(\pi n_x L_x/\lambda)}{\pi n_x L_x/\lambda} \frac{\sin(\pi n_y L_y/\lambda)}{\pi n_y L_y/\lambda} \right]^2. \tag{158}$$

Full beam width between first nulls and full half power beam width of such an antenna are

$$\text{BWFN} = 2\,(\frac{\lambda}{L_x} \text{ and } \frac{\lambda}{L_y}) \text{ (rad)} = 114.^\circ 6\,(\frac{\lambda}{L_x} \text{ and } \frac{\lambda}{L_y}),$$

and

$$\text{HPBW} = 0.88\,(\frac{\lambda}{L_x} \text{ and } \frac{\lambda}{L_y}) \text{ (rad)} = 50.^\circ 4\,(\frac{\lambda}{L_x} \text{ and } \frac{\lambda}{L_y}), \tag{159}$$

respectively.

2.5.10 Circular Aperture Antenna

Now let us consider a circular aperture antenna with diameter D, as shown in Figure 48. Let us assume an axisymmetric illumination around the aperture

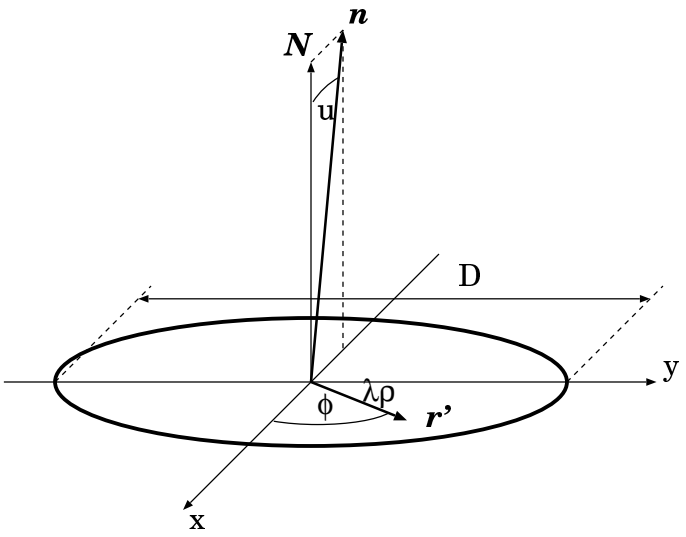

Figure 48: A circular aperture antenna.

axis, which is directed towards a unit vector $\boldsymbol{N}$, perpendicular to the aperture plane, and let us choose a coordinate system, in which a directional unit vector $\boldsymbol{n}$ and a radius vector $\boldsymbol{r}'$ within the aperture plane are given as

$$\boldsymbol{n} \cong (-u,\, 0,\, 1), \tag{160}$$

and

$$\boldsymbol{r}' = (\lambda\rho\cos\phi,\, \lambda\rho\sin\phi,\, 0), \tag{161}$$

where λ is the wavelength, and ρ and ϕ are radial and azimuthal variables.

If we denote the axisymmetric aperture illumination as

$$g(\boldsymbol{r}') = g(\rho), \tag{162}$$

then, according to equation (151), the field pattern at the direction $\boldsymbol{n}$ is expressed in a form:

$$f(u) = \int_0^{2\pi}\int_0^{\infty} g(\rho)\, e^{i2\pi u\rho\cos\phi}\rho d\rho d\phi. \tag{163}$$

Since Bessel function of the zeroth order $J_0(z)$ is defined as

$$J_0(z) = \frac{1}{2\pi}\int_0^{2\pi} e^{iz\cos\phi} d\phi, \tag{164}$$

we have

$$f(u) = 2\pi \int_0^{\infty} g(\rho) J_0(2\pi u\rho)\rho d\rho. \tag{165}$$

Therefore, according to equation (155), the normalized power pattern is

$$P_n(u) = \left[\frac{\int_0^{\infty} g(\rho) J_0(2\pi u\rho)\rho d\rho}{\int_0^{\infty} g(\rho)\rho d\rho}\right]^2. \tag{166}$$

We consider the simplest case of uniform illumination:

$$g(\rho) = \begin{cases} 1 & \text{for } \rho \le \frac{D}{2\lambda}, \\ 0 & \text{else}, \end{cases} \tag{167}$$

Then the normalized power pattern in equation (166) is reduced to

$$P_n(u) = \left[\frac{\int_0^{D/2\lambda} J_0(2\pi u\rho)\rho d\rho}{\int_0^{D/2\lambda} \rho d\rho}\right]^2 = \left[\frac{2\lambda^2}{\pi^2 u^2 D^2}\int_0^{\pi u D/\lambda} J_0(z) z dz\right]^2. \tag{168}$$

From the recurrence formula of Bessel function:

$$\frac{d}{dz}\{z^n J_n(z)\} = z^n J_{n-1}(z), \tag{169}$$

we have

$$x^n J_n(x) = \int_0^x z^n J_{n-1}(z) dz,$$

and hence, in a particular case of $n = 1$,

$$\int_0^x J_0(z) z dz = x J_1(x).$$

Inserting this formula into equation (168), we obtain the normalized power pattern of the uniformly illuminated circular aperture antenna in the far field region:

$$P_n(u) = \left[\frac{2\lambda}{\pi u D} J_1\left(\frac{\pi u D}{\lambda}\right)\right]^2. \tag{170}$$

Figures 49 and 50 show normalized power patterns (left pannel) and field amplitude pattern $\mid f(u) \mid / \mid f_{max} \mid$ (right panel) of a uniformly illuminated circular aperture antenna in rectangular and polar plots. In the rectangular plots (Figure 49), the horizontal axes show offset angle from the aperture axis normalized by λ/D.

Figure 51 shows three–dimensional rectangular and polar plots of the normalized power pattern and the field amplitude pattern of the uniformly illuminated circular aperture antenna.

From numerical values of Bessel function $J_1(x)$, we obtain full beam width between first nulls and full half power beam width of the uniformly illuminated circular aperture antenna as

$$\begin{aligned} \text{BWFN} = 2.44\,\frac{\lambda}{D}\ (\text{rad}) &= 140^\circ\,\frac{\lambda}{D}, \\ \text{and} \\ \text{HPBW} = 1.02\,\frac{\lambda}{D}\ (\text{rad}) &= 58.^\circ 4\,\frac{\lambda}{D}, \end{aligned} \tag{171}$$

respectively.

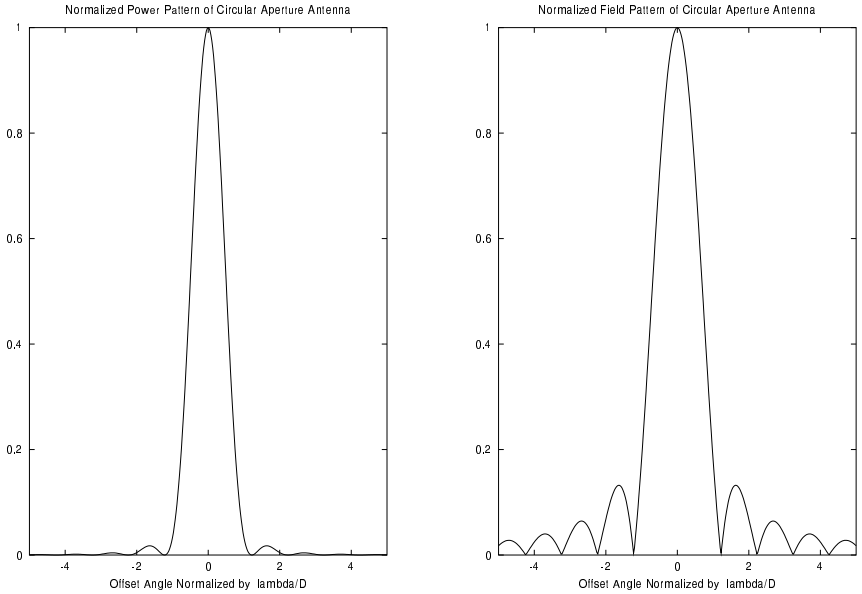

Figure 49: Rectangular plots of normalized power pattern (left) and field amplitude pattern (right) of a uniformly illuminated circular aperture antenna. Horizontal axes show offset angle from the aperture axis normalized by λ/D.

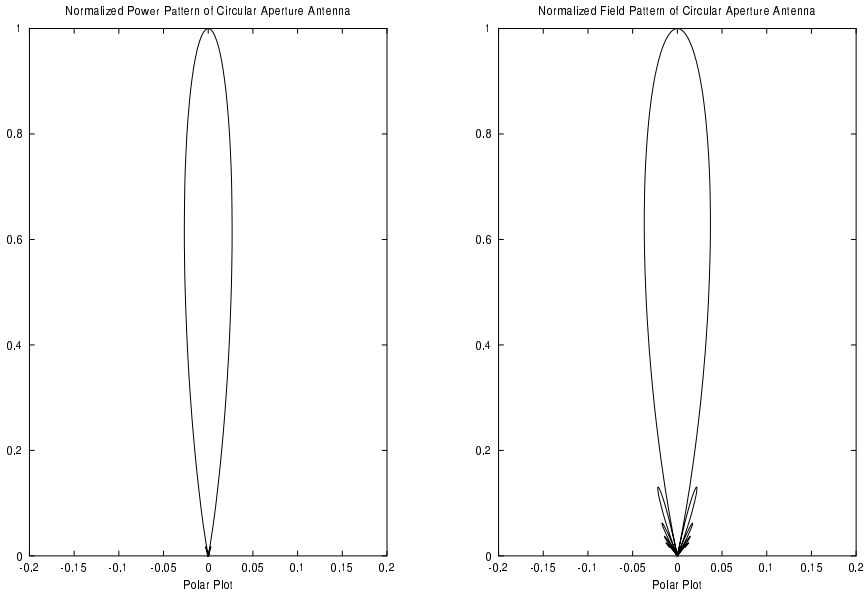

Figure 50: Polar plots of normalized power pattern (left) and field amplitude pattern (right) of a uniformly illuminated circular aperture antenna.

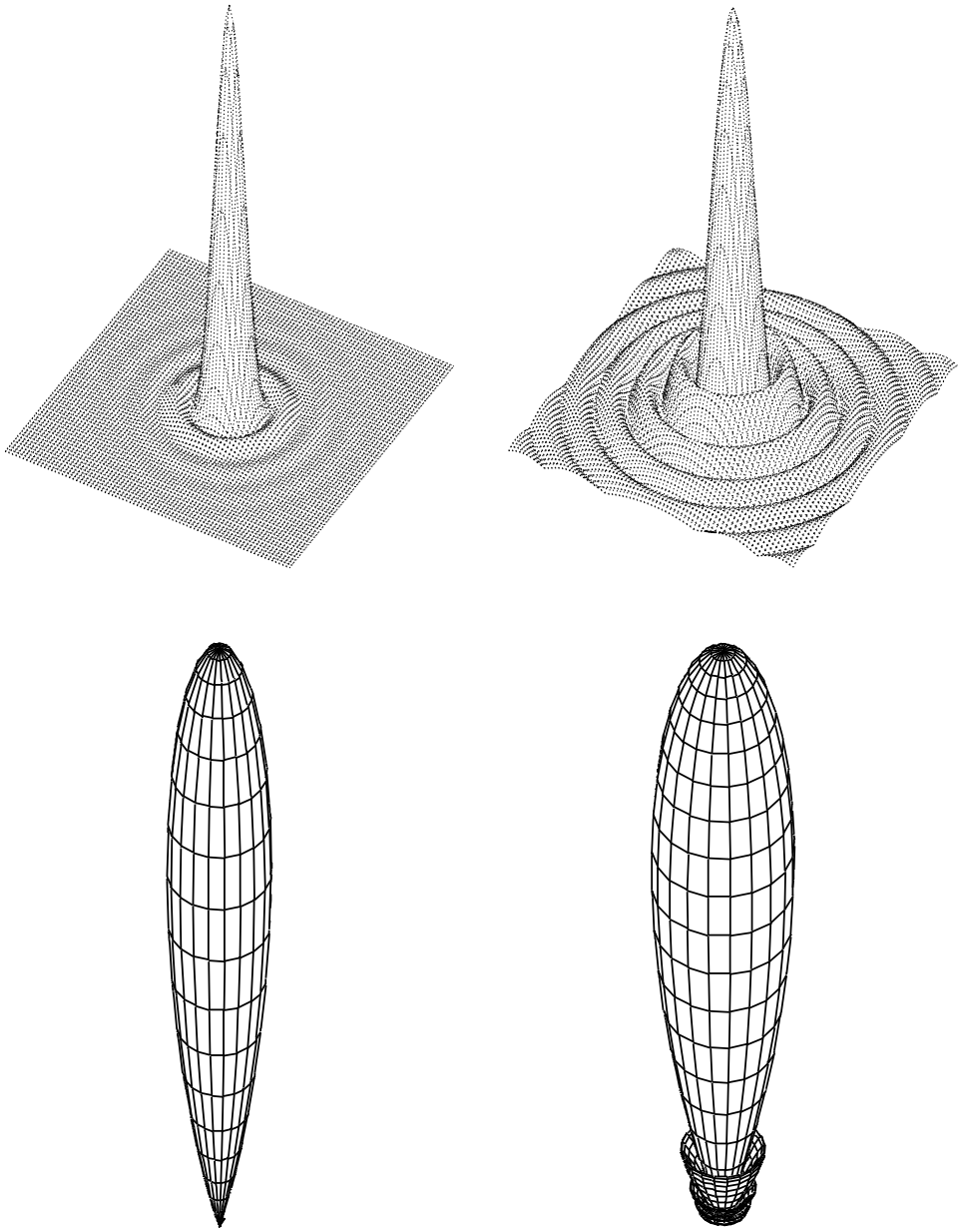

Figure 51: Three–dimensional rectangular (top) and polar (bottom) plots of the normalized power pattern (left) and the field amplitude pattern (right) of a uniformly illuminated circular aperture antenna. Sidelobes are hardly seen in the polar plot of the normalized power pattern.

2.6 Beam Patters of Aperture Antennas

So far, we have discussed and derived field and power patterns of aperture antennas in the transmission mode. Because of the equivalence of field patterns in transmission and reception, which we saw earlier, the field transmission pattern must be the same with the field reception pattern. In the reception case, the field pattern is often called 'voltage reception pattern', as describing the magnitude of the received voltage as a function of direction.

Since the transmitted power is proportional to the square of the electric field, and the received power is proportional to the square of the received voltage, the power reception pattern must also be equivalent to the power transmission pattern.

The antenna beam pattern is a general concept, which means either the power pattern or the field (or voltage–reception) pattern, depending on the context.

2.6.1 Antenna–Fixed Coordinate System

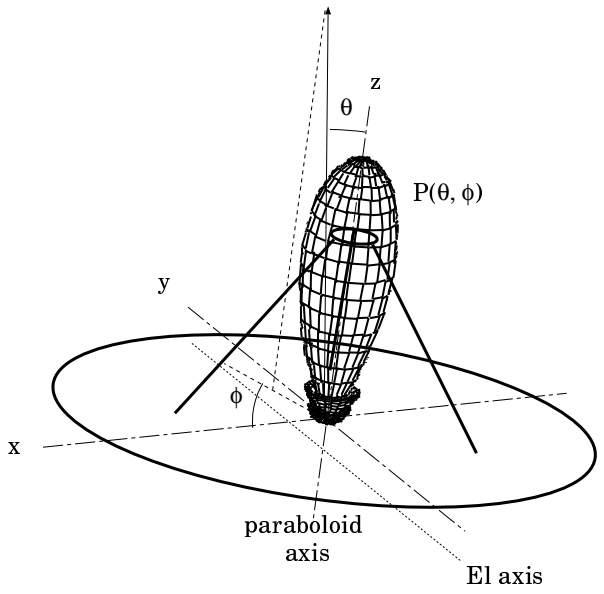

Figure 52: Antenna–fixed coordinate system.

For further convenience, let us introduce an 'antenna–fixed coordinate system' for a paraboloidal antenna.

As the origin of the coordinate system, we choose the center of the paraboloid surface of the main reflector. Then, let us take the aperture axis, or the symmetry axis of the paraboloid, as the polar axis z, and choose the x and y axes towards directions perpendicular and parallel to the elevation axis, as shown in Figure 52. Finally, let us denote a given direction in the sky by an angle from the polar axis θ, and an azimuthal angle ϕ measured from the x–axis towards the y–axis.

Then, the field pattern and power pattern can be described as functions of these angles: such as $f(\theta, \phi)$, and $P(\theta, \phi)$. Hence, the normalized power pattern can be described as

$$P_n(\theta, \phi) = \frac{P(\theta, \phi)}{P_{max}} = \frac{\mid f(\theta, \phi) \mid^2}{\mid f_{max} \mid^2}. \tag{172}$$

2.6.2 A Useful Formula: HPBW $\approx \lambda/D$

From the two simple examples of the aperture antennas, which we saw earlier, it is evident that HPBW $\cong \lambda/D$ (see equations (159) and (171)). This is generally valid for a wide variety of aperture antennas with realistic aperture illuminations. Therefore, the ratio λ/D is often used as a measure of the beam width, or the angular resolution, of radio telescope antennas (table 1).

diameter D	wavelength λ	frequency ν	λ/D
6 m	37.5 mm	8 GHz	0.°36
10 m	7.0 mm	43 GHz	0.°04
21 m	2.3 mm	129 GHz	0.°0063
45 m	2.0 mm	150 GHz	0.°0025

Table 1: Beam widths ($\lambda/D \approx$ HPBW) of antennas with different aperture diameters, observing at different frequencies.

Of course, it is not an easy task to direct a huge radio telescope with 45 m diameter towards an astronomical object with an accuracy much better than 0.0025 degree!

2.6.3 Distance to Fraunhofer Region

As we saw earlier, the antenna beam pattern is a function of distance from the antenna in the Fresnel region, or in the near field, as schematically shown in

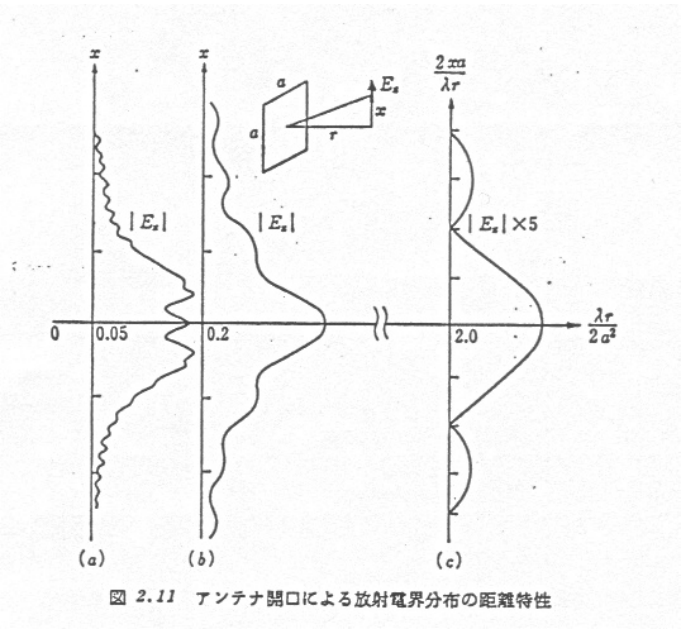

図 2.11 アンテナ開口による放射電界分布の距離特性

Figure 53: Variation of the electric field amplitude pattern with distance from the antenna (S. Adachi, Electromagnetic Wave Engineering, 1985).

Figure 53 (S. Adachi, Electromagnetic Wave Engineering, Korona-sha, 1985, in Japanese). However, in the Fraunhofer region, or in the far field, the beam pattern becomes almost independent of the distance, as illustraited by the rightmost curve of Figure 53.

For a paraboloidal antenna, we can use the aperture diameter as the characteristic size of the wave source region in the Fraunhofer condition given in (89). Therefore, the 'distance to the Fraunhofer region' r_F for a paraboloidal antenna, with the aperture diameter D, receiving or transmitting a radio wave with wavelength λ, is given by

$$r_F \equiv \frac{2D^2}{\lambda}. \tag{173}$$

The Fraunhofer condition $r \gg r_F$ can be graphically illustrated as the condition where the difference between the distance from a distant point p to the edge of the aperture r', and that to the center of the aperture r, must be much smaller than $\lambda/16$ (Figure 54). This could be interpretted as a condition, that a spherical wave coming from the distant point can be practically approximated as a plane wave.

Also, r_F is equal to the twice of the distance r_w to a point, where the width of the far–field beam, measured at HPBW, becomes equal to the width

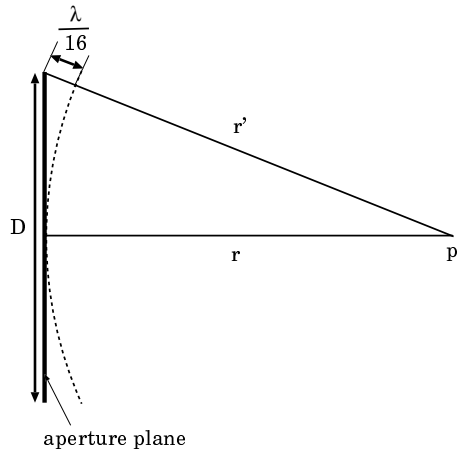

Figure 54: Distance to the Fraunhofer region as a condition for the plane wave approximation of the received wave.

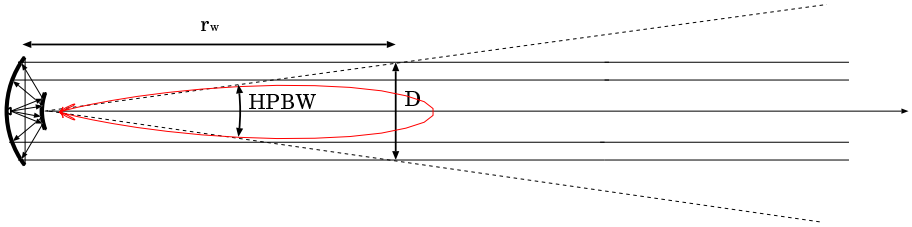

Figure 55: Distance to the Fraunhofer region as a condition for predominance of the far–field beam in the transmitted wave.

of a circular cylinder with the diameter of the aperture D, i.e.,

$$\frac{\lambda}{D} r_w = D, \quad \text{and , hence,} \quad r_w = \frac{D^2}{\lambda} = \frac{r_F}{2},$$

(Figure 55). This could be interpreted as a condition that most of the transmitted power is confined within the far–field beam.

Table 2 shows typical examples of the distances to the Fraunhofer region. Can we measure the far–field patterns of our antennas on the ground?

D	λ	$r_F = 2D^2/\lambda$
6 m	3.8 cm (8 GHz)	1.9 km
10 m	7 mm (43 GHz)	28.6 km
21 m	2.3 mm (129 GHz)	383 km
45 m	2 mm (150 GHz)	2025 km
2300 km	7 mm (43 GHz)	1.51×10^{12} km
VLBI		= 0.16 light year

Table 2: Distances to the Fraunhofer region.

2.7 Illumination Taper (or Gradation)

We can modify antenna beams by changing the electric field distribution on

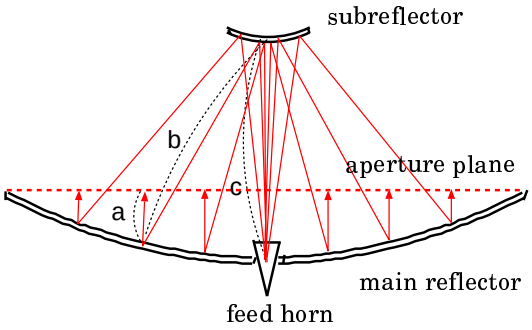

Figure 56: Aperture illumination.

the aperture plane (the illumination pattern). Therefore, in order to realize

a desired beam pattern, we can design a feed horn, so that the horn suitably 'illuminates' the aperture plane (Figure 56).

Generally speaking, it is desirable to have as narrow a main lobe as possible and sidelobe levels as low as possible. However, the two requirements are somewhat contradictory to each other, as shown in Figure 58.

As we saw earlier, the aperture illumination pattern (distribution of the electric field on the aperture plane) and the far–field pattern form a two–dimensional spatial Fourier transform pair. Therefore, the beam patterns are derived from the illumination patterns by simple Fourier transformation.

The main lobe is the narrowest when the aperture illumination is uniform, as shown in Figure 58. However, the sidelobe level is rather high in the uniform illumination case. On the other hand, if we apply a Gaussian illumination, we get fairly low sidelobe levels, but the main lobe becomes broader. This kind of non–uniform illumination, which is usually stronger in the central part of the aperture, and weaker in the outer parts, is called 'tapered' or 'gradated' illumination.

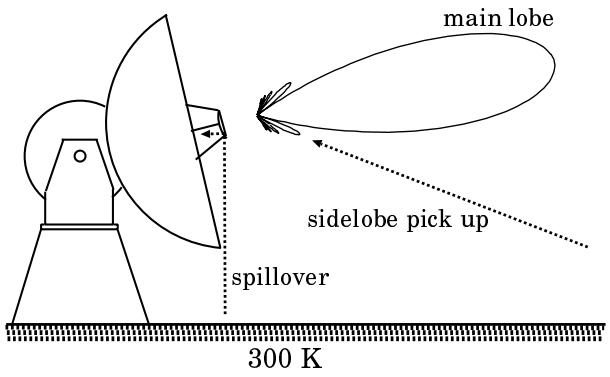

Figure 57: Why is strong tapering required for radio telescopes?

For radio telescopes, which receive very weak signals from astronomical radio sources, the thermal radiation of the ground — with temperature $T \simeq 300$ K — picked up by the sidelobes may seriously degrade the signal–to–noise (S/N) ratio. Likewise, the ground thermal radiation received from the outside of the main reflector edge (which is called 'spillover' in the transmission sense) is also quite harmful (Figure 57).

Therefore, rather strong tapering is required for radio telescopes, although this of course means that we have to sacrifice some portion of their aperture areas as 'ineffective' areas.

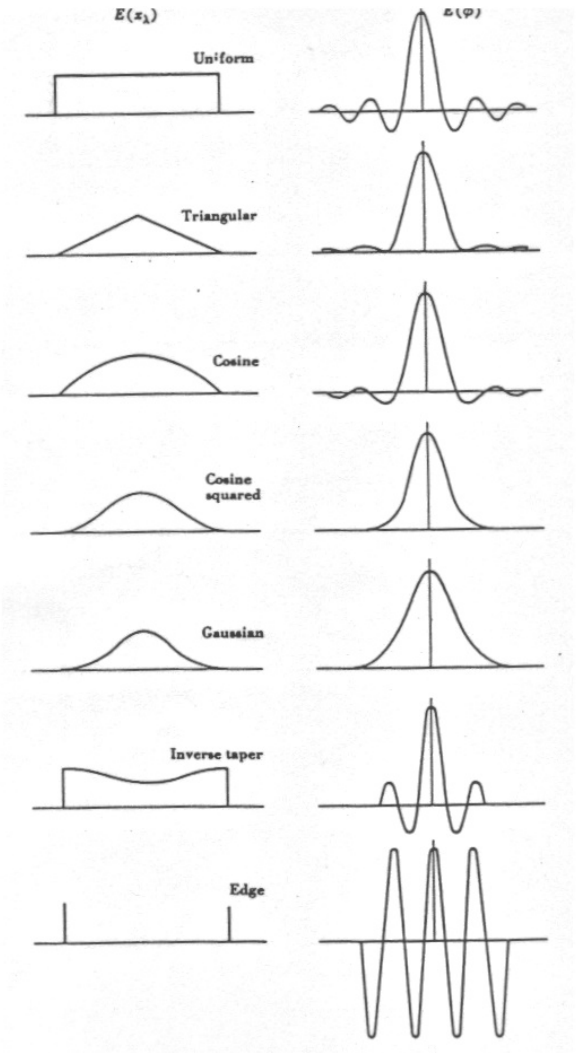

Figure 58: Aperture illumination patterns (left), and the resulting beam patterns (right).

2.8 Spectral Flux Density Received by an Antenna Beam

We defined earlier that the spectral flux density $\mathcal{S}_\nu$ is a portion of the radiation energy from an astronomical source incoming through a cross section of unit area, per unit frequency bandwidth, and per unit time. The spectral flux density was described, through the monochromatic intensity $I_\nu(\theta, \phi)$, by an integral over the source solid angle Ω :

$$\mathcal{S}_\nu = \iint_\Omega I_\nu(\theta, \phi) \cos\theta \sin\theta \; d\theta \; d\phi,$$

where the $\cos\theta$ term arose due to a geometrical effect corresponding to the projection of a unit area onto the plane perpendicular to an incoming ray (left panel of Figure 59).

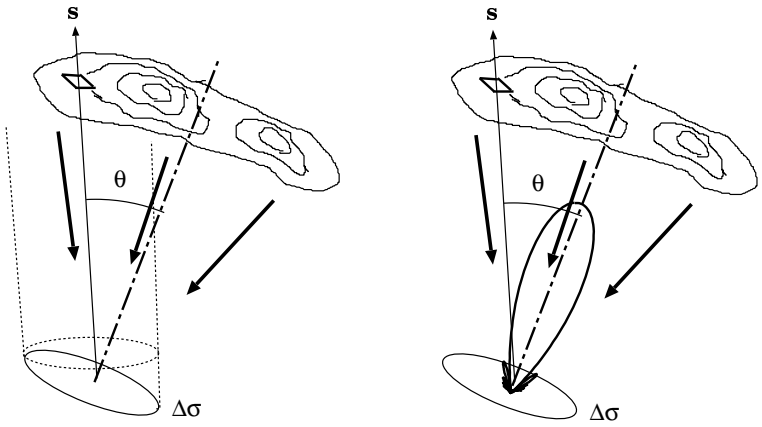

Figure 59: Flux density as received by a unit area in space (left), and the effective flux density as received by a unit area on the aperture plane of a directional antenna (right).

On the other hand, an antenna with a normalized power pattern $P_n(\theta, \phi)$ receives the radiation from the radio source in proportion to $P_n(\theta, \phi)$, which reflects not only the geometrical effect but also the electromagnetic effect. Therefore, the portion of the radiation energy received by a unit area on the antenna aperture plane, per unit frequency bandwidth, and per unit time, is described by

$$\mathcal{S}_\nu = \iint_\Omega P_n(\theta, \phi) \, I_\nu(\theta, \phi) \sin\theta \; d\theta \; d\phi, \tag{174}$$

where θ and ϕ are the angular variables in the antenna–fixed coordinate system (right panel of Figure 59). We call this 'intensity collected by the antenna beam' $\mathcal{S}_\nu$ in equation (174) the 'effective flux density'. The effective flux density is equal to the usual total flux density when the angular size of the source is much smaller than the antenna beam width, and when the beam is exactly directed towards the source. Otherwise, the effective flux is smaller than the total flux. Henceforth, we use the word 'flux density' in the sense of the effective flux density, as far as radio wave reception with radio telescope antennas is concerned.

3 Antenna Characteristics

The performance of a radio wave antenna is characterized by a number of important parameters which usually figure in the major specifications of the antennas (an example of antenna specifications is shown in table 3).

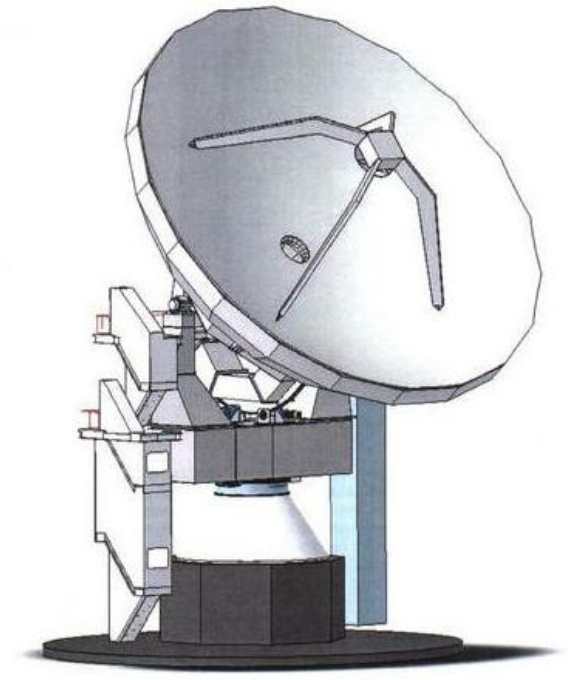

Figure 60: Design drawing of a 21 m antenna.

We will introduce major performance characteristics of antennas, and discuss meanings of numbers in the specifications. We will also discuss ways

Optics	Cassegrain focus, central feed type	
	Primary diameter	21 m
	Primary focal length	7 m
	Secondary diameter	2 m
	Secondary focal length	37 cm
Mechanical operation	Fast switching up to 2.5° separation	
	Azimuth range	±270°
	Elevation range	0° – 91°
	Az max. slew speed	3°/s
	El max. slew speed	3°/s
	Az max. acceleration	$\geq 1°/s^2$
	El max. acceleration	$\geq 1°/s^2$
Surface accuracy	Appropriate for reception at 150 GHz	
	Panels	65 μm rms
	Total	150 μm rms (10 m/s wind)
Pointing performance	Appropriate for reception at 150 GHz	
	Pointing accuracy	0.001° rms (10 m/s wind)
	Az natur. freq.	2.5 Hz
	El natur. freq.	2.5 Hz
	Subref. spar nat. freq.	5 Hz
	Az El axes offset	3 mm
	Az El orthogonality	0.01°
Operational loads	Temperature	-20°C – 40°C
	Wind	10 m/s
Survival Loads	Wind	90 m/s
	Seismic acceleration	0.2 G horiz., 0.1 G vert.

Table 3: An example of antenna specifications.

to estimate important charactersitics such as aperture eficiency, antenna gain, and maximum observable frequency, on the basis of the specifications.

3.1 Directive Gain $G(\theta, \phi)$

The directive gain $G(\theta, \phi)$ of an antenna is defined in terms of the power pattern $P(\theta, \phi)$, in the antenna–fixed coordinate system with the angular variables θ (angular distance from the symmetry axis) and ϕ (azimuthal angle in the aperture plane), as

$$G(\theta, \phi) = \frac{P(\theta, \phi)}{P} = \frac{4\pi P(\theta, \phi)}{\oint P(\theta, \phi) d\Omega}, \tag{175}$$

where

$$P \equiv \frac{1}{4\pi} \oint P(\theta, \phi) d\Omega$$

is the mean value of the power pattern averaged over all directions.

Why is this quantity called directive "gain"? In the transmission case, the directive gain is a quantity describing how well your antenna can emit the radio signal of a given total power towards the direction θ and ϕ, compared with an idealized antenna which would emit the same power of the radio signal isotropically. In the reception case, if you observe a point radio source at θ and ϕ, you would receive $G(\theta, \phi)$ times stronger power with your directional antenna than what you would receive with an idealized isotropic antenna. In both cases, you 'gain' stronger signal with the directional antenna, compared with the idealized isotropic antenna (Figure 61).

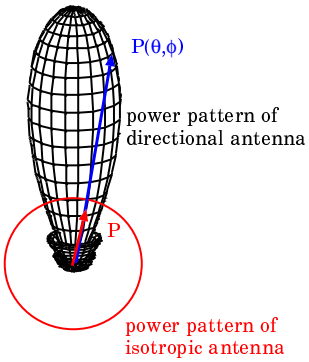

Figure 61: Directive gain: "gain" of a directional antenna compared with an idealized isotropic antenna.

3.2 Beam Solid Angle Ω_A

The beam solid angle Ω_A is a measure of the angular extent of the antenna beam, as defined by the equation :

$$\Omega_A = \oint P_n(\theta,\phi)d\Omega = \frac{\oint P(\theta,\phi)d\Omega}{P_{max}} = \frac{4\pi P}{P_{max}}\,, \tag{176}$$

where $P_n(\theta,\phi)$ is the normalized power pattern :

$$P_n(\theta,\phi) = \frac{P(\theta,\phi)}{P_{max}}\,.$$

If we display the antenna power pattern in a rectangular plot, where the x– and y– coordinates are angles in two perpendicular directions in the aperture plane, and if we approximate the solid angle by an area in this xy– plane, the beam solid angle is equal to the cross–sectional area of a cylinder which has a height of P_{max}, and the same "volume" as the power pattern in this rectangular coordinate system (Figure 62).

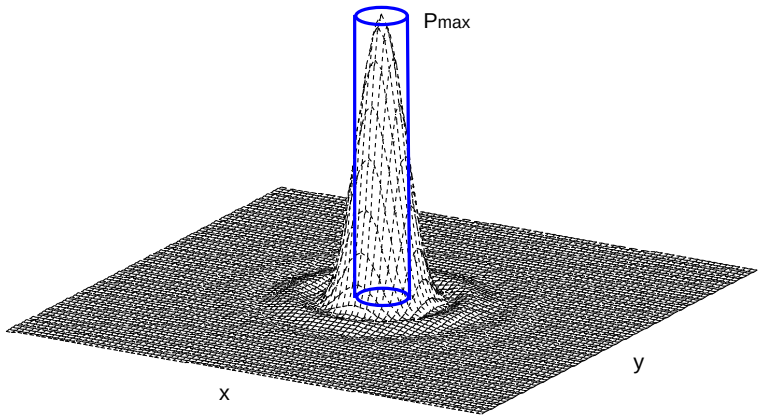

Figure 62: Beam solid angle as the cross-sectional area of a cylinder with a height of P_{max}, and the same volume as the antenna power pattern in this rectangular plot of the beam.

3.3 Main Beam Solid Angle Ω_M

The main beam solid angle is defined in a similar way to the beam solid angle, but now the integration covers the main lobe only:

$$\Omega_M = \int\limits_{main\ \ lobe} P_n(\theta,\phi)d\Omega. \tag{177}$$

In the rectangular plot we saw above, the main beam solid angle is equal to the cross–sectional area of a cylinder which has the height of P_{max}, and the same volume as the main lobe of the power pattern.

3.4 Main Beam Efficiency η_M

The main beam efficiency (or simply the beam efficiency) is the ratio of the main beam solid angle to the beam solid angle :

$$\eta_M = \frac{\Omega_M}{\Omega_A}. \tag{178}$$

η_M is close to 1 if the antenna has a sharp main lobe, and sufficiently low sidelobe levels.

3.5 Directivity, Maximum Directive Gain, or "Gain" $\mathcal{D}$

The directivity is the maximum value of the directive gain, which is usually obtained in the direction of the symmetry axis :

$$\mathcal{D} = G_{max} = \frac{P_{max}}{P} = \frac{4\pi}{\Omega_A}. \tag{179}$$

The directivity describes how much a directional antenna "gains" at the beam center, as compared with the idealized isotropic antenna. This is an important antenna parameter which is also called the "antenna gain", or simply "gain" (not to be confused with the "directive gain" $G(\theta, \phi)$).

The antenna gain is usually expressed in decibels (dB) :

$$\mathcal{D} \text{ (in dB or dBi)} = 10\, log_{10}(\frac{P_{max}}{P}), \tag{180}$$

where log_{10} is the common (i.e. base 10) logarithm, and "**dBi**" here is the same as **dB** but stresses that the comparison is made with the *isotropic* case.

3.6 Antenna Polarization

3.6.1 Some Notes on Polarization of Electromagnetic Wave

Since an electromagnetic wave is a transverse wave, as we saw in the discussions of the plane and spherical waves, the electric or magnetic field vector in the wave oscillates along a certain direction within a plane which is perpendicular to the direction of the propagation, and moves with the wave

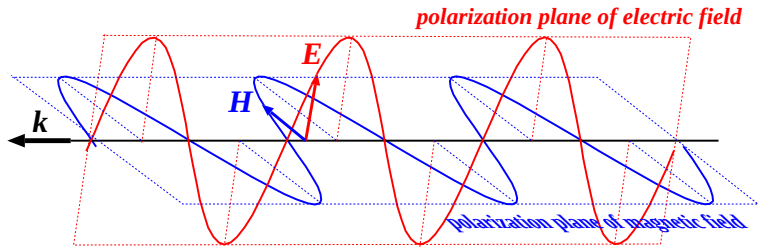

Figure 63: An electromagnetic wave as a polarized wave.

(Figure 63). Such a wave is said to be "polarized". There are two degrees of freedom in the oscillation within the plane. Therefore, any wave with a certain direction of polarization can be expressed as a linear combination of two independent polarization components. Such independent components could be two linearly polarized components in mutually orthogonal directions (horizontal and vertical, say) or two circularly polarized components with mutually opposite rotational senses (RHCP : Right Hand Circular Polarization, and LHCP : Left Hand Circular Polarization), as shown in Figure 64.

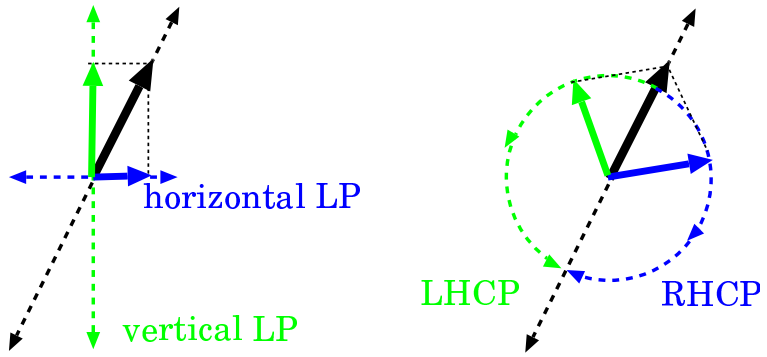

Figure 64: Representation of a polarized wave as a linear combination of two independent linear (left), or circular (right), polarization components.

3.6.2 Polarization characteristics of Antennas

Any antenna (or, better to say: any waveguide to coaxial–cable converter) can receive or transmit only one independent component of the polarization. This can be generally understood in the following way. Although the

electromagnetic wave in free space has two degrees of freedom in the polarization, the oscillation of the voltage or the current in a circuit is essentially one–dimensional. Therefore, we have to lose one degree of freedom, anyway.

For example, if the waveguide to coaxial–cable converter consists of a simple dipole, it receives or transmits only one linear polarization component, with the direction of oscillation parallel to the dipole. If the converter is a helix, on the other hand, it receives or transmits only one circular polarization component. In general, an antenna is designed to receive or transmit a certain polarization component. Therefore, an antenna could be designed, for example, for linear horizontal polarization, or for RHCP, etc.

Of course, the antenna polarization can be relatively easily changed by adjusting the waveguide to coaxial–cable converter or using special optical devices. Moreover, one can install two or more waveguide to coaxial–cable converters, and following receiving systems, in a common main–reflector and sub–reflector system in order to simultaneously receive or transmit the two polarization components by "a single antenna" (which, in actuality, contains two antennas in strict Kraus's sense).

The specification of the antenna polarization performance is usually given as 'polarization isolation'. This shows how well an antenna can receive one polarization component, to which the antenna is designed, without being contaminated by another component, which may arise due to some inperfection in the antenna optical system.

In a majority of radio sources, two independent polarization components are only very weakly correlated ("unpolarized sources"). Therefore, it is quite important for VLBI observations that participating antennas observe radio sources in the same polarization mode, except in the cases when the weak correlation between different polarization components itself is the target of study ("polarization observation"). Wrong setting of polarization has been one of major sources of failures of VLBI observations.

It is not easy to use linear polarization antennas for VLBI observations, because the directions of the linear polarization fixed to the antennas usually correspond to different sky directions, and the difference varies in time as the Earth rotates, for antennas located on different points of the Earth. Therefore, the VLBI antennas in the world are mostly designed for reception of circular polarization components. In geodetic VLBI, RHCP has been conventionally used.

3.7 Effective Aperture A_e

Let us consider that the randomly polarized (or 'unpolarized") radio wave is incident to a radio telescope antenna. The word "unpolarized" here means

that all kinds of waves with different polarization directions are randomly mixed up so that in average no net predominant polarization component is left. Radio waves from most of the astronomical sources are known to be practically unpolarized in this sense.

If an antenna obtains a certain amount of power per unit frequency bandwidth W_ν by observing a randomly polarized radio source with the effective flux density $\mathcal{S}_\nu$, the effective aperture A_e of the antenna is defined by an equation :

$$W_\nu = \frac{1}{2} A_e \mathcal{S}_\nu, \tag{181}$$

where a factor $1/2$ implies that the antenna is able to receive only a half of the power alloted to one independent component of the polarization.

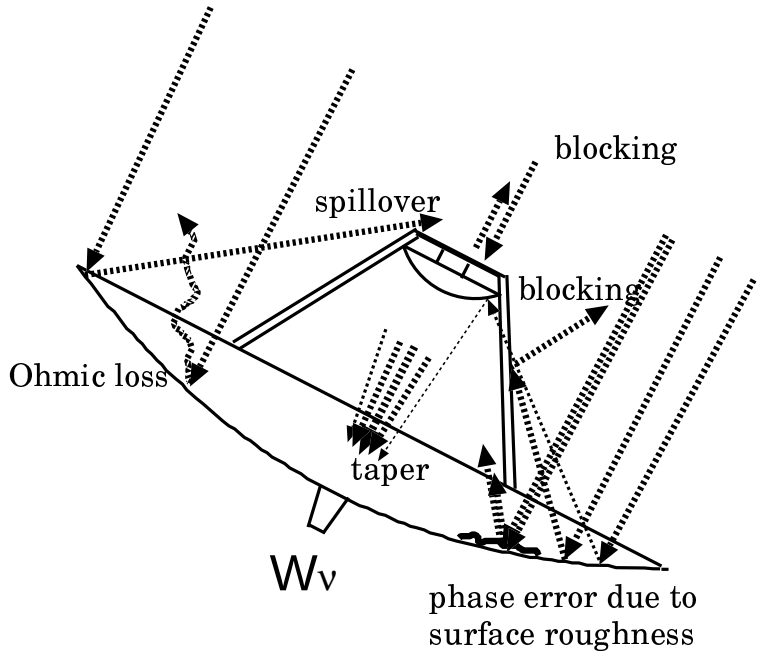

Figure 65: Factors reducing the effective aperture.

The actually received power in equation (181) is smaller than the incident power W_ν^i:

$$W_\nu^i = \frac{1}{2} A_g \mathcal{S}_\nu,$$

for reasons shown in Figure 65, where A_g is the geometrical aperture which is equal to $\pi D^2/4$ for a circular aperture with diameter D. Therefore, the

effective aperture is always smaller than the geometrical aperture.

$$A_e < A_g.$$

3.8 Aperture Efficiency η_A

The ratio of the effective aperture A_e to the geometrical aperture A_g is called aperture efficiency and is usually denoted by η_A.

$$\eta_A = \frac{A_e}{A_g}. \tag{182}$$

The aperture efficiency η_A is always smaller than 1 mainly due to

- the strong illumination taper (or gradation)
- the phase error due to the roughness of the antenna surface
- the blocking by the subreflector and the stays
- the spill over from the subreflector
- the Ohmic dissipation on the surface

as schematically illustrated in Figure 65.

3.9 Effective Aperture and Beam Solid Angle

The effective aperture A_e and the beam solid angle Ω_A are related with each other through the wavelength λ by a simple equation :

$$A_e \Omega_A = \lambda^2. \tag{183}$$

Let us prove the relation assuming a simple case of the thermodynamical equilibrium. Consider a system consisting of an antenna and a surrounding blackbody, and assume that the system is in a complete thermodynamical equilibrium with temperature T (Figure 66).

Since the intensity I_ν of the blackbody radiation in the Reileigh–Jeans approximation :

$$I_\nu = \frac{2\nu^2}{c^2} kT = \frac{2kT}{\lambda^2} \tag{184}$$

is essentially isotropic in this case (I_ν = const for any direction), the effective flux density incoming to the antenna can be expressed as

$$\mathcal{S}_\nu = \oint P_n(\theta, \phi)\, I_\nu \, d\Omega = I_\nu \, \Omega_A = \frac{2kT}{\lambda^2} \, \Omega_A. \tag{185}$$

Therefore, the power per unit frequency bandwidth W_ν received by the antenna is

$$W_\nu = \frac{1}{2} A_e \mathcal{S}_\nu = A_e \frac{kT}{\lambda^2} \Omega_A. \tag{186}$$

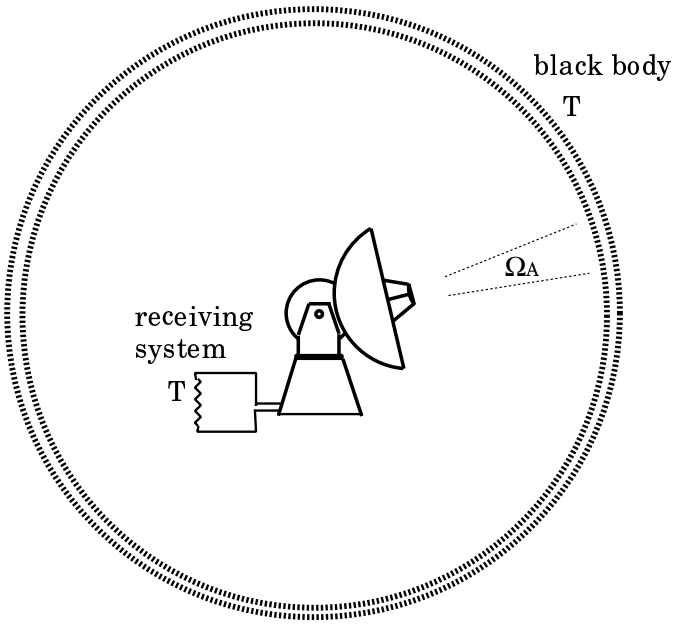

Figure 66: An antenna in a blackbody cavity.

On the other hand, the power L_ν transmitted by the antenna per unit frequency bandwidth due to the noise in the antenna receiving system, which is in the equilibrium with the temperature T, is equal to

$$L_\nu = kT, \tag{187}$$

according to the Nyquist theorem (H. Nyquist, Phys. Rev., **32**, 110–113, 1928). Now, the received power must equal the transmitted power in the detailed balance of the thermodynamical equilibrium. Therefore,

$$L_\nu = W_\nu \quad \longrightarrow \quad \frac{A_e \Omega_A}{\lambda^2} = 1 \quad \longrightarrow \quad A_e \Omega_A = \lambda^2. \tag{188}$$

Although we derived equation (183) in the thermodynamical equilibrium case, the resultant equation does not contain any thermodynamical quantity. Therefore, the equation must generally hold in other cases as well.

3.10 Directivity $\mathcal{D}$ and Aperture Efficiency η_A

As we saw earlier, the directivity (or gain) $\mathcal{D}$ is related to the beam solid angle Ω_A by equation (179) :

$$\mathcal{D} = \frac{4\pi}{\Omega_A}.$$

Now using equation (183) ($A_e\Omega_A = \lambda^2$), we can transform the above equation to

$$\mathcal{D} = 4\pi\frac{A_e}{\lambda^2} = 4\pi\eta_A\frac{A_g}{\lambda^2}, \tag{189}$$

where A_g and λ are the geometrical aperture and the wave length, respectively.

For a circular aperture antenna with diameter D, we have

$$\mathcal{D} = \eta_A\left(\frac{\pi D}{\lambda}\right)^2. \tag{190}$$

If the aperture efficiency η_A is kept nearly constant, we could obtain higher gain with larger diameter and higher frequency. In actuality, however, it becomes increasingly difficult to realize the high aperture efficiency as we go to the higher frequency and/or to the larger diameter.

In performance descriptions of antennas, the gain of an idealized antenna with 100 % aperture efficiency:

$$\left(\frac{\pi D}{\lambda}\right)^2, \tag{191}$$

is often described as '100 % gain', and usually given in dB. The 100 % gain can be directly calculated from the aperture diameter D and the observing wavelength λ. Therefore, if we can estimate the aperture efficiency η_A, we can easily get a value for the gain $\mathcal{D}$, as we will examine later.

3.11 Illumination Taper and Aperture Efficiency

Rather strong illumination taper is required in radio telescope antennas to avoid undesired ground pickups due to the sidelobes or the spill over from the main reflector. The strong taper lowers the sidelobe level. At the same time, however, it increases the beam width as we saw in Figure 58. The increase of the beam width and therefore of the beam solid angle Ω_A decreases the effective aperture A_e as required from equation (183) :

$$A_e = \frac{\lambda^2}{\Omega_A}.$$

In general, the illumination taper improves the main beam efficiency η_M but deteriorates the aperture efficiency η_A. The illumination taper is often the largest factor reducing the aperture efficiency of radio telescope antennas to the level of 50 – 60 %.

3.12 Surface Roughness and Aperture Efficiency

The surface roughness is another important factor which determines the aperture efficiency of radio telescope antennas.

Suppose that a real main reflector deviates from an ideal paraboloid due to the roughness of its surface. Then, the rays, which originally formed a wave front of equal phase in a plane wave, and were reflected at different points on the main reflector, reach at the focus with different path lengths, and, therefore, with different phases, due to the surface roughness. Thus, the rays reached at the focus are summed up with different phases. Therefore, the maximum power pattern P_{max} at the center of the main lobe becomes somewhat lower than the one in the ideal paraboloid case, with perfectly smooth surface. On the other hand, the averaged power pattern P must not change by the surface roughness, since the total power must be conserved.

Therefore, the directivity $\mathcal{D} = P_{max}/P$, and the effective aperture $A_e = \mathcal{D}\lambda^2/(4\pi)$, are reduced, compared with the ideal case. Because of the same reason, the reduced power in the main lobe must be balanced by an increased power in the sidelobes.

Thus, the surface roughness deteriorates both the aperture efficiency η_A and the main beam efficiency η_M.

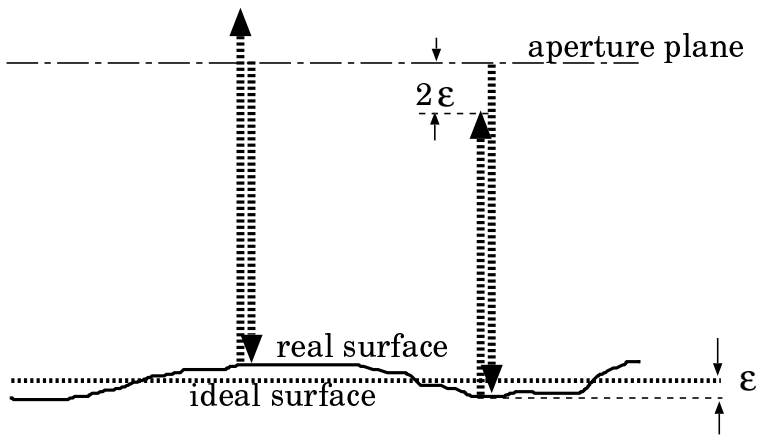

Figure 67: Phase error due to the surface roughness.

Let us calculate how the aperture efficiency is reduced due to the surface roughness of the main reflector, again in the transmission mode. For simplicity, we 'stretch' the actual paraboloid surface and treat it as a plane, parallel to the aperture plane, and regard that the rays are coming from a direction, perpendicular to the 'main–reflector plane', as shown in Figure 67.

This implies that a subreflector, and a primary feed, must be also stretched to a 'plane'. If we adopt such a simplified scheme, all rays transmitted from a primary feed reaches at the aperture plane with the same path length, and, therefore, with equal phase, if the 'main reflector plane' is ideally smooth, just as in the aperture plane of an ideally smooth paraboloid. Such an approximation is permissible in the present discussion, focused only on the effect of the randomly distributed surface roughness in the beam center direction.

Suppose that a ray is reflected at a point of the main reflector surface, at a distance ϵ apart from the ideally smooth surface, as shown in Figure 67. In this case, the path length of the ray differs from the one in the ideal case by 2ϵ, when it reaches at a point $\boldsymbol{r}'$ of the aperture plane. Therefore, we have a phase error of

$$\delta(\boldsymbol{r}') = 2k\epsilon = \frac{4\pi\epsilon}{\lambda}, \tag{192}$$

in the electric field $\boldsymbol{E}_a(\boldsymbol{r}')$ on the aperture plane, where k and λ are the wave number and wavelength, respectively, as before.

Let us assume that the phase error $\delta(\boldsymbol{r}')$ is randomly distributed over the aperture plane with the Gaussian probability density:

$$\mathfrak{f}(\delta) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{\delta^2}{2\sigma^2}}, \tag{193}$$

where $\sigma^2 = \langle \delta^2 \rangle$ is the dispersion of δ, with $\langle \ \ \rangle$ here denoting an ensemble average.

Now, the aperture illumination, taking into account the phase error, is expressed as

$$g(\boldsymbol{r}') = g_0(\boldsymbol{r}') e^{i\delta(\boldsymbol{r}')}, \tag{194}$$

where $g_0(\boldsymbol{r}')$ is an ideal aperture illumination, which would be realized with the perfectly smooth main reflector surface. $g_0(\boldsymbol{r}')$ is a real quantity, since we assume the oscillating electric field with the equal phase throughout the aperture plane, in the idealized case.

Then, the field pattern $f(\boldsymbol{n})$ around the beam center direction in the far–field region is expressed according to equation (150) as

$$f(\boldsymbol{n}) = \int_{\mathcal{A}} g_0(\boldsymbol{r}') e^{-i[2\pi \boldsymbol{n}\cdot\frac{\boldsymbol{r}'}{\lambda} - \delta(\boldsymbol{r}')]} \frac{d\mathcal{S}'}{\lambda^2}. \tag{195}$$

Now, we approximately replace the overall contribution of the phase error δ from the whole aperture plane, by an ensemble average $\langle e^{i\delta} \rangle$. Then, we have

$$\langle f(\boldsymbol{n}) \rangle = \int_{\mathcal{A}} g_0(\boldsymbol{r}') e^{-i2\pi \boldsymbol{n}\cdot\frac{\boldsymbol{r}'}{\lambda}} \langle e^{i\delta} \rangle \frac{d\mathcal{S}'}{\lambda^2}, \tag{196}$$

where $\langle e^{i\delta} \rangle$ can be described, by using the Gaussian probability density in equation (193),

$$\begin{aligned} \langle e^{i\delta} \rangle &= \int_{-\infty}^{\infty} \mathfrak{f}(\delta) e^{i\delta} d\delta = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} e^{i\delta - \frac{\delta^2}{2\sigma^2}} d\delta = e^{-\frac{\sigma^2}{2}} \\ &= e^{-\frac{\langle \delta^2 \rangle}{2}} = e^{-\frac{1}{2}\left(\frac{4\pi\sqrt{\langle \epsilon^2 \rangle}}{\lambda}\right)^2}, \end{aligned} \tag{197}$$

where we used a well known integration formula:

$$\int_{-\infty}^{\infty} e^{iax - x^2} dx = \sqrt{\pi} e^{-\frac{a^2}{4}}, \tag{198}$$

and equation (192). Thus, we obtain

$$\langle f(\boldsymbol{n}) \rangle = f_0(\boldsymbol{n}) e^{-\frac{1}{2}\left(\frac{4\pi\sqrt{\langle \epsilon^2 \rangle}}{\lambda}\right)^2}, \tag{199}$$

where

$$f_0(\boldsymbol{n}) = \int_{\mathcal{A}} g_0(\boldsymbol{r}') e^{-i2\pi \boldsymbol{n} \cdot \frac{\boldsymbol{r}'}{\lambda}} \frac{d\mathcal{S}'}{\lambda^2}, \tag{200}$$

is the ideal field pattern with the perfectly smoooth main reflector in the far field region.

Since the power pattern $P(\theta, \phi)$ in the far field region is proportional to $\mid f(\boldsymbol{n}) \mid^2$, we have

$$P(\theta, \phi) = P_0(\theta, \phi) e^{-\left(\frac{4\pi\sqrt{\langle \epsilon^2 \rangle}}{\lambda}\right)^2}, \tag{201}$$

around the beam center direction, where $P_0(\theta, \phi)$ is the ideal power pattern with the perfectly smooth main refletor. Hence, we have

$$P_{max} = P_{0\,max} e^{-\left(\frac{4\pi\sqrt{\langle \epsilon^2 \rangle}}{\lambda}\right)^2}. \tag{202}$$

Therefore, using equations (179), (189), and (182), we obtain following expressions of the directivity, the effective aperture and the aperture efficiency as functions of the root mean square surface roughness $\sqrt{\langle \epsilon^2 \rangle}$:

$$\mathcal{D} = \mathcal{D}_0 e^{-\left(\frac{4\pi\sqrt{\langle \epsilon^2 \rangle}}{\lambda}\right)^2}, \tag{203}$$

$$A_e = A_{e0} e^{-\left(\frac{4\pi\sqrt{\langle \epsilon^2 \rangle}}{\lambda}\right)^2}, \tag{204}$$

$$\eta_A = \eta_{A0} e^{-\left(\frac{4\pi\sqrt{\langle \epsilon^2 \rangle}}{\lambda}\right)^2}, \tag{205}$$

where the symbols with suffix 0, $\mathcal{D}_0 = P_{0\,max}/P$, $A_{e0} = \mathcal{D}_0\lambda^2/(4\pi)$, and $\eta_{A0} = A_{e0}/A_g$, represent the values for the idealized case when the main reflector surface is perfectly smooth. For a particular case of the circular aperture, the directivity or the gain is expressed in a form:

$$\mathcal{D} = \eta_{A0}\left(\frac{\pi D}{\lambda}\right)^2 e^{-\left(\frac{4\pi\sqrt{\langle\epsilon^2\rangle}}{\lambda}\right)^2}. \tag{206}$$

We derived the above relations for the transmission case. But the results are readily applicable to the reception case as well, in view of the equivalence of the antenna beam patterns in transmission and reception.

The gain curves as functions of wavelength are shown for some existing radio telescopes in Figure 68.

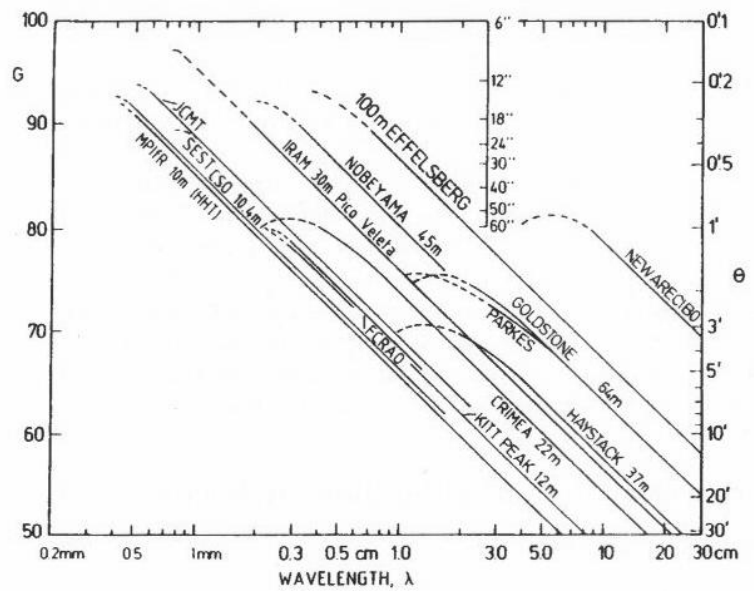

Figure 68: Gain in dB and HPBW of some existing radio telescope antennas as functions of wavelength (from Rohlfs and Wilson, 2000, Tools of Radio Astronomy, 3rd edition, Springer).

3.13 Surface Accuracy and Lower Limit of the Observing Wavelength

A normalized gain curve as a function of $x = \sqrt{\langle\epsilon^2\rangle}/\lambda$:

$$\frac{e^{-(4\pi x)^2}}{x^2}$$

is shown in Figure 69 in the linear scale (in this figure, the rms surface roughness $\sqrt{\langle\epsilon^2\rangle}$ is denoted as ϵ for simplicity). It is evident from the figure

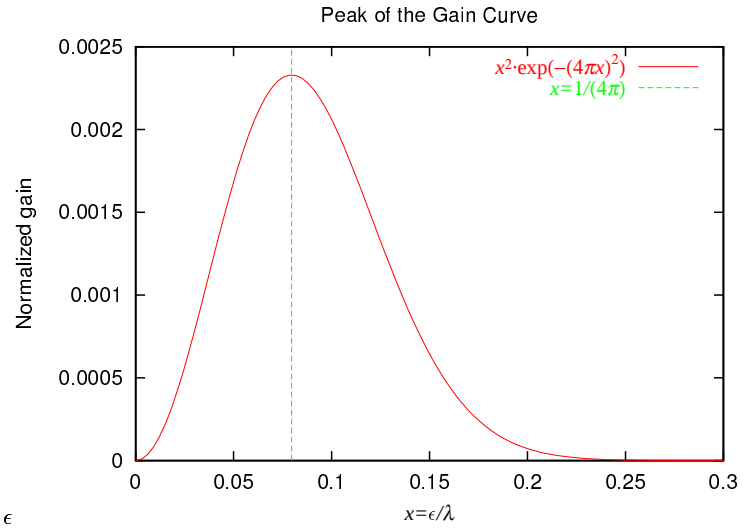

Figure 69: Normalized gain curve and its peak position.

that the maximum gain is obtained at the wavelength :

$$\lambda_{peak} = 4\pi\sqrt{\langle\epsilon^2\rangle}$$

or at the frequency :

$$\nu_{peak} = \frac{c}{4\pi\sqrt{\langle\epsilon^2\rangle}},$$

depending on the rms surface roughness, or otherwise called "surface accuracy", $\sqrt{\langle\epsilon^2\rangle}$.

It may appear from the above discussion that we can use an antenna with the surface accuracy $\sqrt{\langle\epsilon^2\rangle}$ at short wavelength down to around $4\pi\sqrt{\langle\epsilon^2\rangle} \simeq 12.6\sqrt{\langle\epsilon^2\rangle}$. But here we must be a little cautious.

What we actually receive from a radio source is the power :

$$W_\nu = \frac{1}{2}A_e\mathcal{S}_\nu = \frac{1}{2}\eta_A A_g\mathcal{S}_\nu$$

(see equation (181)), and equation (189) shows that

$$\mathcal{D} = 4\pi\frac{A_e}{\lambda^2}, \quad \text{and hence,} \quad A_e = \frac{\mathcal{D}\lambda^2}{4\pi}. \tag{207}$$

Therefore, the received power is

$$W_\nu = \frac{1}{8\pi}\mathcal{D}\lambda^2\mathcal{S}_\nu. \tag{208}$$

Consequently, the wavelength dependence of the power is uniquely determined, i.e. without further dependence on the wavelength, by the wavelength dependence of the gain:

$$\mathcal{D} = 4\pi\eta_A\frac{A_g}{\lambda^2}$$

(see equation (189)), only when the flux density $\mathcal{S}_\nu$ is proportional to λ^{-2}, just like in the case of a thermal blackbody radiation source with temperature T and solid angle Ω_s in the Rayleigh–Jeans region :

$$\mathcal{S}_\nu = \frac{2\nu^2}{c^2}kT\Omega_s = \frac{2kT}{\lambda^2}\Omega_s.$$

Therefore, the wavelength dependence of the gain, as shown in Figure 68, is really a good measure of the antenna sensitivity, when we observe thermal radio sources with the spectral index, which we defined as $\mathcal{S}_\nu \propto \nu^{-\alpha} \propto \lambda^\alpha$, being -2.

For non–thermal radio sources, such as Synchrotron radiation sources, however, the spectral index α is not equal to -2, but usually positive. Therefore, the wavelength dependence of the received power W_ν is no longer uniquely determined by the gain $\mathcal{D}$ only.

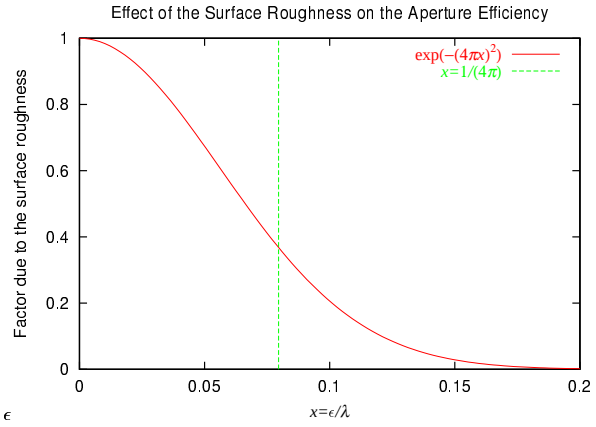

Figure 70: Dependence of the aperture efficiency on the ratio of the surface accuracy ϵ to the wavelength λ.

Consequently, it is more appropriate to use the wavelength dependence of the aperture efficiency η_A shown in equation (205) as a measure of the

sensitivity of an antenna with a given diameter and a given surface accuracy as illustrated in Figure 70 (here again the surface accuracy $\sqrt{\langle\epsilon^2\rangle}$ is simply denoted as ϵ). As we see from the figure, the aperture efficiency is reduced by a factor of around 0.37 when the wavelength $\lambda \simeq 4\pi\sqrt{\langle\epsilon^2\rangle}$. Although we could still use an antenna with aperture efficiency $\leq 37\%$, and we actually do in some cases, it is usually accepted as a reasonable lower limit of the observing wavelength

$$\lambda_{min} \simeq 20\sqrt{\langle\epsilon^2\rangle}, \tag{209}$$

where the reduction factor is around 70 %. This means that the surface accuracy specificatin of 150 μm rms, shown in table 3, is good enough for an observing wavelength as short as 3 mm (or, as high as 100 GHz in frequency).

The surface accuracy specification is often the most decisive factor in the cost estimation of an antenna, since tolerances in manifacturing of various parts of the antenna, including supporting structures and materials for main reflector panels, are largely determined by the required surface accuracy. Therefore, the surface accuracy is carefully specified under certain environmental conditions, such as, 'at night time, under wind speed less than 10 m/s, and at elevation 45 degrees'. For high frequency radio telescopes, astronomers make regular efforts to improve the surface accuracy by precisely adjusting surface panels.

An example of the gain calculation, taking into account the surface accuracy, as well as illumination taper, blocking, and spill over effects, is given in table 4. One can easily confirm values of 100 % gain by equation (191), and values of surface accuracy effects for main reflector and subreflector by equation (205) (though we derived equation (205) for a main reflector, the same discussion is applicable for a subreflector as well).

3.14 Pointing Accuracy

The main lobe axis of a radio telescope antenna may not be exactly oriented towards a required direction due to the mechanical inaccuracy and the deformations due to the gravity and the wind pressure. The angular offset ϵ_θ between the actual and required directions is called "pointing error" or "tracking error" (Figure 71). The rms value of the pointing error is often called as "pointing accuracy" which we denote here as σ_θ

$$\sigma_\theta = \sqrt{< \epsilon_\theta^2 >}. \tag{210}$$

Let us consider the loss of the received power from a radio source due to the pointing error. For simplicity, we assume a Gaussian power pattern $P(\theta)$

Items		Frequency in GHz				Remarks
		2.2	8.2	22.2	43.1	
Illumination taper		-2.39	-0.92	-1.30	-1.46	
Blocking	Subref.	-0.16	-0.16	-0.16	-0.16	
	Stay	-0.35	-0.35	-0.35	-0.35	
Spill over	Main ref.	-0.27	-0.07	-0.01	-0.01	
	Subref.	-0.97	-0.51	-0.18	-0.09	
Surface accuracy	Main ref.	-0.01	-0.13	-0.94	-3.54	0.5mm rms
	Subref.	neg.	-0.02	-0.15	-0.57	0.2mm rms
	Setting	neg.	-0.04	-0.18	-0.66	Subref–feed
Horn cover loss		-0.07	-0.06	-0.11	-0.24	
Aperture efficiency	in dB	-4.24	-2.26	-3.38	-7.08	
	in %	37.6%	59.4%	45.9%	19.6%	
100 % gain in dB		47.2	58.6	67.3	73.1	
Gain in dB		43.0	56.4	63.9	66.0	
Feed & VSWR loss		-2.0	(NAO)	(NAO)	(NAO)	
Total gain in dB		41.0	(56.4)	(63.9)	(66.0)	
Total efficiency in %		23.8%	(59.4%)	(45.9%)	(19.6%)	
Nominal gain in dB		40.7	56.1	63.6	65.7	at horn neck

Table 4: An example of the gain calculation for a 10 m diameter antenna. Numerical values are given in dB, unless otherwise stated.

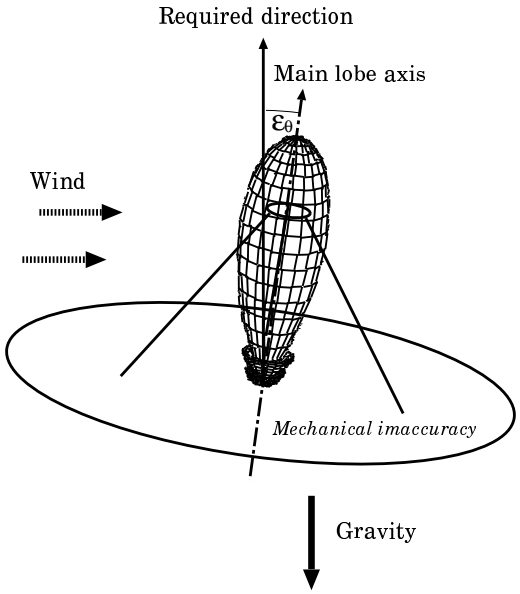

Figure 71: Pointing error.

with the peak power P_0 and the HPBW Θ :

$$P(\theta) = P_0 e^{-ln2\left(\frac{2\theta}{\Theta}\right)^2} \tag{211}$$

If we assume that the main lobe axis is offset from the required direction in average by σ_θ, the received power from the radio source at the required direction is reduced in average by a factor of

$$\frac{P(\sigma_\theta)}{P_0} = e^{-4ln2\left(\frac{\sigma_\theta}{\Theta}\right)^2} = e^{-2.7726\left(\frac{\sigma_\theta}{\Theta}\right)^2}. \tag{212}$$

Figure 73 shows the power loss calculated by equation (212) as a function of the pointing accuracy σ_θ (which is denoted as σ in Figure 73 for simplicity).

It is evident from the figure that the pointing accuracy must be

$$\sigma_\theta \leq \frac{1}{10}\Theta \tag{213}$$

to keep the power loss smaller than several %. Table 5 shows the pointing accuracy requirement given in equation (213) for some typical cases of antenna diameter D and observing wavelength λ on the basis of the approximate formula $\Theta \simeq \lambda/D$.

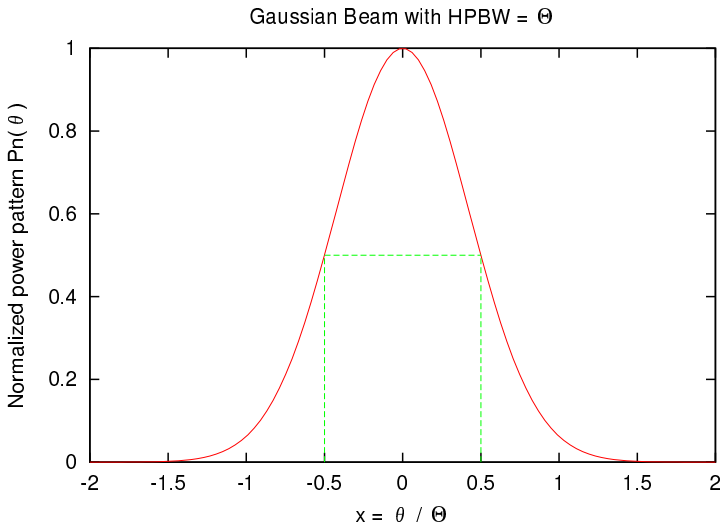

Figure 72: Gaussian beam.

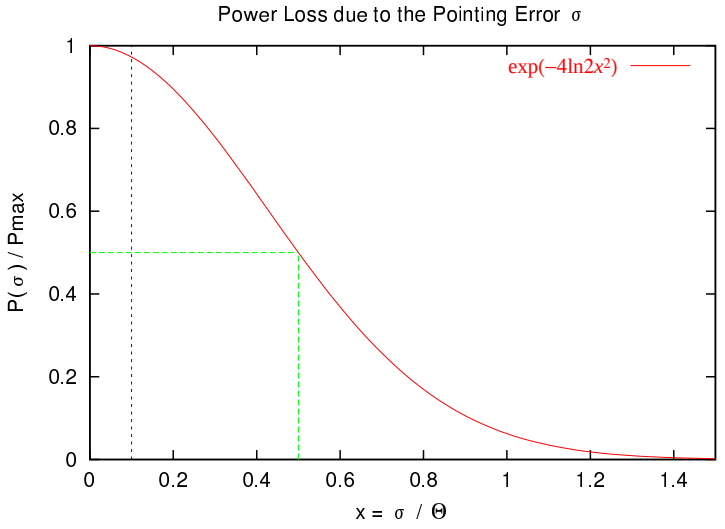

Figure 73: Power loss due to the pointing error.

D	10 m	20 m	45 m
λ	7 mm	2 mm	1.5 mm
$\sigma_{\theta\, max}$	0.°004	0.°00057	0.°00019

Table 5: Requirements to the pointing accuracy.

Needless to say, a state–of–the–art technology of mechanical construction and good enough materials are required to precisely keep the orientaion of a huge radio telescope antenna, with the accuracy of 0.001 degree or higher, under variable wind and heat conditions. The pointing accuracy is another cost–impacting factor, along with the surface accuracy.

3.15 Design of the Feed System

In order to efficiently receive or transmit the radio signal of wave length λ with a paraboloidal antenna of Cassegrain design, the diameter D_S of the subreflector, the diameter of the feed horn D_F and the distance L between the feed horn and the subreflector must satisfy a relation for avoiding a significant spill over from the edge of the subreflector :

$$\frac{\lambda}{D_F} < \frac{D_S}{L} \tag{214}$$

as illustrated in Figure 74 (note that the subreflector is located in the near field, or in the Fresnel region, of the feed horn, and therefore the beam pattern of the feed horn is more complicated than the one in the far field, i.e. in the Fraunhofer region, where a sharp main lobe with HPBW $\simeq \lambda/D_F$ is dominating).

Suppose that we use a paraboloidal antenna at several observing wave lengths. If we denote the longest observing wavelength as λ_{max} and the diameter of the feed horn for the longest wave length as $D_{F\,max}$, the product $D_S D_{F\,max}$ must be well larger than $\lambda_{max} L$ according to equation (214). In order not to make either of D_F and D_S too large, it is natural to select that $D_S \simeq D_{F\,max}$. Thus the equation (214) is reduced to $\lambda_{max} L < D_{F\,max}^2$. In actuality, a more stringent condition:

$$\frac{D_{F\,max}^2}{\lambda_{max} L} \geq 10 \tag{215}$$

is usually considered as an optimal condition for the size $D_{F\,max}$ of the feed horn and the subreflector for the longest wavelength. For example, if we

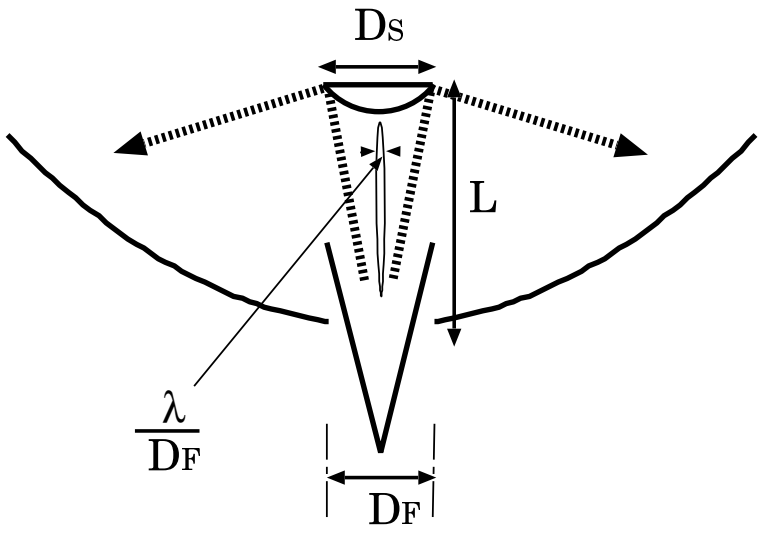

Figure 74: Diameters of the feed horn D_F and the subreflector D_S and their mutual distance L.

adopt this condition, and assume $L = 10$ m for definiteness, we have a following table of feed horn sizes for several cases of the longest wavelength.

λ_{max}	$D_{F\,max}$
3 mm (100 GHz)	0.55 m
7 mm (43 GHz)	0.84 m
1.4 cm (22 GHz)	1.2 m
3.8 cm (8 GHz)	2.0 m
15 cm (2 GHz)	3.9 m
21 cm (1.4 GHz)	4.6 m

Table 6: Maximum observing wavelengths vs. diameters of the feed horns.

It is evident from the table 6 that at an observing frequency as low as 1.4 GHz or 2 GHz the sizes of the feed horns and sureflectors must be fairly large. In practice, feed horn sizes are often designed to be smaller than subreflector sizes, and the optimal condition in equation (215) is significantly released, in order to place feed horns in a finite volume of the receiver cabin, by sacrificing

a part of the aperture efficiency. Even though, it is difficult to cover too wide frequency range within a simple Cassegrain design only.

An alternative way to receive the low frequency radio waves without too much increasing the feed–horn and subreflector sizes is to use the primary focus (Figure 75). In the primary focus case, a feed horn must 'illuminate' the whole area of a large main reflector. Therefore, a wide beam, and therefore a small diameter of the feed horn, is rather preferable.

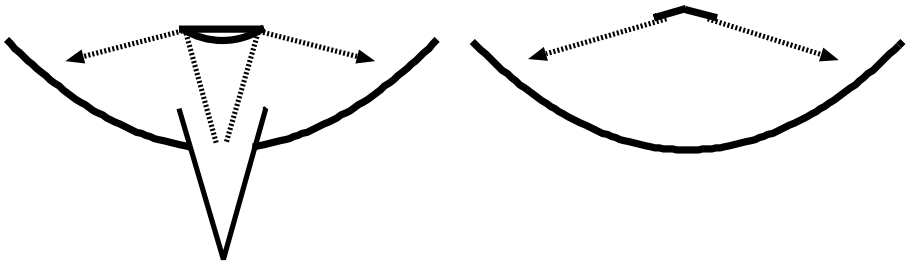

Figure 75: A large feed horn and subreflector system (left) or a primary feed system (right) is needed for the low frequency observations.

The large feed horns and subreflectors, or the feed and receiver blocks on the primary focus, required for the low frequency observations, might cause serious difficulties for achieving the high aperture efficiency and the high pointing accuracy in the high frequency observations. Nevertheless, it is a natural desire of radio astronomers to observe at as many frequency bands as possible, for example, from the 1.4 GHz line of the neutral hydrogen to the 115 GHz CO line, say, with a single radio telescope. Hence some sophisticated solutions are required (see examples in Figures 76 and 77).

3.16 Other Characteristics

3.16.1 Range of Motion

For an Alt–Azimuth mount antenna, the range of azimuth and elevation angles must cover at least from $-180°$ to $+180°$ and from $0°$ to $90°$, respectively, in order to be able to point toward any direction in the sky. However, we have to also take into account that a radio source may cross the meridian line during an observation, from east to west, or west to east, in case of circumpolar sources, depending on their locations with respect to the celestial pole. In order to smoothly track such diurnal motions of radio sources, the azimuth motion range is usually chosen from $-270°$ to $+270°$ with $0°$ selected

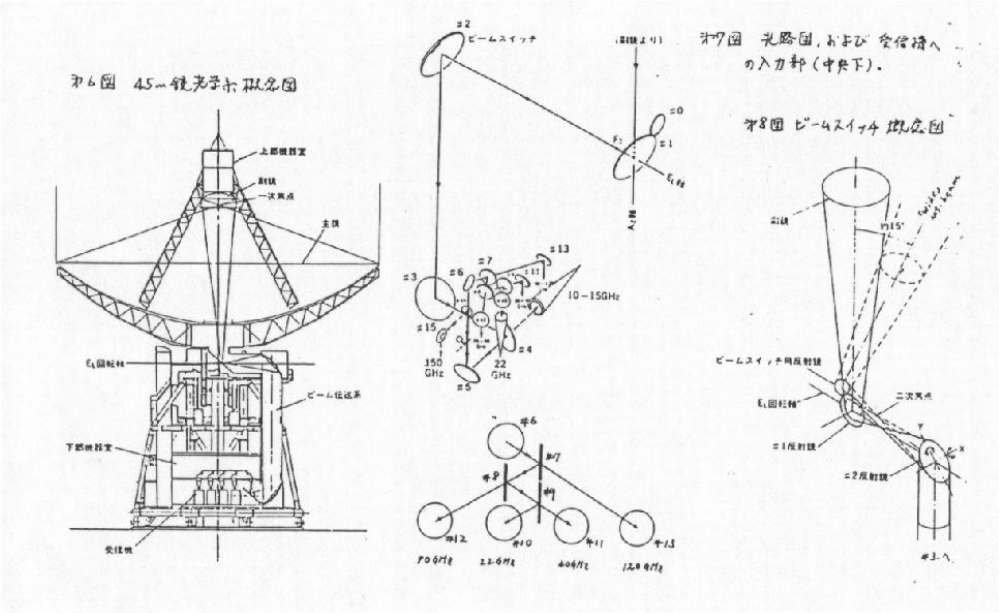

Figure 76: The initial design of the wave–beam–guide feed system of Nobeyama 45 m mm–wave telescope.

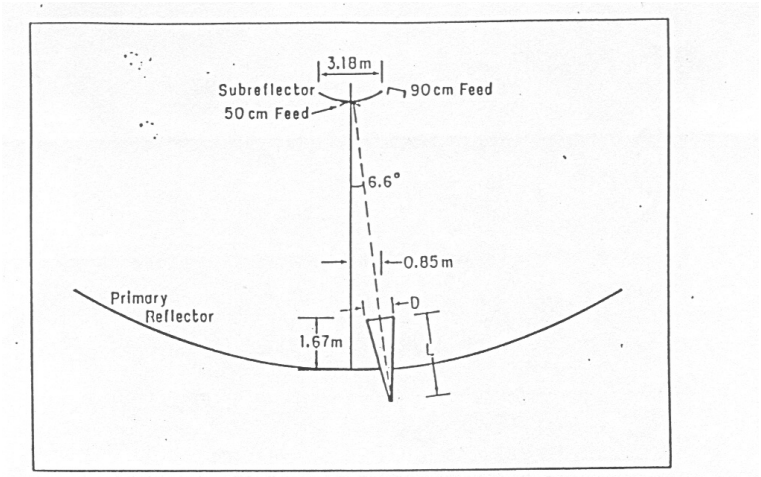

Fig. IV-2a. Proposed Cassegrain geometry, the length, L, and opening, D, of each feed horn is given in Table IV-2.

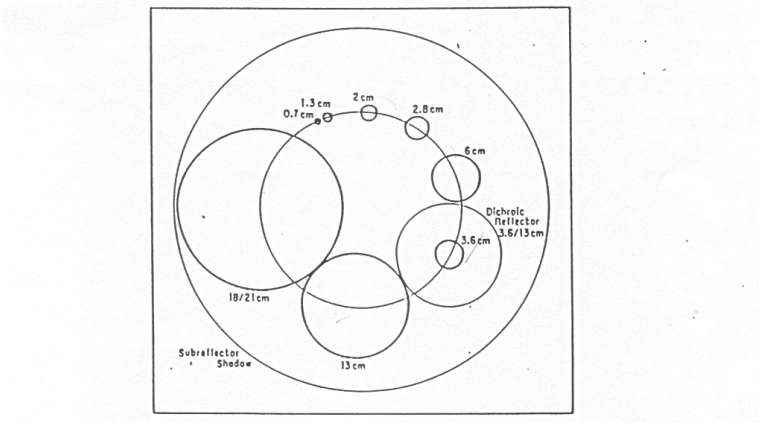

Figure 77: The tilted subreflector and feed horns arranged on a circle in 25 m radio telescopes of VLBA (Very Large Baseline Array).

at South (in the conventional azimuth angle, measured eastward from North, the same range corresponds to -90° – $+450^\circ$).

3.16.2 Slewing Speed

For VLBI antennas, especially for geodetic VLBI antennas, high slewing speeds both in azimuth and elevation are preferable, since the high slewing speeds allow us to effectively observe many sources on different points of sky within a limited telescope time, without losing much time for antenna repointing from one source to another. Generally, slew speeds larger than 2°/sec in both axes are desirable.

In recent VLBI observations, a technique called 'fast switching' is often used. In this technique, orientation of a telescope is switched between two closely spaced radio sources (typically, separated by less than a few degrees) within a period of time as short as 30 to 40 seconds, for compensating irregular atmospheric refraction effects. For the fast switching, not only the slewing speeds, but also their accelerations must be sufficiently high, to realize the short switching cycle. For example, 3°/sec slewing speeds and $3^\circ/\mathrm{sec}^2$ accelerations in both axes are desirable to achieve 20 second switching cycle for sources separated by 2 to 3 degrees.

3.16.3 Operational and Survival Loads

Operational and survival loads characterize environmental conditions (temperature, wind speed, etc.) under which antennas can be normally used for observations, and can survive in a stable stow position at rest, respectively. Operational loads must be determined depending on the actual conditions of the observing sites. For mid–latitude stations with moderate wind, usually operational temperature condition of around -20 – 40 °C and operational maximum wind speed condition of $7 \sim 10$ m/sec are appropriate. Survival wind load condition could be around 90 m/sec or 60 m/sec, depending whether there are strong typhoons (cyclons, or hurricanes) or not, in the station area.

3.16.4 Axes Offset and Axes Orthogonality

For both Alt–Azimuthal and Equatorial mount antennas, it is desirable that two axes (Az– and El–axes, or Right Ascension– and Declination–axes) exactly intersect at a point, and are mutually orthogonal. After actual mecahnical construction of antennas, the two axes usually show some finite offsets and non–orthogonalities. Tolerances to these quantities are often required to be less than a few mm and 0.01°, respectively.

Even though such requirements are satisfied, they may be still large enough to affect observational performances. The axes offsets may affect high–precision geodetic VLBI measurements of antenna positions, and the axes non–orthogonalities may affect antenna pointings. Therefore, we must estimate and calibrate the offsets and non–orthogonalities, based on the geodetic VLBI measurements and antenna pointing measurements.

4 Antenna Temperature and Single Dish Imaging

4.1 What Is the Antenna Temperature T_A?

"Antenna temperature" is an important but again a little confusing quantity often used in radio astronomy, as well as in VLBI geodesy, to describe the power received by an antenna from a radio source. Actually, the antenna temperature is nothing but the power W_ν per unit frequency bandwidth, centered at ν, received by an antenna, which is regarded as a noise power added to a detector, and converted to the "temperature of the noise" T_A according to the Nyquist theorem :

$$W_\nu = kT_A. \tag{216}$$

Therefore, the amtenna temperature is the "temperature sensed by an antenna" and has nothing common with the temperature of the antenna body which could be measured by a thermometer.

If we denote the effective aperture of the antenna as A_e and the effective flux density of the source as $\mathcal{S}_\nu$, then W_ν is expressed by equation (181) :

$$W_\nu = \frac{1}{2} A_e \mathcal{S}_\nu.$$

If the normalized power pattern of the antenna is $P_n(\theta, \phi)$ and the source intensity distribution is $I_\nu(\theta, \phi)$ in the antenna–fixed coordinate system, the above equation is transformed to

$$W_\nu = \frac{1}{2} A_e \mathcal{S}_\nu = \frac{1}{2} A_e \iint_\Omega P_n(\theta, \phi)\, I_\nu(\theta, \phi) \sin\theta \, d\theta \, d\phi,$$

in view of the equation (174). We can formally extend the above range of the integration to the whole sky, since $P_n(\theta, \phi) \simeq 0$ outside the narrow beam

anyway. Thus, using the solid angle element $d\Omega = \sin\theta d\theta d\phi$, we express the antenna temperature in equation (216) as

$$T_A = \frac{1}{2k} A_e \oint P_n(\theta, \phi)\, I_\nu(\theta, \phi)\, d\Omega, \tag{217}$$

(see Figure 78).

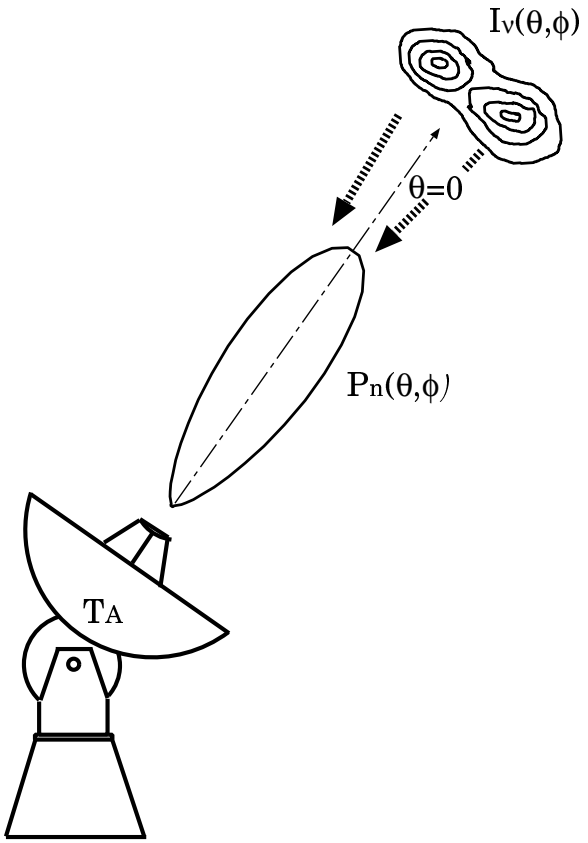

Figure 78: Antenna temperature.

Taking into account the equation (183) which relates the effective aperture A_e, the beam solid angle Ω_A and the wavelength λ

$$A_e \Omega_A = \lambda^2,$$

and the relation

$$\Omega_A = \oint P_n(\theta, \phi) d\Omega,$$

we can transform the equation (217) into

$$T_A = \frac{\lambda^2}{2k} \frac{\oint P_n(\theta,\phi)\, I_\nu(\theta,\phi)\, d\Omega}{\oint P_n(\theta,\phi) d\Omega}. \tag{218}$$

Furthermore, if we express the intensity I_ν through the brightness temperature T_B :

$$I_\nu = \frac{2k}{\lambda^2} T_B,$$

we obtain

$$T_A = \frac{\oint P_n(\theta,\phi)\, T_B(\theta,\phi)\, d\Omega}{\oint P_n(\theta,\phi) d\Omega}. \tag{219}$$

Therefore, the antenna temperature is the brightness temperature averaged over the antenna beam with the weighting function $P_n(\theta,\phi)$, which is the normalized power pattern.

4.2 Imaging with the Single Dish Radio Telescope

Based on the above discussions, we can interpret that the single dish radio telescope is a huge power meter measuring the weighted mean brightness temperature at a certain direction of the sky with the antenna power pattern as the weight. The single dish radio telescope can draw the image of the radio source by scanning from point to point meshed on the sky measuring the distribution of the brightness (intensity) averaged over the beam. It is thus naturally understood that the angular resolution of the image obtained by the single dish radio telescope is determined by the beam size (Figure 79).

Let us examine three simplest cases of the source brightness distribution.

1. Point source $I_\nu(\boldsymbol{s}) = \mathcal{S}_\nu \delta(\boldsymbol{s} - \boldsymbol{s}_0)$, where $\mathcal{S}_\nu$ is the source flux density, while $\boldsymbol{s}$ and $\boldsymbol{s}_0$ are unit vectors, and $\boldsymbol{s}_0$ is parallel to the beam axis. In this case we have

$$T_A = \frac{1}{2k} A_e \oint P_n(\boldsymbol{s})\, \mathcal{S}_\nu\, \delta(\boldsymbol{s} - \boldsymbol{s}_0)\, d\Omega = \frac{1}{2k} A_e\, \mathcal{S}_\nu. \tag{220}$$

   We can measure the effective aperture A_e or the aperture efficiency η_A by observing a point source of known flux density on the basis of this equation.

2. Source of finite extent which slightly exceeds the main lobe :

$$T_A = T_B \frac{\int\limits_{main\ lobe} P_n(\theta,\phi) d\Omega}{\oint P_n(\theta,\phi) d\Omega} = \frac{\Omega_M}{\Omega_A} T_B = \eta_M T_B. \tag{221}$$

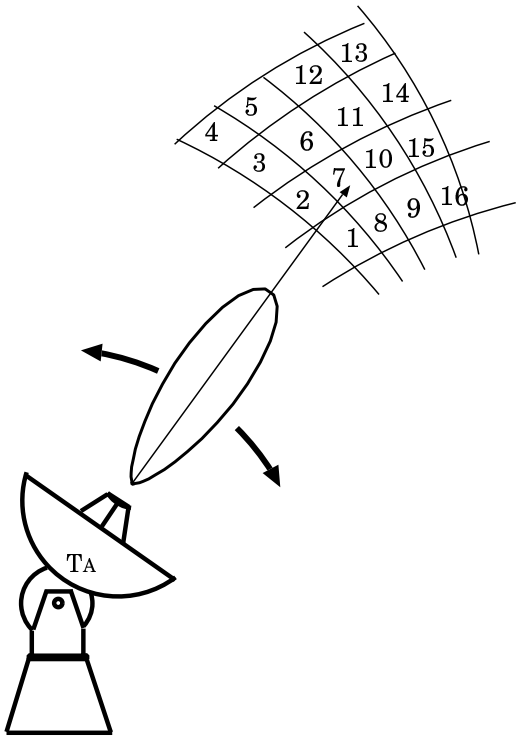

Figure 79: Scanning observation for radio source imaging with a single dish radio telescope.

We can measure the main beam efficiency η_M by observing this kind of source with known brightness temperature (for example, a planet) on the basis of the above equation.

3. Widely extended homogeneous source :

$$T_A = T_B. \tag{222}$$

For example, we always receive the antenna temperature of 2.7 K from the cosmic backgroud radiation.

Usually, the antenna temperature added by an astronomical radio source is very low, typically ranging from milli–Kelvins to Kelvins depending on the source and the telescope. For example, if we receive a relatively strong radio source with the effective flux density of 1 Jy ($= 10^{-26}$ W m^{-2} Hz^{-1}) using an 20 m diameter antenna with the aperture efficiency of 60 %, the antenna temperature is only about

$$T_A = \frac{1}{2k} A_e \mathcal{S}_\nu = \frac{1}{8k} \eta_A \pi D^2 \mathcal{S}_\nu \simeq 0.068 \text{ K}.$$

This is of course much lower than the physical temperature of an antenna which is always $\simeq 300$ K.

Q : Why can't a radio telescope give us an image of the radio source just by a single observation like any optical telescope does?
A : Unlike the photo plate or CCD plate, a radio telescope is usually equipped with only one detecting element, the receiver. So what you can sense is whether the received power is strong or weak in the beam direction. If you have many receivers on the focal plane, then you will get the two–dimensional image at once. Now so–called multi–beam receiving systems are intensively developed at many radio telescope observatories of the world for that purpose.

5 Receiving Systems

The very weak high frequency radio signal from an astronomical radio source collected by the main reflector is amplified and frequency converted by the receiving system and sent to the signal processing unit as the tractable signal of low frequency and with enough strength.

At the same time, the receiving system inevitably adds some noise to the received signal, which may make it difficult to extract scientifically useful information with sufficiently high signal–to–noise ratio. Also, it is difficult

to keep the ultra–high amplification ratio of the receiving system very stable in time. The gain variation effect must be well removed from the observed data for any measurement of the received power from a radio source.

So, the topics here will cover the system noise temperature, signal–to–noise ratio, frequency conversion and removal of the gain–variation effect.

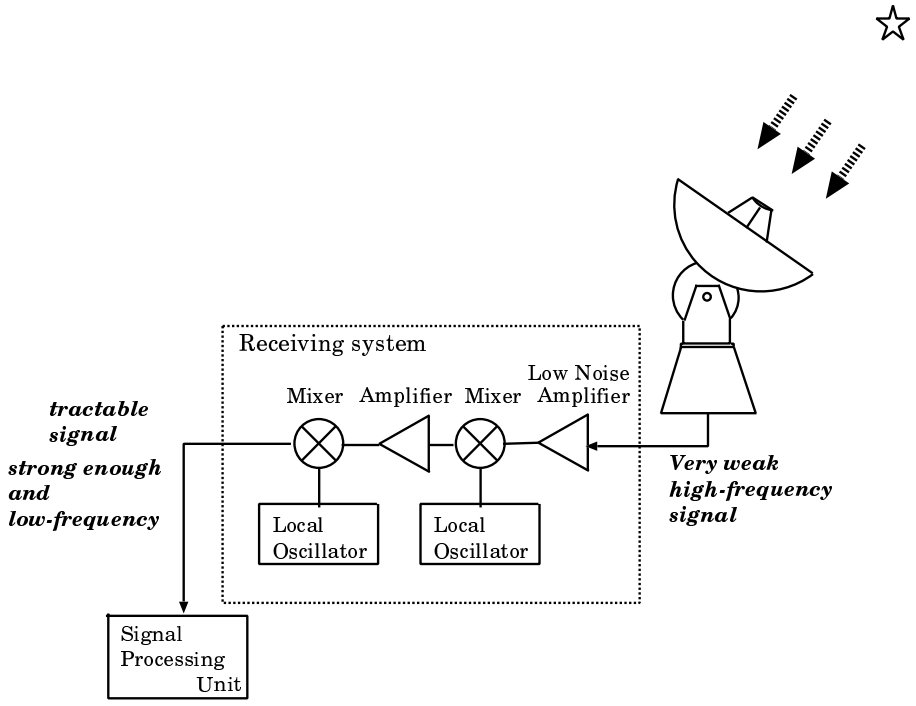

Figure 80: A schematic view of a radio telescope receiving system.

5.1 System Noise Temperature

The "input equivalent noise"

The word "system noise" means the "input–equivalent" of the sum of all noises generated in the receiving system (in the amplifiers, attenuators, connectors, ...) as well as in the ground pickups due to the sidelobes and the spill overs and also in the thermal radiation of the atmosphere. The "input equivalent" implies here to treat something occuring within and/or outside the system as all coming from the outside of the system as schematically illustrated in Figure 81. If our system has a total gain G and if the noise power per unit frequency bandwidth at the output of our system is $W_{N\nu}$,

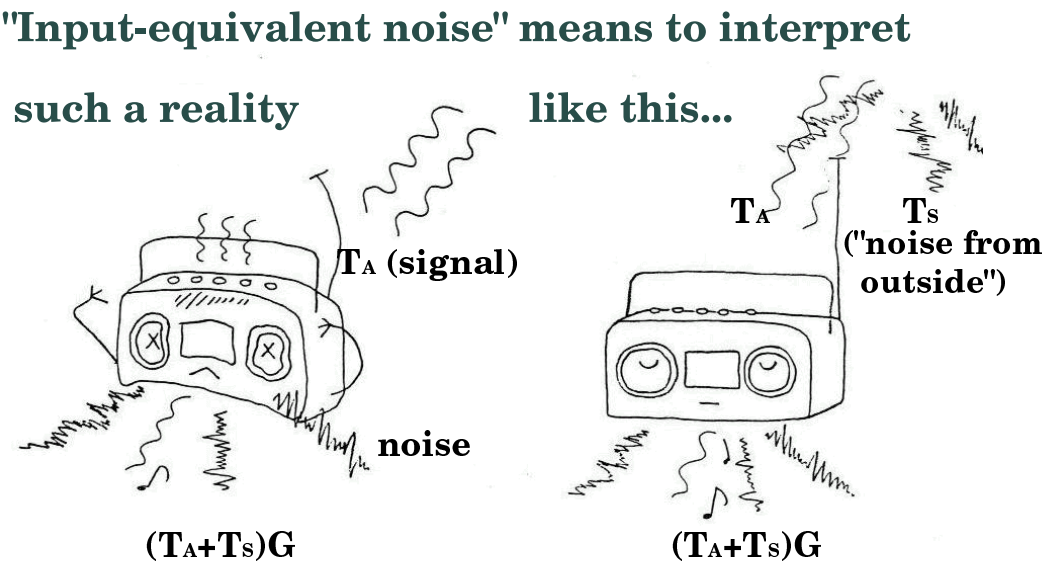

Figure 81: What is the "input equivalent noise"?

then the power $W_{S\nu}$ of the input–equivalent system noise is

$$W_{S\nu} = \frac{W_{N\nu}}{G}.$$

The "system noise temperature" T_S is a temperature representing the power $W_{S\nu}$ of the system noise in terms of the Nyquist theorem $W_{S\nu} = kT_S$.

Signal attenuation and thermal noise

Let us consider how the thermal noise emerges within the receiving system using a model based on the radiative transfer equation, since it is most likely that the absorption and emission of the radio wave occur in a similar way both in the wave guide systems of the receivers and in the interstellar clouds.

Let us apply the radiative transfer equation :

$$\frac{dI_\nu}{dl} = -\kappa_\nu I_\nu + \epsilon_\nu \tag{223}$$

to some "lossy piece" of our receiving system with temperature T (Figure 82). For simplicity, we ignore the amplification for a while and consider the effect of the attenuation only. In equation (223), I_ν again stands for the intensity of the radio wave, l is the length along the optical path, and κ_ν and ϵ_ν are the opacity (or absorption coefficient) and the emissivity, respectively.

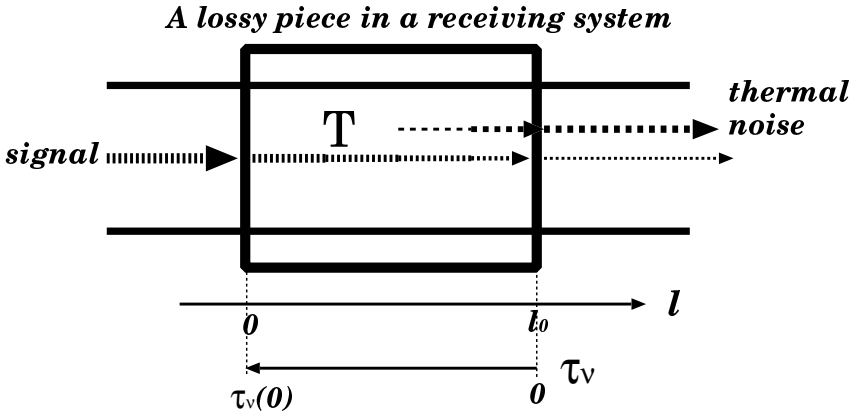

Figure 82: Signal attenuation and thermal noise emission in a lossy piece of the receiving system.

Since it is likely that every element of the lossy piece, for example attenuator, wave guide tube, connector, etc., is in the thermodynamical eqilibrium with each other, giving rise to the Boltzmann distribution among energy levels, we assume the LTE(Local Thermodynamic Equilibrium) in the lossy piece. This means that Kirchhoff's law

$$\frac{\epsilon_\nu}{\kappa_\nu} = \mathcal{B}_\nu(T)$$

holds as a good approximation, although the radiation is not in the thermodynamic equilibrium with the elements of the lossy piece

$$I_\nu \neq \mathcal{B}_\nu(T),$$

where

$$\mathcal{B}_\nu(T) = \frac{2h\nu^3}{c^2}\frac{1}{e^{\frac{h\nu}{kT}} - 1} \simeq \frac{2\nu^2}{c^2}kT$$

is the Planck function for the intensity of the blackbody radiation in the environmental temperature T of the lossy piece. Using the Kirchhoff's law and introducing the optical depth τ_ν

$$d\tau_\nu = -\kappa_\nu dl,$$

we can reduce equation (223) into

$$\frac{dI_\nu}{d\tau_\nu} = I_\nu - \mathcal{B}_\nu(T).$$

If we express the intensity I_ν through some "temperature" T_B as

$$I_\nu = \frac{2\nu^2}{c^2} k T_B$$

which is an equivalent of the brightness temperature in the interstellar–medium case, the equation (223) is further reduced to

$$\frac{dT_B}{d\tau_\nu} = T_B - T. \tag{224}$$

In the isothermal environment (T = const), a solution of equation (224) within the lossy piece, where $0 \le l \le l_0$ and $\tau_\nu(0) \ge \tau_\nu \ge 0$ as shown in Figure 82, is

$$T_B(l) = T_B(0) e^{\tau_\nu - \tau_\nu(0)} + T(1 - e^{\tau_\nu - \tau_\nu(0)}), \tag{225}$$

where $T_B(0)$ now can be interpreted as representing the power of the incoming signal at the input (left–most edge in Figure 82) of the lossy piece. In particular, a solution at the output (right–most edge in Figure 82) is

$$T_B(l) = T_B(0) e^{-\tau_\nu(0)} + T(1 - e^{-\tau_\nu(0)}). \tag{226}$$

The first term in the RHS of equation (226) corresponds to the power of the input signal attenuated by a factor of $e^{-\tau_\nu(0)}$ in the lossy piece, while the second term shows the power of the newly added thermal noise.

Note that, if $\tau_\nu(0) = 0$, the second term of the RHS of equation (226) is also zero, so no thermal noise is added. Thus this general discussion leads to an important conclusion that **every receiver element adds thermal noise as far as it attenuates the input signal**.

Let us denote $\varepsilon = e^{-\tau_\nu(0)}$ and call it "transmission efficiency". $\varepsilon = 1$ means perfect lossless transmission and $\varepsilon = 0$ means complete attenuation or zero efficiency.

The thermal noise T_N added in the system in temperature T with a transmission efficiency ε is $T_N = T(1 - \varepsilon)$ according to equation (226). Therefore, the input–equivalent system noise temperature T_S of such a system is by definition

$$T_S = \frac{T_N}{\varepsilon} = \frac{T(1 - \varepsilon)}{\varepsilon}. \tag{227}$$

In order to get a low–noise device, one must achieve a high transmission efficiency ε and/or cool the whole system in order to lower the environmental temperature T. This is the reason why modern highly sensitive radio telescope receivers are usually cooled to a very low temperature close to 10 K using cryogenic systems.

Q : Sometimes we see in the advertisements of the satellite broadcasting receivers phrases like "65 K low noise amplifier in the room temperature". Isn't such a statement against the principles of the thermodynamics?
A : "65 K" here implies the system noise temperature which is the input–equivalent temperature of the noise added in the system. Since well made systems could transmit signals with high efficiency ($\varepsilon \simeq 1$) even in the room temperature $T \simeq 290K$, there is no violation of the thermodynamics.
Q : If we slightly misalign a wave guide tube, the system noise temperature of our receiving system could be as high as several thousand Kelvin. But why the purely passive element like wave guide tube can increase the temperature to such a high value?
A : You could increase the input–equivalent noise temperature to an arbitrarily high value if you only degrade the transmission efficiency. In fact, if $\varepsilon = 0$, then $T_N = T$ but $T_S = \infty$!

System noise temperature of the multistage receiving system

Receiving systems of the radio telescopes are usually composed of serially connected several amplifiers, frequency converters, and other related devices. Such a multistage design is required because it is extremely difficult to realize by a single amplifier the ultra–high amplification of the very weak signal of an astronomical radio source to the level of the tractable signal. In addition, it is also difficult to convert the high frequency signal to the low frequency signal by just one frequency converter.

For example, the power received from a source of the effective flux density $\mathcal{S}_\nu = 1$ Jy by an antenna with the diameter D =30 m, the aperture efficiency η_A =0.6 and the frequency bandwidth $B = 200$ MHz is

$$\begin{aligned} W &= \frac{1}{2} A_e \mathcal{S}_\nu B = \frac{1}{8} \eta_A \pi D^2 \mathcal{S}_\nu B \\ &= \frac{1}{8} \times 0.6 \times 3.145 \times 30^2 \times 1 \times 10^{-26} \times 200 \times 10^6 = 4.2 \times 10^{-16} \text{ W}. \end{aligned}$$

Even if we apply 100 dB (i.e., 10^{10} times) amplification to this signal, we obtain 4.2 μW only!

Therefore, a receiving system can be usually represented as a series of linear devices each having a gain G_i and a system noise temperature T_{Si}. A passive device like a transmission cable can be regarded as having a gain (= transmission efficiency ε) smaller than 1. For a device which both amplifies the signal and generates the noise, we can conceptually regard that a passive attenuator first reduces the signal power by a factor of ε_i and adds the thermal noise $T(1-\varepsilon_i)$, and then a lossless amplifier increases both the signal and the noise by a factor of G_i/ε_i, so that the total gain and the system noise temperature are described as G_i and $T_{Si} = T(1-\varepsilon_i)/\varepsilon_i$, respectively.

We do not include to this linear multistage receiving system the square law detectors or correlators because they are essentially non–linear devices.

Because of the linearity, the total gain G of an n–element multistage receiving system is expressed as

$$G = \prod_{i=1}^{n} G_i.$$

If we describe the gain in dB, then we have

$$G = \sum_{i=1}^{n} G_i.$$

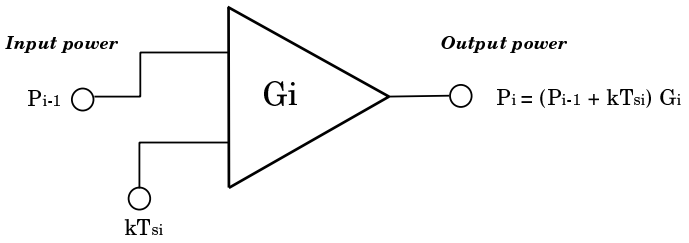

Figure 83: Input and output of a linear device with gain G_i and system noise temperature T_{Si}.

Now let us describe the output power P_i of i-th device in the n–element multistage receiving system through its gain G_i, its system noise power kT_{Si} and the output power of (i–1)–th device P_{i-1} as

$$P_i = (P_{i-1} + kT_{Si})G_i, \tag{228}$$

(see Figure 83). Then if we denote the input power to the first device as $P_0 = kT_A$, where T_A is the antenna temperature due to an astronomical source, the output power of n–th device is expressed as

$$\begin{aligned} P_n &= \{\cdots\{[(P_0 + kT_{S1})G_1 + kT_{S2}]G_2 + kT_{S3}\}G_3 + \cdots + kT_{Sn}\}G_n \\ &= kT_A \prod_{i=1}^{n} G_i + kT_{S1} \prod_{i=1}^{n} G_i + kT_{S2} \prod_{i=2}^{n} G_i + kT_{S3} \prod_{i=3}^{n} G_i \\ &\quad + \cdots + kT_{Sn}G_n, \end{aligned} \tag{229}$$

(Figure 84).

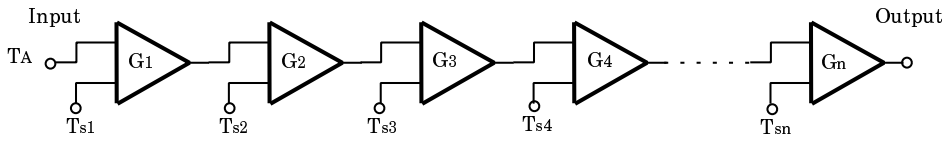

Figure 84: Multistage receiving system.

Denoting the total gain of the system as $G = \prod_{i=1}^{n} G_i$ and equating P_n in equation (229) to the sum of the input power and the sytem noise power multiplied by the total gain

$$P_n = k(T_A + T_S)G, \tag{230}$$

we obtain an expression for the total system noise temperature :

$$T_S = T_{S1} + \frac{T_{S2}}{G_1} + \frac{T_{S3}}{G_1 G_2} + \cdots + \frac{T_{Sn}}{G_1 G_2 \cdots G_{n-1}}. \tag{231}$$

Since usually $G_i \gg 1$, the system noise temperature of a device in a later stage does not contribute much to the total system noise temperature T_S. On the other hand, the system noise temperature of the first stage device directly contributes to T_S and practically determines the system noise performance of the whole system. Therefore, **it is important to make the first stage device (mostly amplifiers) as low–noise as possible**. For this purpose, the first stage device is often cooled down to $\sim$ 15 K or lower. First stage amplifiers are frequently called "low noise amplifiers (LNA)".

What we showed above can be summarized as replacing the actual multistage receiving system shown in the upper panel of Figure 85 by a simple total system shown in the lower panel.

The total system noise temperature T_S figuring in the output of the whole system as shown in equation (230) or in the lower panel of Figure 85 is relatively easily measured and provides the ratio T_A/T_S which is important in the signal–to–noise ratio considerations.

Receiver noise temperature T_{RX}

The system noise temperature of the receiving system only (i.e., not including other effects like atmospheric or ground pickup) is often called as "receiver noise temperature" and is denoted by T_{RX}.

Historically, a variety of low noise devices were used for the first stage which practically determines T_{RX}. Some examples of receiver noise temperatures of various devices are shown in Figure 86 as functions of observing frequency.

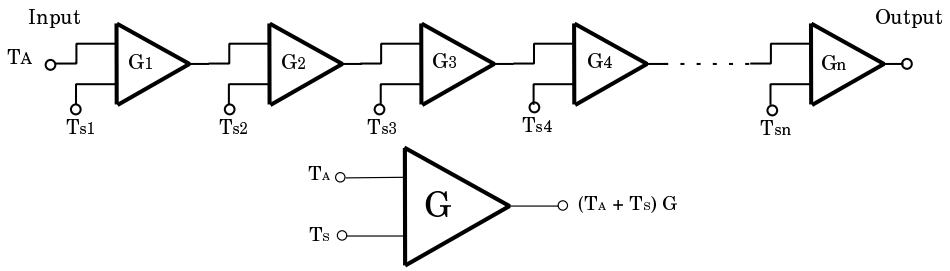

Figure 85: Total gain and total system temperature of the multistage receiving system.

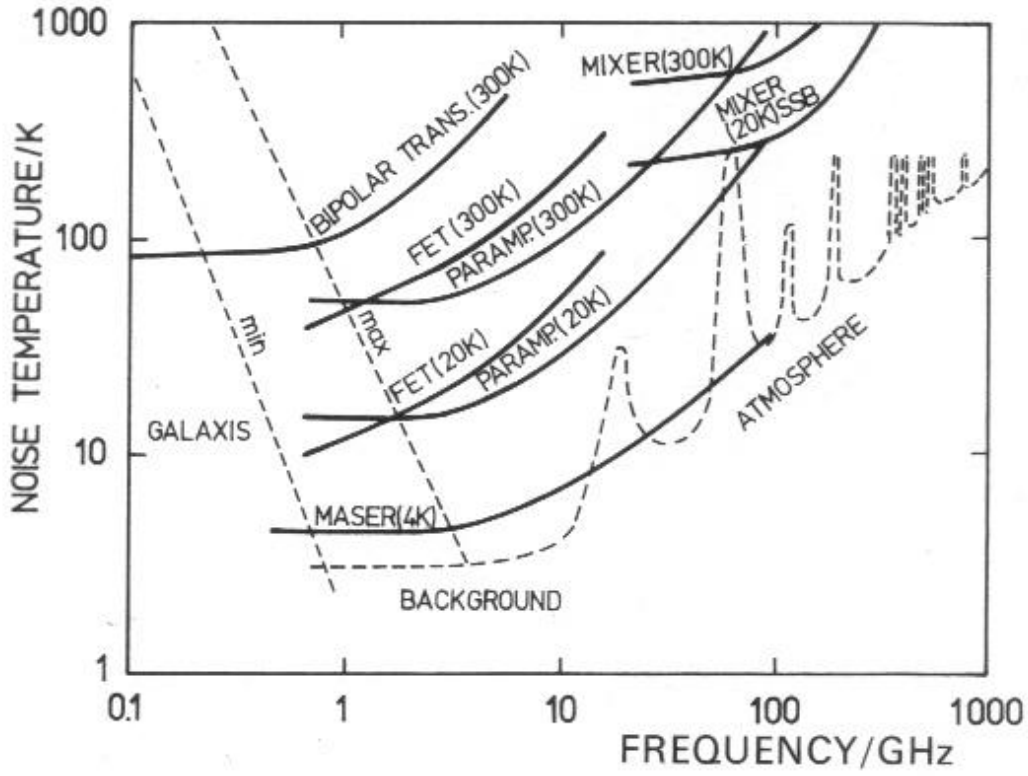

Figure 86: Receiver noise temperatures T_{RX} presented in the 1st edition of 'Tools of Radio Astronomy' by K. Rohlfs (1986).

Nowadays, low noise amplifiers (LNA's) made of HEMT (High Electron Mobility Transistor) are mostly used in a wide frequency range lower than 100 GHz, because HEMT amplifiers are stable, easy to handle and capable of receiving sufficiently wide bandwidth. HEMT amplifiers for 22 GHz or higher frequency are usually cooled by He–gas cryogenic system (Figure 88). At 8 and 2 GHz which are used in the current geodetic VLBI observations, He–cooling is rarely applied since the HEMT amplifiers are good enough at these frequency ranges even in the room temperature or in the electronically cooled system to provide T_{RX} as low as several tens of Kelvins.

At frequency higher than 100 GHz, He–cooled SIS (superconductor–insulator–superconductor) mixers instead of the amplifiers are usually used as the first stage device, since it is still difficult to make reliable low noise amplifiers for the high frequency range (Figure 89).

Figure 87 shows present day performance of the HEMT amplifiers and SIS mixers.

Antenna noise temperature T_{ant} and atmospheric contribution to the system noise temperature

Besides the thermal noise generated within the receiving system, we have other contributions to the system noise.

The thermal radiation from the ground, which is picked up by sidelobes or spill overs from the main reflector edge, and all other noises which could be picked up or generated by the antenna structure, can be regarded as forming some "antenna related" noise power. We call a temperature which is calculated from the power using the Nyquist theorem "antenna noise temperature" and denote as T_{ant}. Since the antenna noise temperature T_{ant} comes from outside of the antenna receiving system, it can be treated just as an input noise to the first stage of the receiving system. This means that T_{ant} can be simply added to T_{S1} in the RHS of equation (231). Of course, the antenna noise temperature must strongly depend on the elevation of the antenna beam since the ground pickup must be stronger at lower elevation.

The noise temperature T_{Nair} due to the thermal radiation of the atmosphere of temperature T_{atm} is simply expressed by the radiative transfer theory as

$$T_{Nair} = T_{atm}(1 - e^{-\tau_{atm}}), \tag{232}$$

where τ_{atm} is the optical depth of the atmosphere at the observing frequency and at the direction of the antenna beam. Note that τ_{atm} and therefore T_{Nair} must vary significantly with the elevation of the antenna beam. The antenna noise temperature T_{Nair} can be also just added to T_{S1} in the RHS of equation (231).

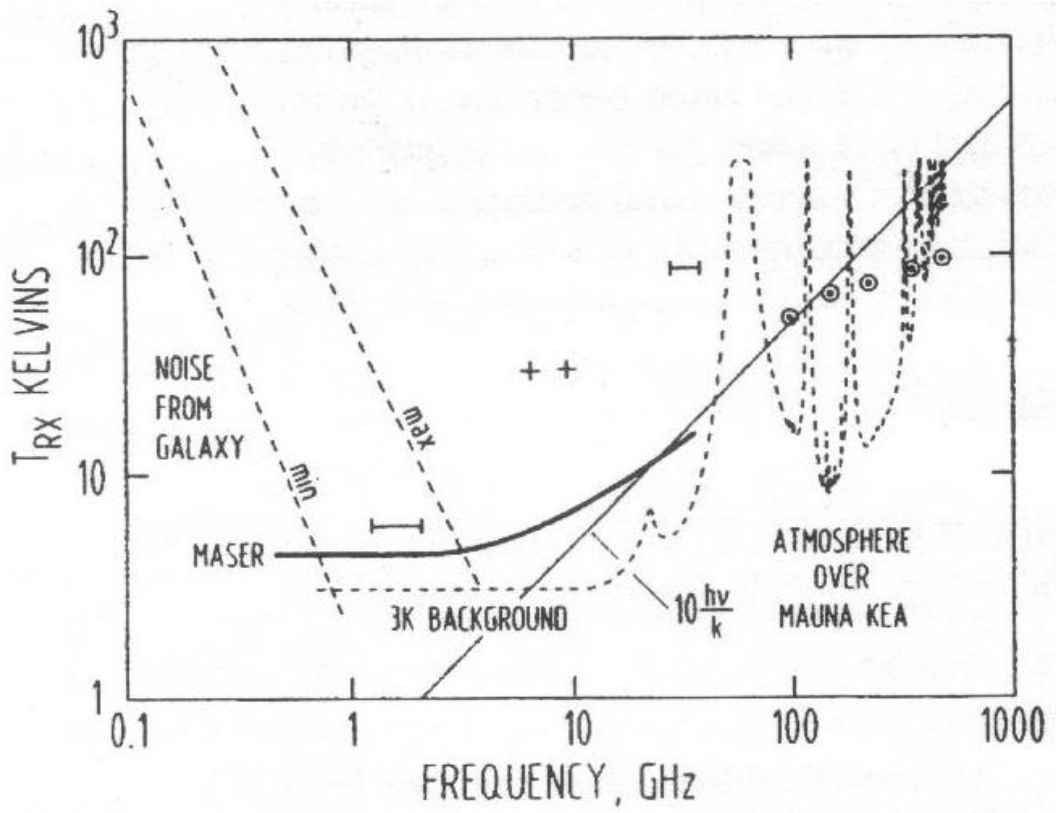

Figure 87: Receiver noise temperatures T_{RX} v.s. frequency diagram brought from 3rd edition of 'Tools of Radio Astronomy' by K. Rohlfs and T.L. Wilson (2000). Note that many devices shown in Figure 86, including FET and Parametric amplifiers, are no longer shown in this Figure, because they are largely replaced by HEMT amplifiers.

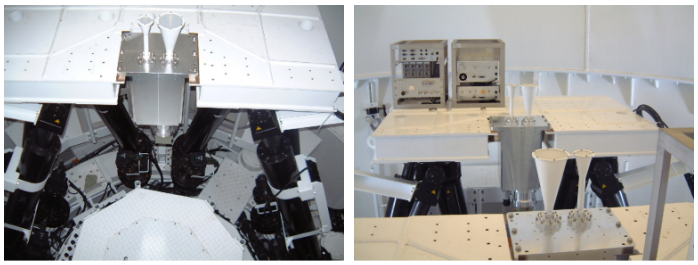

Figure 88: He–cooled 22 GHz and 43 GHz HEMT amplifiers placed in common cryogenic dewers and mounted on VERA dual–beam receiving system.

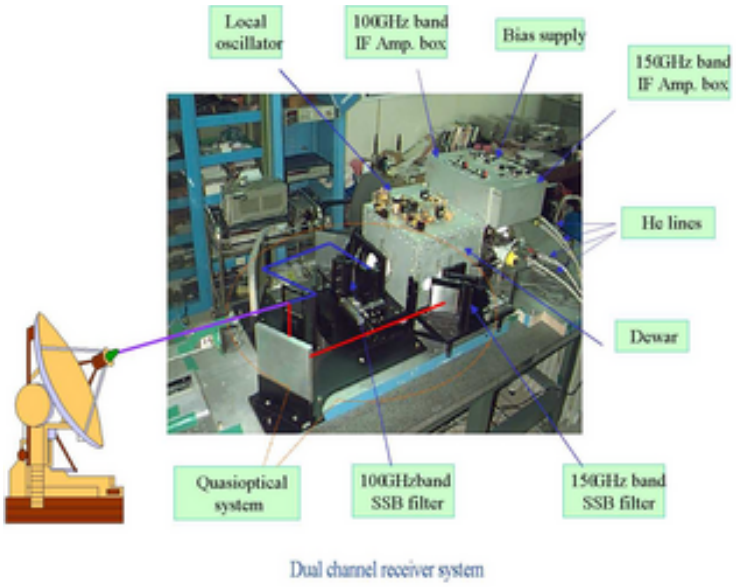

Figure 89: He–cooled SIS–mixer in the 100 GHz and 150 GHz dual channel receiver system of Taedok Radio Astronomy Observatory, Korea.

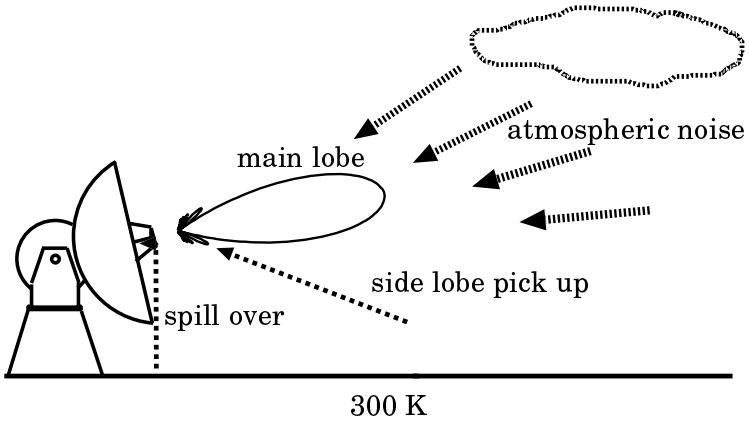

Figure 90: Atmospheric and ground pickup contributions to the system noise temperature.

Therefore, the total system noise temperature T_S now including the atmospheric and the ground pickup effects can be expressed as

$$T_S = T_{RX} + T_{ant} + T_{atm}(1 - e^{-\tau_{atm}}). \tag{233}$$

5.2 Frequency Conversion

If a radio signal is band limited, that means it has non–zero spectral power within a certain frequency range with a finite bandwidth only, we can shift the central frequency of such a signal (usually to the lower side) without losing any spectral information (Figure 91). This is the principle of the frequency conversion.

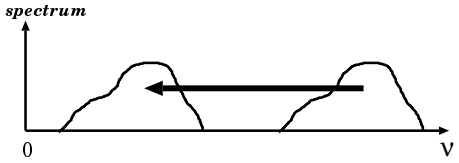

Figure 91: Frequency shift of the band limited signal.

Frequency conversion is needed in the radio telescope receiving system because

1. the received signal is much less attenuated in transmission cabeles and is much easily filtered, detected by a square–law detector, digitized, and so on, in the low frequency (lower than 1 GHz, say) than in the high frequency, and

2. it is necessary that the input and output signals are in different frequency bands before and after the enormous amplification of several tens to hundreds dB for avoiding possible leakage of a small amount of output power back to the input which could terribly oscillate the system.

Technical terms in the frequency conversion

A little peculiar technical terms are used in descriptions of the frequency conversion. They are

- **radio frequency (RF)** : the high frequency involved in the band which is directly received by an antenna, sometimes called also as "sky frequency",

- **intermediate frequency (IF)** : the low frequency obtained by the frequency conversion,
- **local oscillator (LO)** : an oscillator which provides a sinusoidal reference signal with a specified frequency,
- **mixer** : a nonlinear device which multiplies the reference signal to the received signal,
- **bandpass filter (BPF)** : a filter which passes only necessary band of the intermediate frequency and cuts off all other frequency components in the output from the mixer,
- **down converter** : a unit composed of the LO, mixer and BPF which converts RF ν to IF $\nu - \nu_{LO}$, where ν_{LO} is the frequency of the reference signal provided by LO which is often called "LO frequency",
- **superheterodyne receiver** : a receiver based on the frequency conversion technique.

Figure 92 shows elements of the frequency conversion. For example, in a typical Mark III–type geodetic VLBI observation at 8 GHz, we can choose 8180 MHz $\sim$ 8600 MHz as the RF band and 8080 MHz as the LO frequency ν_{LO} to get 100 MHz $\sim$ 520 MHz as the IF band.

What is the mixer?

The main part of the mixer is a device (for example, a mixer diode) which has a nonlinear current (I)–voltage (V) relation :

$$I = a_0 + a_1 V + \boldsymbol{a_2 V^2} + a_3 V^3 + \cdots, \tag{234}$$

where a_i's are constant coefficients. The received signal and the reference signal from LO are first summed up before entering to the nonlinear device. So, denoting the voltage of a frequency component with the angular frequency $\omega = 2\pi\nu$ in the RF band as $V_s \cos(\omega t + \phi)$ and the voltage of the reference signal with LO angular frequency ω_{LO} as $V_{LO} \cos(\omega_{LO} t + \phi_{LO})$, where ϕ and ϕ_{LO} are initial phases, we have the summed input voltage V to the nonlinear device :

$$V = V_s \cos(\omega t + \phi) + V_{LO} \cos(\omega_{LO} t + \phi_{LO}). \tag{235}$$

Therefore, we obtain a term proportional to $V_s V_{LO}$ in the output current from the nonlinear device due to the second order term $a_2 V^2$ of the I–V relation

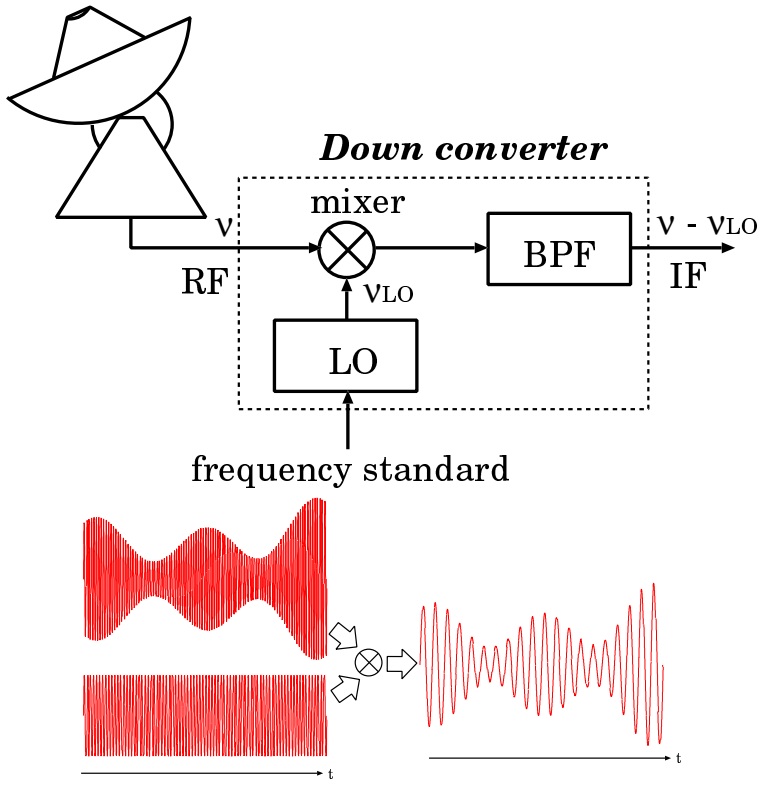

Figure 92: Elements and technical terms used in the frequency conversion.

in equation (234), which is

$$\begin{aligned} & a_2 V_s V_{LO} \cos(\omega t + \phi) \cos(\omega_{LO} t + \phi_{LO}) \\ = \; & \frac{1}{2} a_2 V_s V_{LO} \cos[(\omega + \omega_{LO})t + \phi + \phi_{LO}] \\ + \; & \frac{1}{2} a_2 V_s V_{LO} \cos[(\omega - \omega_{LO})t + \phi - \phi_{LO}]. \end{aligned} \tag{236}$$

The first and second terms in the RHS of equation (236) correspond to up–converted and down–converted frequencies, respectively. So, we can realize the desired frequency conversion by picking up only the down–converted frequency with a suitably designed BPF which passes the down–converted frequency and cuts off the up–converted as well as all other frequencies produced by the I–V relation in equation (234). Note that the linearity of the received signal is preserved in this procedure though we used the nonlinear device.

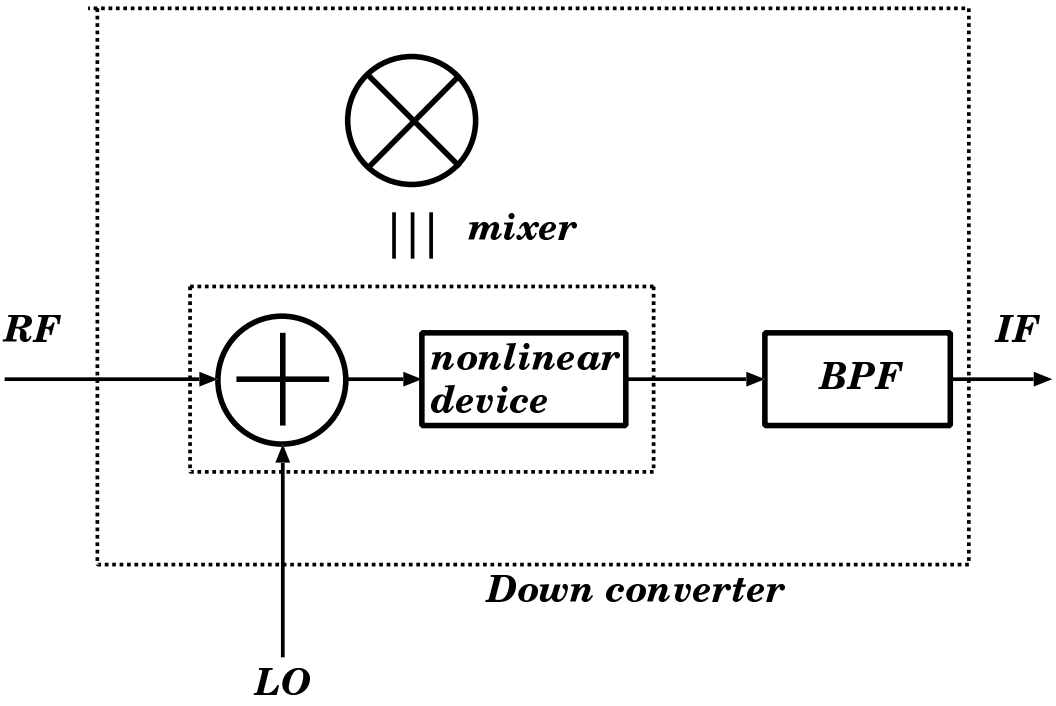

Figure 93: General concept of the mixer.

Upper sideband and lower sideband

If the LO frequency is chosen within the received RF band, two frequencies ω_U and ω_L satisfying relations

$$\omega_U - \omega_{LO} = \omega_{IF} \tag{237}$$

and

$$\omega_{LO} - \omega_L = \omega_{IF}, \tag{238}$$

both pass through the BPF and have the same IF ω_{IF}. In fact, they both have the same $\cos(\omega_{IF}t + \cdots)$ form in the mixer output as the above discusion on the mixer principle shows. The frequency higher than the LO frequency forms "upper sideband (USB)" and the lower one forms "lower sideband (LSB)". When the both sidebands contribute to the IF band, the IF spectrum turns out to be a superposition of the down–converted spectra of the USB and the frequency–inverted LSB as illustrated in Figure 94. Frequencies in

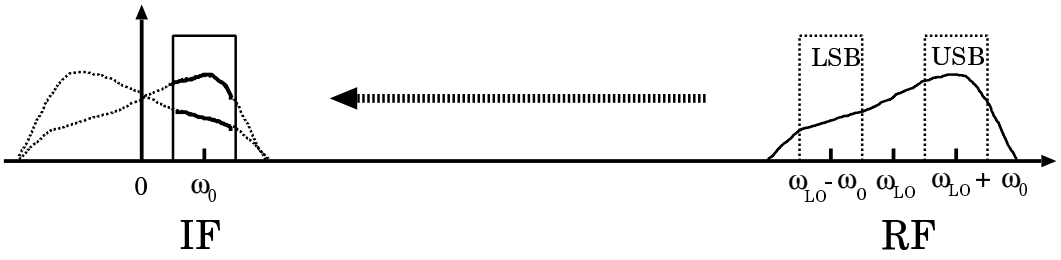

Figure 94: Upper and lower sidebands are converted to the same IF band.

different sidebands corresponding to the same IF frequency are called "mirror frequencies" or "image frequencies".

Sideband rejection

A superheterodyne receiving system which accepts both USB and LSB contributions to the IF band is called "double sideband (DSB)" receiver, while a system which receives only one of the two sidebands is called "single sideband (SSB)" receiver.

For observations of continuum sources, the DSB receivers offer better sensitivity, while for observations of spectral line sources, SSB receivers are preferred. In VLBI observations, SSB receivers are mostly used even for continuum sources for avoiding complication of the data processing.

The process to remove unnecessary sideband for a SSB observation is called "sideband rejection" or "image rejection". The rejection is most simply realized by a suitable BPF in the RF range which limits the RF band of the mixer input in such a way that the LO frequency falls outside of the input band (Figure 95).

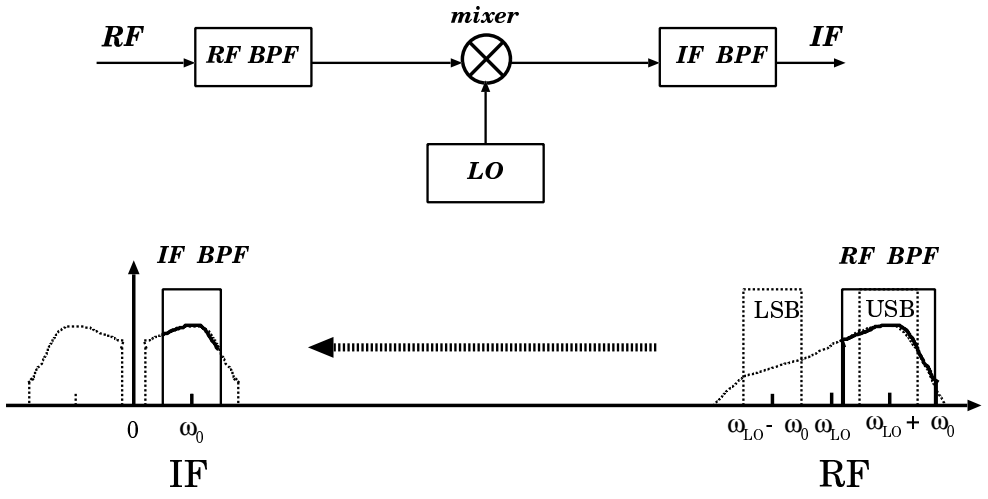

Figure 95: Sideband rejection in terms of the BPF in the RF band.

5.3 Signal–to–Noise Ratio of the Single–Dish Radio Telescope

The radio signal from an astronomical source received by a radio telescope antenna is transmitted through the receiving system as a voltage or current signal as shown in Figure 96. After several steps of amplification and down–conversion a unit like a square–law detector squares the voltage (or current) signal to make a power signal which is eventually displayed to observers by a powermetor or other devices.

The power signal is the squared voltage averaged for some "integration time". A statistical theory of the dispersion of the time–averaged squared random variable shows that the signal–to–noise ratio S/N of the detected power is expressed in terms of the antenna temperature T_A, system noise temperature T_S, bandwidth B and integration (or averaging) time τ_a by a formula :

$$S/N = \frac{T_A}{T_S}\sqrt{B\tau_a}. \tag{239}$$

Or, if we introduce again the Boltzmann constant k, the effective aperture A_e, the effective flux $\mathcal{S}_\nu$, the aperture diameter D and the aperture efficiency η_A, we have

$$S/N = \frac{A_e\mathcal{S}_\nu}{2kT_S}\sqrt{B\tau_a} = \frac{\pi\eta_A D^2\mathcal{S}_\nu}{8kT_S}\sqrt{B\tau_a}. \tag{240}$$

As we mentioned earlier, even if we receive a relatively strong radio source with the effective flux density of 1 Jy ($= 10^{-26}$ W m^{-2} Hz^{-1}) using an

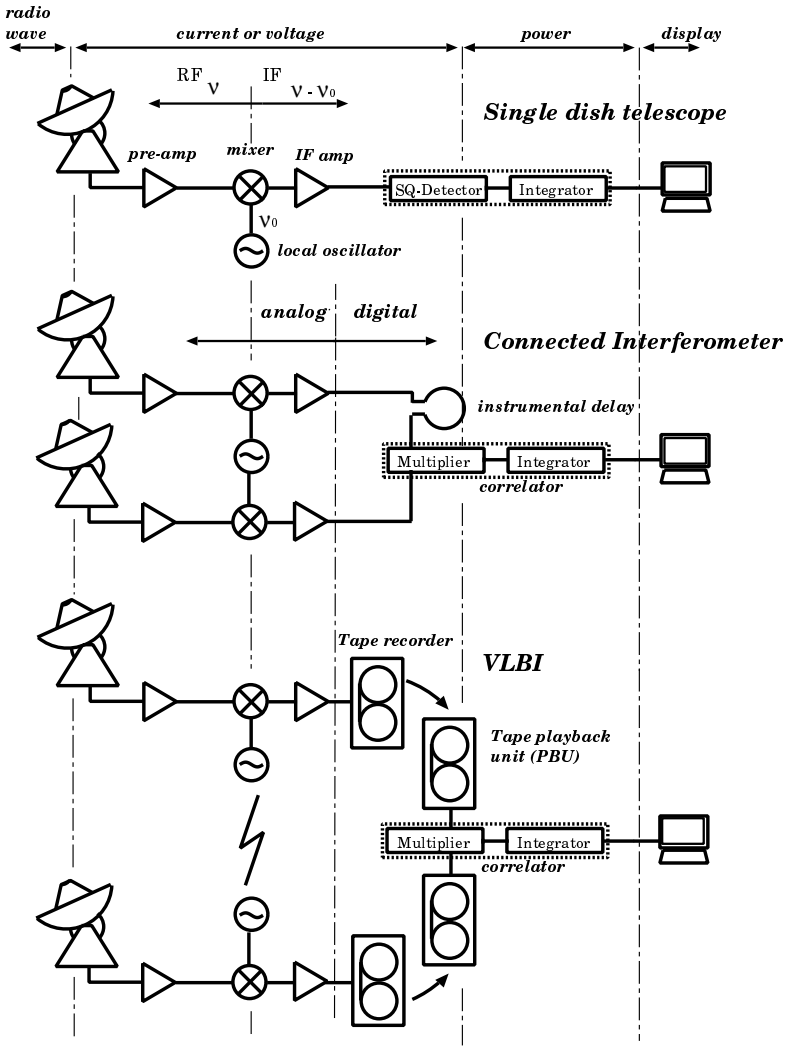

Figure 96: Superheterodyne receiving systems in single dish telescope, connected interferometer and VLBI.

20 m diameter antenna with the aperture efficiency of 60 %, the antenna temperature is only about

$$T_A = \frac{1}{2k} A_e \mathcal{S}_\nu = \frac{1}{8k} \eta_A \pi D^2 \mathcal{S}_\nu \simeq 0.068 \text{ K}.$$

Therefore, if the system noise temperature is 100 K (which is not a bad figure, by the way),

$$\frac{T_A}{T_S} = 6.8 \times 10^{-4} \text{ !}$$

Nevertheless, if we observe the source with the receiving bandwidth $B = 100$ MHz and integrate for 1 second, then the signal–to–noise ratio is

$$S/N = 6.8,$$

and therefore well detectable! This is the way how radio astronomers receive the radio source signals.

5.4 Gain Variation of the Receivers and Switching Observations

The high signal–to–noise ratio does not necessarily guarantee the detection of a radio source since there might be disturbing effects other than the thermal noise. One of such effects is the gain variation of the receiving system. At frequencies higher than $\simeq$ 15 GHz, the time variation of the atmospheric thermal radiation is another serious problem.

In order to understand why the gain variation may prevent the detection of the signal from the astronomical source, we consider again the output power of the receiving system :

$$W = k(T_A + T_S)GB, \tag{241}$$

where k is the Boltzmann constant, T_A is the antenna temperature due to the astronomical source, T_S is the system noise temperature, G is the total gain and B is the receiving bandwidth. Suppose that G shows some short–term (a few minutes, say) time variation ΔG. Then received power would vary as much as

$$\Delta W_G = k(T_A + T_S)\Delta G\, B. \tag{242}$$

Since the system temperature T_S is usually much larger than the antenna temperature T_A, the variation $\Delta W_G \simeq kT_S\, \Delta G\, B$ could easily exceed $kT_A GB$ itself, making the detection of the source almost impossible (Figure 97).

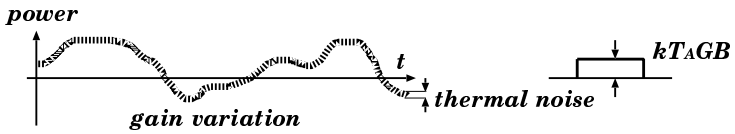

Figure 97: The gain variation prevents the detection of the astronomical radio source even when the S/R is sufficiently high.

The simple observation measuring the received power as described so far is called "total power mode observation" or "total power radiometry".

The above discussion shows that a condition

$$\frac{T_A}{T_S} > \frac{\Delta G}{G} \tag{243}$$

must hold, in order for a source to be detected against the gain variation in the total power mode observation. For a source with $T_A \simeq 10^{-4} T_S$, this implies that we must suppress the gain variation of the receiving system $\Delta G/G$ within a level as small as

$$\frac{\Delta G}{G} < 10^{-4} \tag{244}$$

which is practically quite difficult.

Therefore, radio astronomers apply switching techniques to the single dish telescope observations in order to avoid the gain variation effect and firmly detect the weak astronomical radio sources.

Dicke mode switching observation

A switching method was proposed by R.H. Dicke in 1946 and has been widely used. In this method, two input signals (voltages), one from the antenna feed horn and another from a noise source (ex., a cooled resistor) of a constant temperature T_R are quickly switched in front of the receiving system and, at the same time, the sign of the output power from the square–law detector is reversed with the same switching timing which are then summed up (i.e., the powers from the two inputs are subtracted) and averaged as shown in Figure 98.

The square–law detector yields a power

$$W_A = k(T_A + T_S)GB, \tag{245}$$

in the upper switch position in Figure 98, while in the lower switch position,

$$W_R = k(T_R + T_{RX})GB, \tag{246}$$

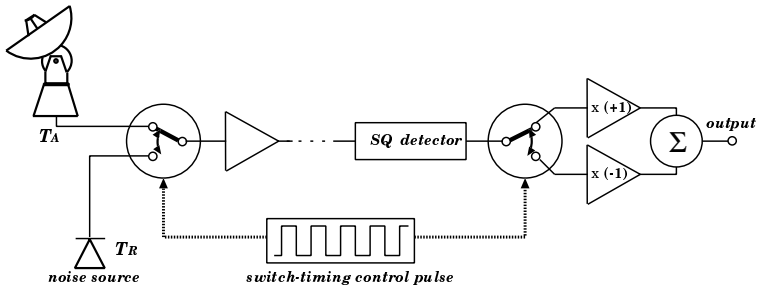

Figure 98: A general scheme of the Dicke mode receiver.

where T_{RX} is the receiver noise temperature. At relatively low observing frequency ≤ 15 GHz, the atmospheric contribution to the system noise temperature is much smaller than T_{RX} and the ground pickup effects are also small as far as the elevation is not very low. Consequently, $T_S \cong T_{RX}$ in equation (233) for the low frequency observations. Therefore, the difference output from the Dicke mode reciever gives us

$$W_A - W_R \cong k(T_A - T_R)GB, \tag{247}$$

provided that the switching interval is short enough compared with the time scale of the gain variation. Thereofre, the effect of the gain variation to this difference output is

$$\Delta(W_A - W_R) \cong k(T_A - T_R)\,\Delta G\,B \tag{248}$$

and hence the condition for the source detection is now

$$\frac{T_A}{T_A - T_R} > \frac{\Delta G}{G} \tag{249}$$

instead of equation (243). This new condition is easily satisfied if we suitably choose the temperature T_R low enough to be $T_R \sim T_A$.

Instead of the noise source, we could use a small antenna like a horn antenna looking at the sky for the switching observation (Figure 99).

For spectral line sources, quick "frequency switching" between inside and outside of the spectral lines can be used for detecting the sources against the gain variation of the receiving system.

Beam switch

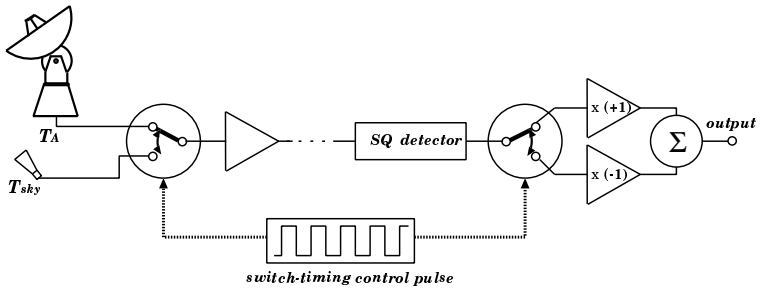

Figure 99: A switching mode observation using a small antenna.

At frequencies higher than $\simeq$ 15 GHz, the radio signal from an astronomical radio source is badly masked by the time variation of the thermal radiation of the atmosphere along with the gain variation of the receiving sysytem. Also, it is not desirable at the high frequencies to install the switching mechanism in the RF signal transmision system since the switches themselves could become additional noise sources.

In this circumstance, another technique called "beam switching" is effective. The beam switching is a quick switching of the antenna beam between the on–source and the off–source directions and measuring the difference of the received powers at the two directions (Figure 100).

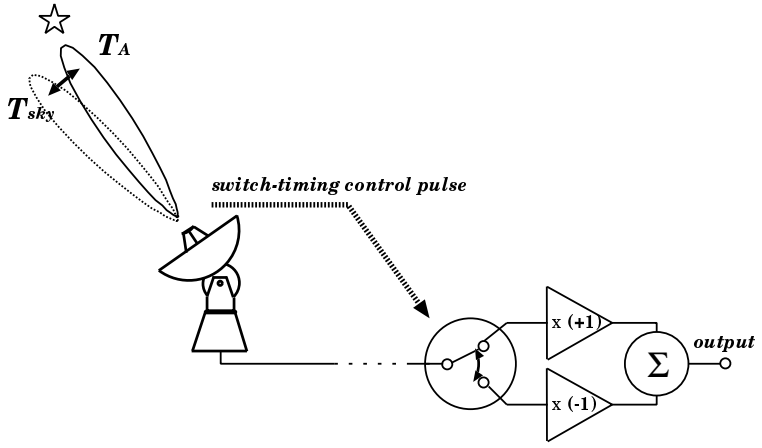

Figure 100: Beam switching between the on–source and the off–source directions.

Since the on–source output power is

$$W_{source} = k(T_A + T_S)GB, \tag{250}$$

while the off–source output power is

$$W_{sky} = kT_S GB, \tag{251}$$

the power difference yields

$$W_{source} - W_{sky} = kT_A GB. \tag{252}$$

Therefore, we obtain the antenna temperature T_A by a formula

$$T_A = \frac{W_{source} - W_{sky}}{W_{sky}} T_S, \tag{253}$$

if we know the system noise temperature T_S. The atmospheric thermal radiation effects are well cancelled out in the above procedure, as far as the two beam directions are not too much separated on the sky.

Since it is mostly impossible to shake a big antenna with a frequency of several to tens Hz, as usually required, a vaned–wheel like mirror is rotated in front of the feed horn for effectively altering the beam direction. The rotating mirror, which itself is a small antenna of the offset–paraboloidal shape, periodically intersects and cuts off the beam from the sub reflector and instead guides the reflected radio wave from another direction of the sky to the feed horn (Figure 101). Although the beam pattern for the mirror–reflected radio wave must be wide, distorted and thus quite different from that of the main beam, the pattern must be good enough to collect the widespread thermal radiation of the atmosphere (see equations (219) and (222)). The timing of the sign reversal of the output power is synchronized to the mechanical rotation of the mirror. The off–source beam is usually designed to be shifted in the azimuthal direction from the on–source beam for avoiding effects of the elevation dependence of the atmospheric thermal radiation.

Since we observe the astronomical source only for about a half of the total observing time τ_a in the above switching mode observations, the signal–to–noise ratio is now

$$S/N \simeq \frac{T_A}{T_S} \sqrt{\frac{B\tau_a}{2}}, \tag{254}$$

which is by a factor of about $\sqrt{2}$ smaller than that of the total power mode observation (equation (239)). But the $\sqrt{2}$ times loss is certainly better than nothing.

The above switching techniques are inevitable for measurements of the received power (radiometry) from astronomical radio sources by the single dish radio telescopes. On the other hand, for VLBI observations, and for

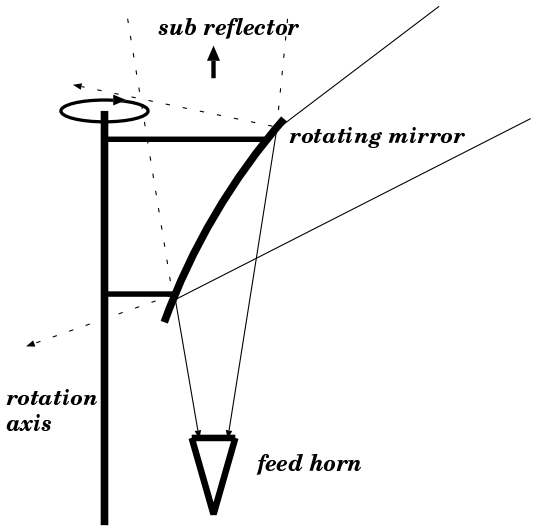

Figure 101: Beam switching using the vaned–wheel like rotating mirror.

interferometry observations in general, the switching techniques are usually not used in a direct sense. This is because the receiver noises in the voltages from different telescopes are not correlated with each other, and therefore the system noise temperature term T_S does not figure in the interferometer analog of equation (241) obtained in a correlator output. Nevertheless, the techniques are still important for VLBI in checking and adjusting the pointing performance of VLBI telescopes and in measuring single–dish flux densities of the sources which are necessary for interpreting the VLBI results.

6 Measurements of Antenna Performance

For proper interpretaion of the results of the VLBI observations, accurate knowledge of the basic performance parameters of the component telescopes is necessary. Therefore, the performance must be measured. Moreover, the measurements must be repeated periodically, since the parameters are more or less time variable depending on the environmental conditions, the system improvement works and the ages of the telescopes. The measurements are also important for quickly finding and repairing possible mulfunctions in the antenna system.

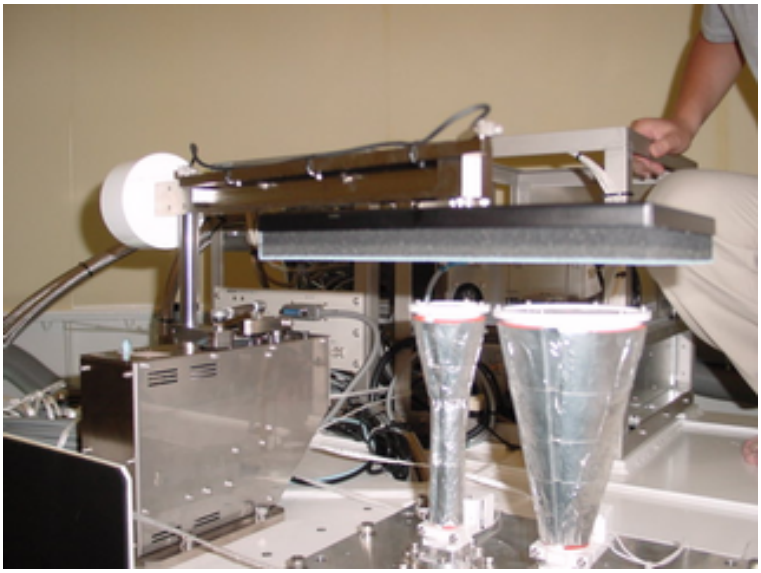

Figure 102: Room–Sky measurement device in a upper cabin of a VERA antenna.

6.1 System Equivalent Flux Debsity (SEFD) and other parameters

The most important parameters to be regularly measured are the system noise temperature T_S and the aperture efficiency η_A which figure in a quantity called "system equivalent flux density" often denoted as SEFD. The SEFD is defined as

$$SEFD = \frac{2kT_S}{A_e} = \frac{2kT_S}{\eta_A A_g} = \frac{8kT_S}{\eta_A \pi D^2}, \tag{255}$$

where the last term is for a circular aperture antenna of diameter D, so that

$$\frac{T_A}{T_S} = \frac{\mathcal{S}_\nu}{SEFD}, \tag{256}$$

where again k is the Boltzmann constant, A_e is the effective aperture, A_g is the geometrical aperture, T_A is the antenna temperature and $\mathcal{S}_\nu$ is the effective flux density of an astronomical radio source. Therefore, if we know both $\mathcal{S}_\nu$ of a source and SEFD of a telescope, we can easily calculate an expected signal–to–noise ratio to be obtained when we observe the source by the telescope on the basis of a suitable equation like equation (239). The SEFD values are widely listed in the system documents of the VLBI networks as indicators of the sensitivities of the component telescopes. For example,

SEFD of an antenna with diameter 20 m, aperture efficiency 0.6 and system noise temperature 100 K is 1465 Jy. Of course, the smaller the SEFD, the better the system performance is.

Other important quantities are the receiver noise temperature T_{RX} and the pointing accuracy σ_θ of the telescope, which must be carefully monitored for checking the system status. Also, the optical depth of the atmosphere τ_{atm} is regularly measured as an parameter showing the observational condition of the telescope site.

6.2 Measurement of the Receiver Noise Temperature T_{RX}

The receiver noise temperature T_{RX} is measured usually in a scheme shown in Figure 103 known as "hot–cold method".

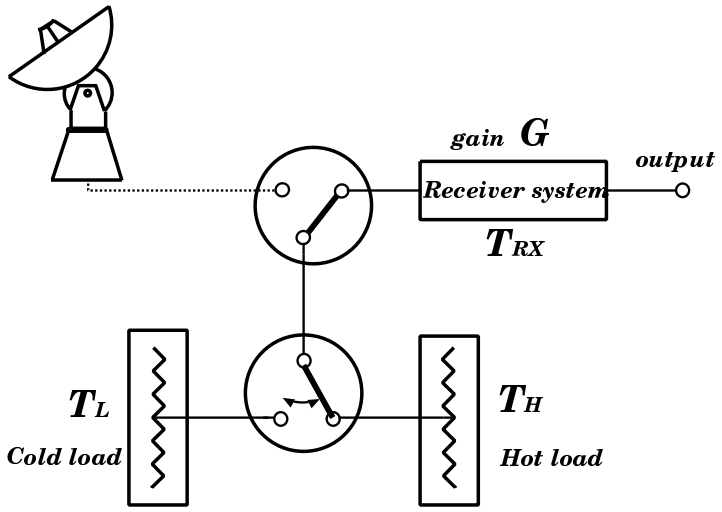

Figure 103: Measurement of the receiver noise temperature using hot and cold loads with known temperature values.

When the hot and cold loads with known temperatures T_H and T_L (for example, two resistors one in the room temperature and another cooled by the liquid nitrogen) are switched, the output power values W_L and W_H are

$$\begin{aligned} W_L &= k(T_L + T_{RX})GB \\ W_H &= k(T_H + T_{RX})GB, \end{aligned} \tag{257}$$

where G and B are the total gain of the receiving system and the observed bandwidth, respectively. Therefore, introducing a factor

$$y \equiv \frac{W_H}{W_L} = \frac{T_H + T_{RX}}{T_L + T_{RX}},$$

we obtain the T_{RX} and GB by equations

$$\begin{aligned} T_{RX} &= \frac{T_H - yT_L}{y-1} \\ GB &= \frac{W_H - W_L}{k(T_H - T_L)}. \end{aligned} \tag{258}$$

In geodetic VLBI antennas observing at 2 GHz and 8 GHz, switching schemes such as shown in Figure 103 are often builtin in the receiving system for regular monitoring of the receiver noise temperature. In high frequency antennas, two absorbers one in the room temperature and another immersed in the liquid nitrogen are often used to cover a feed horn one by one, instead of the switching mechanism shown above.

6.3 Measurement of the System Noise Temperature

In many VLBI antennas, noise diodes with known noise temperatures are used for measurements of the system noise temperature. The noise power of temperature T_{ND} from a diode is intermittently added through a wave guide coupler to the input signal of the antenna looking at the blank sky, as shown in Figure 104.

The output powers of the receiving system when the noise power is added and when it is off are

$$\begin{aligned} W_{ND} &= k(T_{ND} + T_S)GB \\ W_{sky} &= kT_SGB, \end{aligned} \tag{259}$$

and hence

$$\frac{W_{ND} - W_{sky}}{W_{sky}} = \frac{T_{ND}}{T_S} \tag{260}$$

where the system noise temperature T_S includes contributions from the receiver noise, the ground pickup and the atmospheric thermal radiation as shown in equation (233). Therefore, introducing a factor y defined now as

$$y \equiv \frac{W_{ND}}{W_{sky}},$$

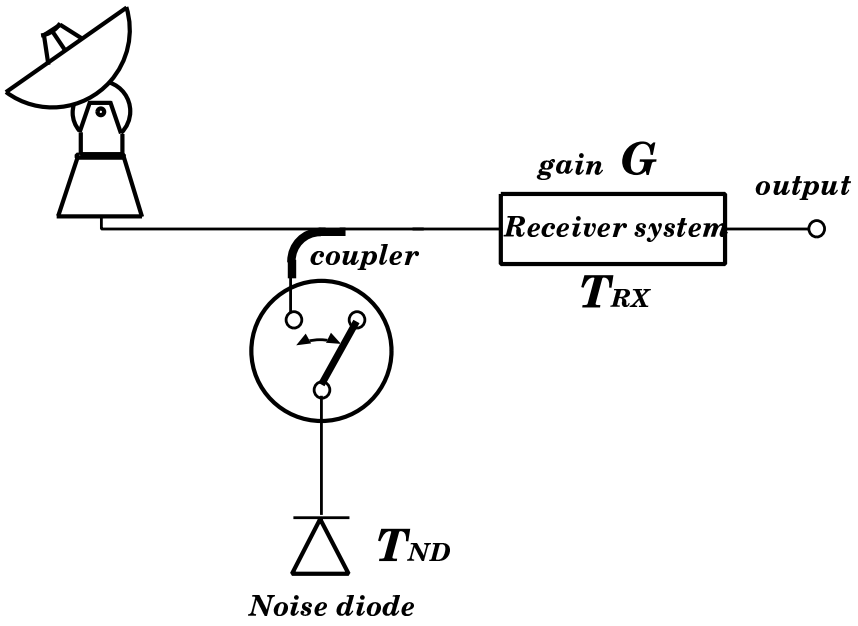

Figure 104: Measurement of the system noise temperature using a noise diode of the known noise temperature.

we obtain the system noise temperature

$$T_S = \frac{T_{ND}}{y-1}. \tag{261}$$

The system noise temperature as given in the input–equivalent form of equation (230) can be easily measured indeed. In VLBI observations, the measurements of the system noise temperature are usually highly automated and conducted in real time for every scan of radio sources.

6.4 System Noise Temperature and Antenna Temperature Referred to the Outside of the Atmosphere T_S^* and T_A^*

The system noise temperature given in equation (233) and measured using the noise diode as described above is "referred to the antenna on the ground", that means the input point for both T_A and T_S is chosen at the front side of the antenna. In VLBI observations at frequency higher than $\simeq$ 15 GHz, however, the signal from a radio source received by an antenna could be heavily attenuated by a factor of $e^{-\tau_{atm}}$ due to the atmosphere with an optical depth τ_{atm} which often reaches 0.3 ($e^{-\tau_{atm}} \simeq 74\%$) or worse. It is much nicer for many astronomers to know the antenna temperature "referred to the outside of the atmosphere", which means not attenuated by the atmosphere, in order to study the intrinsic properties of the radio source. But if we shift the input point for the antenna temperature to the outside of the atmosphere, we have to refer also the system noise temperature to the outside of the atmosphere, too. They are usually denoted as T_S^* and T_A^* and related to the

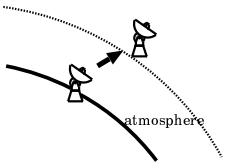

Figure 105: Input points within and outside the atmosphere.

ordinary T_S and T_A as

$$\begin{aligned} T_S &= T_S^* e^{-\tau_{atm}} \\ T_A &= T_A^* e^{-\tau_{atm}}, \end{aligned} \tag{262}$$

or

$$\begin{aligned} T_S^* &= T_S e^{\tau_{atm}} = (T_{RX} + T_{ant}) e^{\tau_{atm}} + T_{atm}(e^{\tau_{atm}} - 1) \\ T_A^* &= T_A e^{\tau_{atm}}. \end{aligned} \tag{263}$$

6.5 Measurement of the T_S*

There is a clever method to directly measure the system noise temperature T_S^* and then the antenna temperature T_A^* referred to the outside of the atmosphere. In fact, the T_S^* can be measured by a simple method called as "R–Sky" or "Room–Sky" method where the feed horn of the antenna is covered and uncovered by an absorber at the room temperature T_{room} as shown in Figure 106.

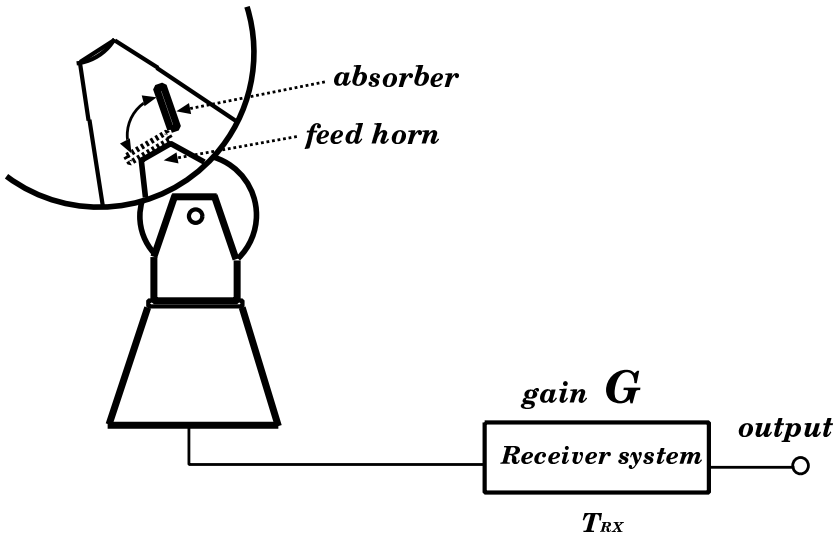

Figure 106: Measurement of the system noise temperature referred to the outside of the atmosphere by the R–Sky method.

For simplicity, we ignore the ground pickup effect T_{ant} in euation (233). In a well designed radio telescope antenna, the side lobe level is usually well suppressed so that $T_{ant} \leq 10$ K even at the low elevation.

When the absorber is taken off from the feed horn, the antenna looks at the blank sky and produces an output power

$$W_{sky} = kT_S^* e^{-\tau_{atm}} GB = k[T_{RX} + T_{atm}(1 - e^{-\tau_{atm}})]GB, \tag{264}$$

where T_{atm} is the physical temperature of the atmosphere. On the other hand, when the feed horn is covered by the absorber of temperature T_{room}, the output power is

$$W_{room} = k(T_{RX} + T_{room})GB. \tag{265}$$

If we adopt here a reasonable assumption that $T_{atm} = T_{room}$ (this must be correct with an accuracy usually better than a few percent since both of them must be around 300 K), then we have

$$W_{room} - W_{sky} = kT_{room}e^{-\tau_{atm}}GB, \tag{266}$$

and therefore

$$\frac{W_{sky}}{W_{room} - W_{sky}} = \frac{T_S^*}{T_{room}}. \tag{267}$$

If we introduce a factor y defined as

$$y \equiv \frac{W_{room}}{W_{sky}}, \tag{268}$$

we obtain the system noise temperature referred to the outside of the atmosphere

$$T_S^* = \frac{T_{room}}{y - 1}. \tag{269}$$

6.6 Measurement of the Aperture Efficiency η_A

If we switch the telescope beam between directions of a source and the blank sky using the beam switch mechanism we saw earlier, we obtain the difference of the output powers

$$\begin{aligned} W_{source} &= k(T_A^* + T_S^*)e^{-\tau_{atm}}GB \\ W_{sky} &= kT_S^*e^{-\tau_{atm}}GB. \end{aligned} \tag{270}$$

Therefore, we easily get the antenna temperature referred to the outside of the atmosphere if we know the sytem noise temperature referred to the outside of the atmosphere T_S^* by an equation :

$$T_A^* = \frac{W_{source} - W_{sky}}{W_{sky}}T_S^*. \tag{271}$$

If the radio source observed is a standard source with known effective flux density $\mathcal{S}_\nu$ and is much more compact than the beamwidth, we can calculate the effective aperture A_e from equation (220) as

$$A_e = \eta_A A_g = \frac{2kT_A^*}{\mathcal{S}_\nu}, \tag{272}$$

where A_g is the geometrical aperture, and therefore the aperture efficiency

$$\eta_A = \frac{2kT_A^*}{A_g \mathcal{S}_\nu} = \frac{8kT_A^*}{\pi D^2 \mathcal{S}_\nu}, \tag{273}$$

where the expression in the last term is for a circular aperture antenna with diameter D.

If the source is so bright that any short term variation of the receiver gain or the atmospheric noise can be ignored, we can just drive the telescope to switch the directions when the beam switching mechanism is not available.

The best standard sources are bright planets like Jupiter or Venus if the telescope beam is not narrower than their sizes, since their brightness temperatures are reliably known. Strong extragalactic AGN's like 3C84, 3C273, ... could also be used if their flux densities of the date are known.

6.7 Chopper Wheel method for Precise Measurement of the Antenna Temperature T_A^*

In order to accurately measure the antenna temperature T_A^* and aperture efficiency η_A, it is desirable to simultaneously measure the system noise temperature T_S^*, too. For this purpose, one can use a method known as "absorbing disk method" or "chopper wheel method". In this method, a rotating vaned–wheel like device is used which is much alike with the one used in the beam switch, but an absorber of the known temperature T_{room} is attached to one of the vanes in addition to the mirror. Thus the telescope can now look at the source direction, the mirror–reflected blank sky, and the absorber in turn within a short period of one revolution which is well less than 1 second (Figure 107). Here we again ignore the small ground pickup effect T_{ant}. Since

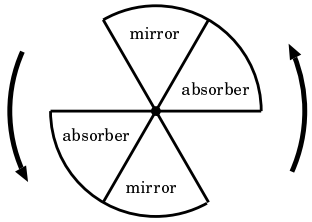

Figure 107: Absorbers and mirrors in the rotating vaned–wheel used in the chopper–wheel method.

we now have three output powers nearly simultaneously

$$\begin{aligned} W_{source} &= k[T_A^* e^{-\tau_{atm}} + T_{RX} + T_{atm}(1 - e^{-\tau_{atm}})]GB, \\ W_{sky} &= k[T_{RX} + T_{atm}(1 - e^{-\tau_{atm}})]GB = kT_S^* e^{-\tau_{atm}} GB, \\ W_{room} &= k(T_{room} + T_{RX})GB, \end{aligned} \tag{274}$$

we directly obtain the antenna temperature referred to the outside of the atmosphere

$$T_A^* = \frac{W_{source} - W_{sky}}{W_{room} - W_{sky}} T_{room}, \tag{275}$$

if we again assume $T_{atm} = T_{room}$.

6.8 Measurement of the Optical Depth of the Atmosphere τ_{atm} — $\sec z$ Method

The attenuation of the radio signal due to the atmosphere can be well represented by a model which assumes a plane–parallel atmospheric layer ignoring the curvature of the Earth's surface (Figure 108). In such a model, the optical depth of the atmosphere τ_{atm}, which is approximately proportional to the path length within the atmosphere, depends on the zenith distance z (which is the angle of a direction from the zenith direction) as

$$\tau_{atm} = \tau_0 \sec z. \tag{276}$$

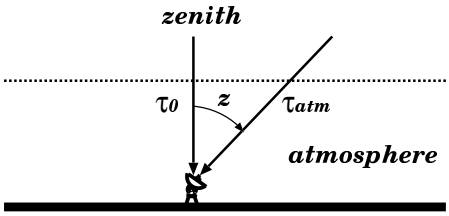

Figure 108: $\sec z$ dependence of the optical depth τ_{atm} in the plane–parallel model of the atmosphere.

Let us use the R–Sky method to measure the output powers when the feed horn is covered and uncovered by the absorber with the room temperature T_{room}, ignoring again the small T_{ant}. Then we have

$$\begin{aligned} W_{sky} &= k[T_{RX} + T_{atm}(1 - e^{-\tau_{atm}})]GB, \\ W_{room} &= k(T_{RX} + T_{room})GB. \end{aligned} \tag{277}$$

and, assuming again $T_{room} = T_{atm}$,

$$W_{room} - W_{sky} = kT_{room}e^{-\tau_{atm}}GB. \tag{278}$$

Therefore, taking the logarithm of equation (278), we obtain

$$-\ln(W_{room} - W_{sky}) = \tau_0 \sec z - \ln(kT_{room}GB), \tag{279}$$

in view of equation (276).

Therefore, if we measure the $W_{room} - W_{sky}$ at several values of zenith distance z, and if we assume that $T_{room}GB$ is constant during the measurement, then we obtain the zenith optical depth τ_0 as the inclination of a straight line fitted to the measured data in the $\sec z - [-\ln(W_{room} - W_{sky})]$ plane, as shown in Figure 109.

Knowing the τ_0 and therefore $\tau_{atm} = \tau_0 \sec z$, we can estimate the receiver noise temperature T_{RX} as well, using measured values of $T_S^* = T_{RX}e^{\tau_{atm}} + T_{atm}(e^{\tau_{atm}} - 1)$.

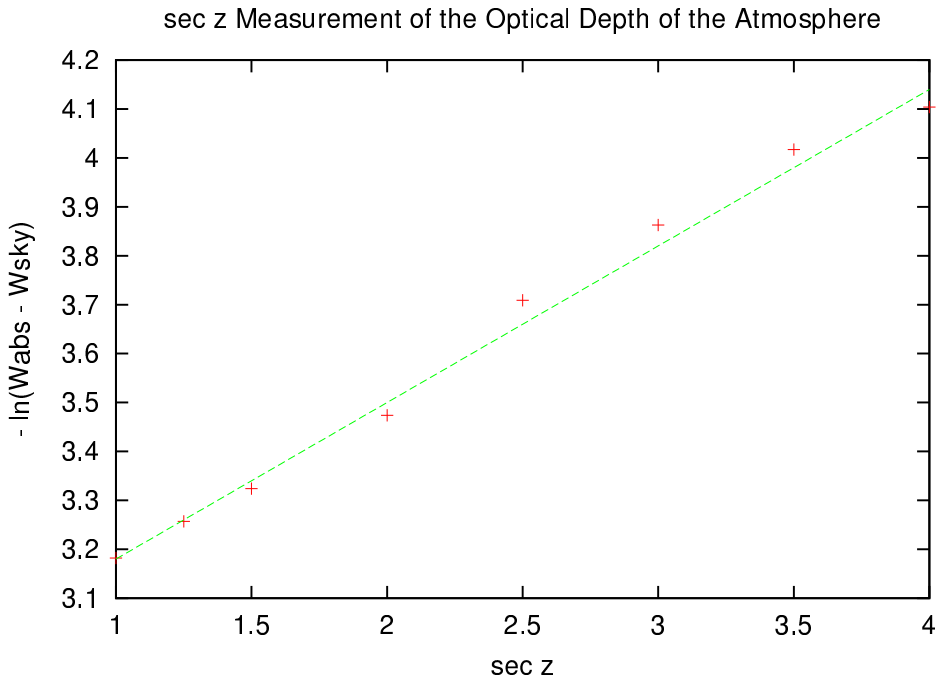

Figure 109: An example of the $\sec z - [-\ln(W_{room} - W_{sky})]$ plot of the measured data.

It is desirable to conduct the hot–cold measurement of the receiver noise temperature in parallel with the $\sec z$ measurement, in order to monitor the

possible gain variation effect during the measurement and check the consistency of the T_{RX} values estimated in the $\sec z$ and hot–cold methods.

It is highly desirable also to regularly measure the atmospheric optical depth, hopefully every day, for statistical study of the observational condition of the site. In particular, it will be useful to compare the measured relationship between the system temperature T_S^* and the optical depth $-\tau_{atm}$ (or the transparency $e^{-\tau_{atm}}$) during 1 year, say, with the theoretical prediction given in equation (263) and illustrated in Figure 110. We will then be able to see if the underlying assumptions of the equation well hold and T_{RX} is sufficiently stable for a year or so.

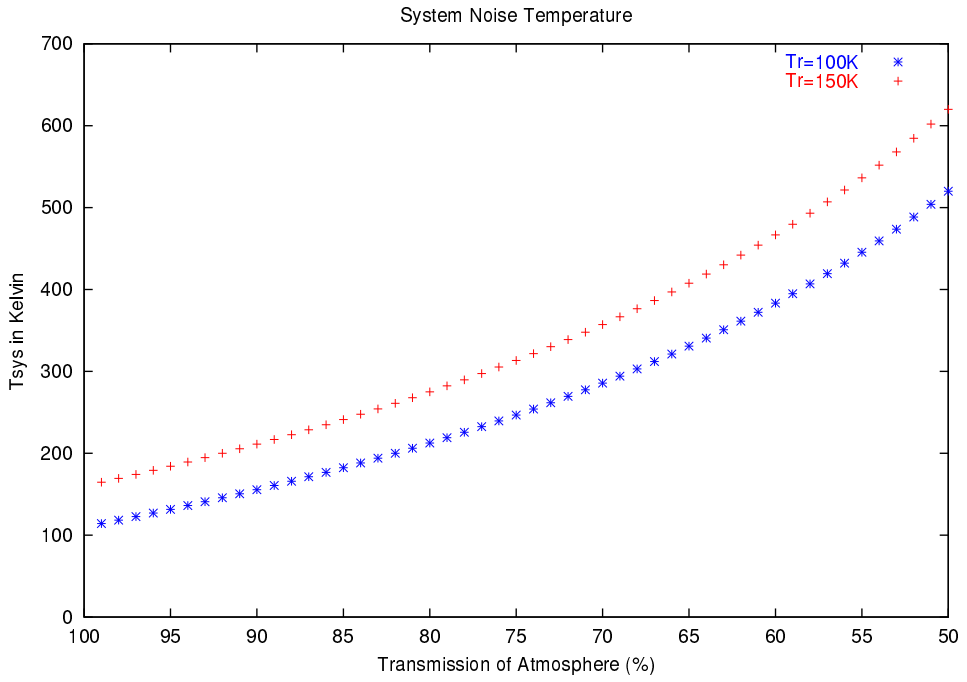

Figure 110: T_S^* – transparency $e^{-\tau_{atm}}$ relation expected from equation (263).

For efficient $\sec z$ measurements, we need a good automatic operation software, which would point the radio telescope to desired sky directions with different elevation angles, drive the "R–Sky" measuring instrument, read power–meter outputs at "room" and "sky" states, plot figures such as Figure 109, calculate τ_{atm}, T_S^* and T_{RX}, and store the results to an appropriate database.

6.9 Pointing Calibration and Pointing Accuracy σ_θ

Even in the well manufactured antennas, the azimuth axes may not be exactly parallel to the vertical lines at their sites. Likewise, the elevation axes may not be strictly perpendicular to the azimuth axes. In the antenna specifications, tolerance values for the axis–alignments are usually set to be at $0.^\circ 01$ level, but such a large axis–offset could easily cause pointing errors (offsets) exceeding the accuracy requirement of $\sim 0.1\lambda/D$ in equation (213). Also, the gravitational yielding of the antennas causes noticeable pointing errors mainly in the elevation direction. These pointing errors are "systematic" in the sense, that we obtain the same error values repeatedly in the same direction of the sky. But this means that they are predictable as functions of the azimuth and the elevation of the telescope beam direction.

Therefore, these systematic errors must be estimated and calibrated before starting regular scientific observations. The pointing accuracy σ_θ, which characterizes more or less random errors of the telescope pointing due to mechanical inaccuracy of gears, bearings, as well as unpredictable deformations caused by the gravity and the wind, must be determined after removing the systematic deviations. Actually, the pointing accuracy is calculated from the residuals of the least–squares fitting of a number of parameters, characterizing systematic errors, to the measured values of the pointing errors. Hereafter, we call these parameters as 'offset parameters'.

Since the offset parameters may vary in time due to the mechanical change in the antenna structure, the measurements must be repeated periodically once a year or more, for keeping the best system performance.

6.9.1 Pointing Model

Let us consider a model which describes a functional dependence of the systematic pointing error on offset parameters, assuming that we observe a radio source located in azimuth Az and elevation Ez with an Alt–Azimuth mount antenna. Here we adopt a convention that the azimuth is measured eastward from the North, i.e., $Az = 0^\circ$ at the North, and $Az = 90^\circ$ at the East.

Following 8 quantities are usually regarded as most important offset parameters for Alt–Azimuth mount antennas:

1. tilt angle of the azimuth axis in the North–South direction : a
2. tilt angle of the azimuth axis in the East–West direction : b
3. offset of the origin of the azimuth encoder (angle detection device), plus tilt angle of the elevation axis in the azimuthal direction : c

4. tilt angle of the elevation axis in the elevational direction : d
5. offset of the origin of the elevation encoder, plus tilt angle of the beam axis in the elevational direction : e
6. tilt angle of the beam axis in the azimuthal direction : f
7. cosine coefficient of the gravitational yielding in elevation : g
8. sine coefficient of the gravitational yielding in elevation : h.

Parameters $a - f$ represent axis offset effects, i.e., effects of misalignments of telescope axes, while parameters g and h represent elastic deformation effects. In the present level of antenna manufacturing technology, these offset parameters are quantities of the order of a few minutes of arc, or smaller. Therefore, it is sufficient to consider the pointing model in a linear approximation with respect to the small offset parameters.

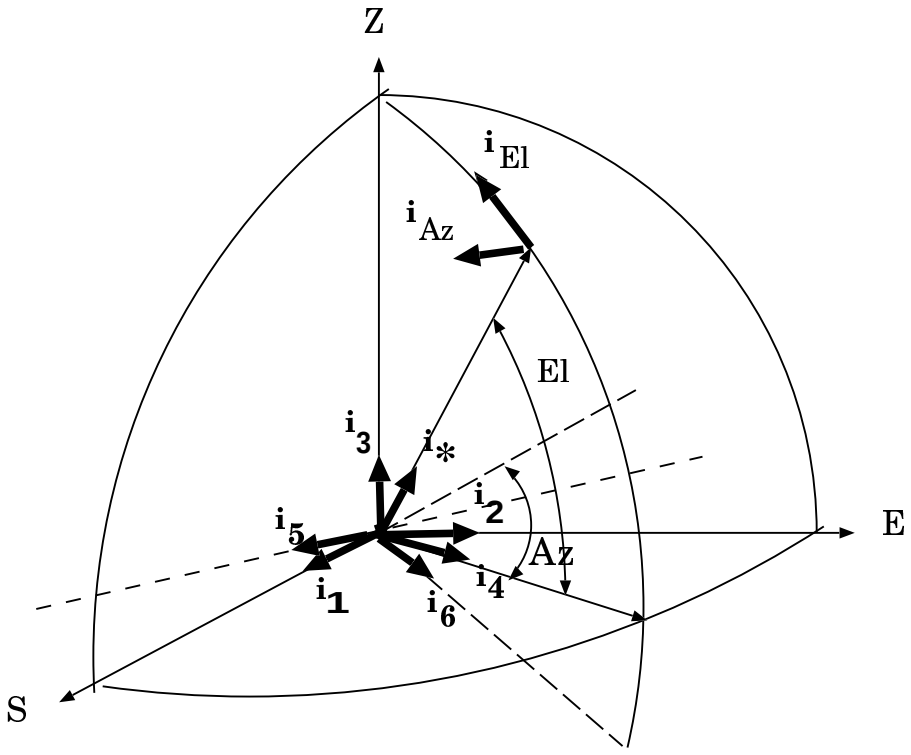

Figure 111: Unit vectors for describing the axis–offsets of an antenna.

The axis–offset angles a, b, c, d, e and f correspond to small rotation angles around directions shown by the unit vectors $\boldsymbol{i_1}$, $\boldsymbol{i_2}$, $\boldsymbol{i_3}$, $\boldsymbol{i_4}$, $\boldsymbol{i_5}$ and $\boldsymbol{i_6}$ in

Figure 111. Here we adopt the 'right–handed screw rule' for the direction of the rotation, i.e., the positive rotation angle corresponds to the clock–wise rotation, viewed towards the direction of the unit vector.

In Figure 111, $\boldsymbol{i_1}$, $\boldsymbol{i_2}$, and $\boldsymbol{i_3}$ are basis vectors of the horizontal coordinate system at the antenna site. $\boldsymbol{i_3}$ is directed towards the zenith, while $\boldsymbol{i_1}$ and $\boldsymbol{i_2}$ are directed towards the South and the East, respectively, in the horizontal plane. The $\boldsymbol{i_3}$ is the vertical axis, towards which the azimuth axis of the antenna must be directed, in an ideal case when there is no systematic axis offset.

$\boldsymbol{i_4}$ is directed towards azimuth Az in the horizontal plane, which is the azimuth of the radio source at the time of the observation.

$\boldsymbol{i_5}$ is chosen towards azimuth $Az + 90^\circ$ in the horizontal plane. The $\boldsymbol{i_5}$ is the direction, along which the elevation axis of the antenna must be aligned, in the ideal case when there is no systematic axis offset.

$\boldsymbol{i_6}$ is chosen in a direction with azimuth Az and elevation $El - 90^\circ$, where El is the elevation of the radio source at the time of the observation.

$\boldsymbol{i_5}$, $\boldsymbol{i_6}$, and a unit vector $\boldsymbol{i_\star}$, which is directed towards the radio source with azimuth Az and elevation El, form a right–handed orthogonal triad. The $\boldsymbol{i_\star}$ is the direction, towards which the beam axis of the antenna must be oriented, in the ideal case when there is no systematic axis offset.

Unit vectors $\boldsymbol{i_{Az}}$ and $\boldsymbol{i_{El}}$, which are directed towards azimuthal and elevational directions in a plane tangent to the celestial sphere at the radio source position, are related to $\boldsymbol{i_5}$ and $\boldsymbol{i_6}$, as

$$\begin{aligned} \boldsymbol{i}_5 &= \boldsymbol{i}_{Az}, & (280)\\ \boldsymbol{i}_6 &= -\boldsymbol{i}_{El}, & (281) \end{aligned}$$

as evident from Figure 111.

In the linear approximation, the pointing error $\Delta\boldsymbol{i_\star}$ is expressed in terms of the offset of the unit vector $\boldsymbol{i'_\star}$, which represents the direction of the actual telescope beam axis, from the one in the desired direction $\boldsymbol{i_\star}$:

$$\Delta\boldsymbol{i}_\star = \boldsymbol{i}'_\star - \boldsymbol{i}_\star. \qquad (282)$$

The azimuthal and elevational components of the pointing error can be expressed, in the linear approximation, as:

$$\begin{aligned} \cos El\, \Delta Az &= \Delta\boldsymbol{i}_\star \cdot \boldsymbol{i}_{Az},\\ \Delta El &= \Delta\boldsymbol{i}_\star \cdot \boldsymbol{i}_{El}. & (283) \end{aligned}$$

These are the quantities which we can measure in "pointing measurements", as we will see later.

Now let us derive the functional dependence of these azimuthal and elevational components of the pointing error on the offset parameters, which is nothing but the pointing model we are looking for.

In the linear approximation, we can separately consider the effects of the axis offsets and the elastic deformation, and then sum up the results to obtain the total effect. Therefore, we first consider the axis offset effects, assuming that the radio telescope antenna is a rigid body.

From the theory of the rotational dynamics of the rigid body, we know that, if a rigid body rotates around an axis $\boldsymbol{i}$, where $\boldsymbol{i}$ is a unit vector, by a small angle $\Delta\theta$, any vector $\boldsymbol{x}$ fixed to the rigid body is displaced by:

$$\Delta\boldsymbol{x} = \Delta\theta\boldsymbol{i} \times \boldsymbol{x}, \tag{284}$$

in a linear approximation with respect to $\Delta\theta$. We can introduce here a vector of small angle rotation $\Delta\boldsymbol{\Theta}$ as:

$$\Delta\boldsymbol{\theta} = \Delta\theta\,\boldsymbol{i}. \tag{285}$$

Then, equation (284) can be expressed as

$$\Delta\boldsymbol{x} = \Delta\boldsymbol{\theta} \times \boldsymbol{x}. \tag{286}$$

It is known that the vectors of small angle rotations can be summed up as ordinary vectors in the linear approximation with respect to the small angles. Since, in our present problem, a rotational displacement of the radio telescope antenna is composed of rotations by small angles a, b, $\cdots$, f around the six unit vectors $\boldsymbol{i}_1$, $\boldsymbol{i}_2$, $\cdots$, $\boldsymbol{i}_6$, the total rotation angle vector of the antenna $\Delta\boldsymbol{\Theta}$ is expressed by:

$$\Delta\boldsymbol{\Theta} = a\boldsymbol{i}_1 + b\boldsymbol{i}_2 + c\boldsymbol{i}_3 + d\boldsymbol{i}_4 + e\boldsymbol{i}_5 + f\boldsymbol{i}_6. \tag{287}$$

The angular offset of the telescope beam axis $\Delta\boldsymbol{i}_\star^a$ from the desired direction $\boldsymbol{i}_\star$, due to the superposed axis offset effects, is then expressed as:

$$\Delta\boldsymbol{i}_\star^a = \Delta\boldsymbol{\Theta} \times \boldsymbol{i}_\star. \tag{288}$$

Now, it is evident from Figure 111 and equations (280), (281) that

$$\boldsymbol{i}_3 = \sin El\,\boldsymbol{i}_\star + \cos El\,\boldsymbol{i}_{El}, \tag{289}$$

$$\boldsymbol{i}_4 = \cos El\,\boldsymbol{i}_\star - \sin El\,\boldsymbol{i}_{El}, \tag{290}$$

and

$$\begin{aligned}
\boldsymbol{i}_1 &= \cos(180^\circ - Az)\,\boldsymbol{i}_4 + \sin(180^\circ - Az)\,\boldsymbol{i}_5 \\
&= -\cos Az\,\boldsymbol{i}_4 + \sin Az\,\boldsymbol{i}_5 \\
&= -\cos Az\,\cos El\,\boldsymbol{i}_\star + \cos Az\,\sin El\,\boldsymbol{i}_{El} + \sin Az\,\boldsymbol{i}_{Az}, \qquad (291) \\
\boldsymbol{i}_2 &= \sin(180^\circ - Az)\,\boldsymbol{i}_4 - \cos(180^\circ - Az)\,\boldsymbol{i}_5 \\
&= \sin Az\,\boldsymbol{i}_4 + \cos Az\,\boldsymbol{i}_5 \\
&= \sin Az\,\cos El\,\boldsymbol{i}_\star - \sin Az\,\sin El\,\boldsymbol{i}_{El} + \cos Az\,\boldsymbol{i}_{Az}. \qquad (292)
\end{aligned}$$

Taking into account that

$$\begin{aligned}
\boldsymbol{i}_\star \times \boldsymbol{i}_\star &= 0, \\
\boldsymbol{i}_{Az} \times \boldsymbol{i}_\star &= \boldsymbol{i}_{El}, \qquad (293) \\
\boldsymbol{i}_{El} \times \boldsymbol{i}_\star &= -\boldsymbol{i}_{Az},
\end{aligned}$$

we obtain, from equations (280), (281), (289), (290), (291), and (292),

$$\begin{aligned}
\boldsymbol{i}_1 \times \boldsymbol{i}_\star &= -\cos Az\,\sin El\,\boldsymbol{i}_{Az} + \sin Az\,\boldsymbol{i}_{El}, \qquad (294) \\
\boldsymbol{i}_2 \times \boldsymbol{i}_\star &= \sin Az\,\sin El\,\boldsymbol{i}_{Az} + \cos Az\,\boldsymbol{i}_{El}, \qquad (295) \\
\boldsymbol{i}_3 \times \boldsymbol{i}_\star &= -\cos El\,\boldsymbol{i}_{Az}, \qquad (296) \\
\boldsymbol{i}_4 \times \boldsymbol{i}_\star &= \sin El\,\boldsymbol{i}_{Az}, \qquad (297) \\
\boldsymbol{i}_5 \times \boldsymbol{i}_\star &= \boldsymbol{i}_{El}, \qquad (298) \\
\boldsymbol{i}_6 \times \boldsymbol{i}_\star &= \boldsymbol{i}_{Az}. \qquad (299)
\end{aligned}$$

Therefore, in view of equations (287) and (288), we have the following expression for the beam offset due to the axis offset effects:

$$\begin{aligned}
\Delta\boldsymbol{i}^a_\star &= \Delta\boldsymbol{\Theta} \times \boldsymbol{i}_\star \\
&= a\,(-\cos Az\,\sin El\,\boldsymbol{i}_{Az} + \sin Az\,\boldsymbol{i}_{El}) \\
&\quad + b\,(\sin Az\,\sin El\,\boldsymbol{i}_{Az} + \cos Az\,\boldsymbol{i}_{El}) \\
&\quad - c\,\cos El\,\boldsymbol{i}_{Az} + d\,\sin El\,\boldsymbol{i}_{Az} + e\,\boldsymbol{i}_{El} + f\,\boldsymbol{i}_{Az}. \qquad (300)
\end{aligned}$$

Consequently, we obtain the azimuthal and elevational components of the beam offset due to the axis offset effects, using equation (283),

$$\begin{aligned}
\cos El\,\Delta Az^a &= \Delta\boldsymbol{i}^a_\star \cdot \boldsymbol{i}_{Az} \\
&= -a\,\cos Az \sin El + b\,\sin Az \sin El - c\,\cos El + d\,\sin El + f, \\
\Delta El^a &= \Delta\boldsymbol{i}^a_\star \cdot \boldsymbol{i}_{El} \\
&= a\,\sin Az \cos Az + b\,\cos Az + e. \qquad (301)
\end{aligned}$$

The main terms of the elastic deformation effect, which are the cosine and sine of elevation terms due to the gravitational yielding of the radio telescope antenna, are just added to the elevational component of the above equations (301), and therefore, we obtain the total beam offset:

$$\begin{aligned} \cos El\,\Delta Az &= -a\,\cos Az \sin El + b\,\sin Az \sin El - c\,\cos El + d\,\sin El + f, \\ \Delta El &= a\,\sin Az \cos Az + b\,\cos Az + e + g\cos El + h \sin El. \end{aligned} \quad (302)$$

Note that signs of the terms in the above equations may differ in different literature, depending on senses of the rotation angles.

6.9.2 Pointing Measurement

Now, if we measure the azimuthal and elevational offsets of the actual beam direction from the desired one at many directions, we can fit the equation (302) to the measured values to estimate the 8 parameters from a to g and, from the residuals, the pointing accuracy σ_θ, by the method of the least squares. An example of the sky coverage of the water maser sources observed in a pointing measurement at 22 GHz is shown in Figure 112.

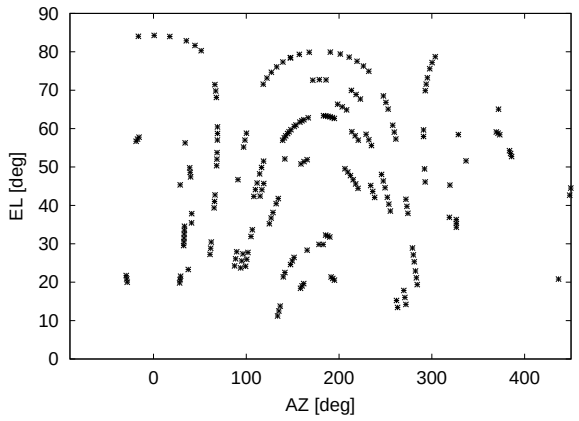

Figure 112: An example of the sky coverage of the sources in a pointing measurement.

Sometimes it is difficult to solve all 8 parameters due to the strong correlations among some of the parameters. In such a case, we have to assume reasonable values for some of the parameters and solve the least squares equations for rest of the parameters.

It is empirically known, for some of the wheel & track type antennas, that inclusion of additional terms with $\cos(2Az)$ and $\sin(2Az)$ dependences

into the pointing model better fits the results of the pointing measurements. In such a case, we have to increase the number of offset parameters to 12, or so.

The measurements are usually conducted by so–called "5 points scan" method. In this method, power from a radio source with well determined celestial position is measured at a telescope direction determined by a certain set of values of the azimuth– and elevation encoders (or angular scales) of the telescope, which are predicted by the source tracking software and instructed by the antenna control software, and at 4 more directions in North, South, East and West sides of the instructed one with the angular separation of roughly a half of the HPBW (Figure 113).

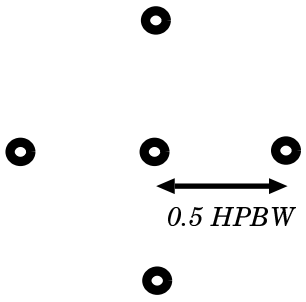

Figure 113: 5 points scan.

Given that the celestial positions and predicted Az and El values are accurate enough (at a level better than 1 arcsecond, say), the power must be strongest at the instructed direction and equally lower at four other directions, if the telescope does not have any systematic axis–offsets and gravitational yielding. If not, however, we will get an asymmetrical distribution of the measured powers at the 5 points. It is a usual practice to fit a two–dimensional Gaussian beam model to the measured power values to derive the offsets $\cos El \Delta Az$ and ΔEl between the actual and required beam directions which will be fitted by the model given in equation (302).

To achive the required accuracy of the prediction, the source tracking software must properly take into account major astrometric and geophysical effects including the abberation, the precession–nutation and the atmospheric refraction effects.

The radio sources used for the pointing measurements must be compact, strong enough, and must have well determined values of the position coordinates. Strong astronomical maser sources are often used for frquency bands, where such masers are available. For maser sources, we can easily obtain the antenna temperature value by comparing the 'in–line" part and

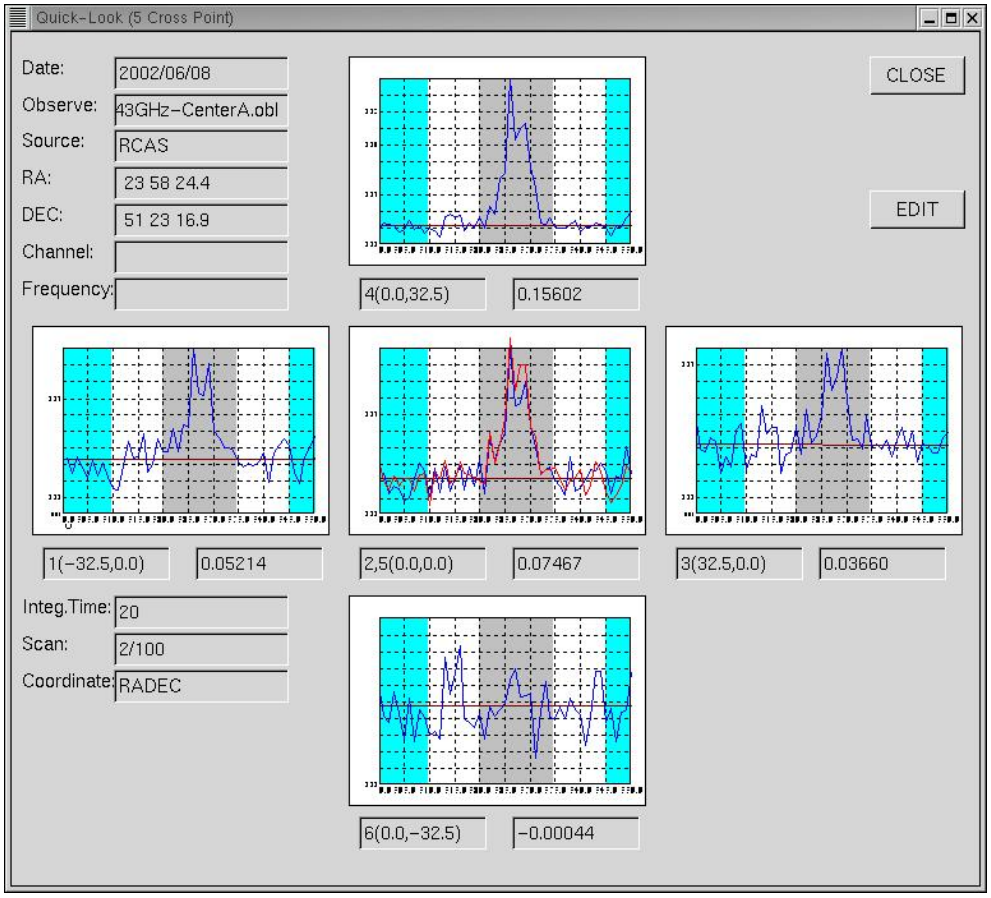

Figure 114: 5 points scan result at 43 GHz for SiO maser source R Cas obtained at Iriki station of VERA which shows southward offset of the beam.

the 'out–of–line" part of the frequency spectrum, as shown in Figure 114. For other frequency bands, strong continuum sources such as quasars or supernova remnants are used with appropriate antenna temperature measuring instruments.

For efficient pointing measurements, we need a good scheduling software and a well automated antenna operation software, which would enable us to quickly observe 'pointing sources' one by one at different sky directions, performing 5 points scans and automatically getting pointing offset values by the Gaussian fitting. In the case of the maser sources, we must correctly choose the "in–line" frequency range of the maser spectrum, where a desired line component is located. Since the range periodically shifts in frequency due to the Doppler effect assocaited with the orbital motion of the Earth, we need a software which calculates the "in–line" frequency range of date, from the local–standard–of–rest frequency value (or from the radial velocity v_{LSR}, referred to the Local Standard of Rest).

In massive star forming regions, maser features are often spread over a wide area of the sky, exceeding tens of arcseconds in size. Therefore, we have to choose the "in–line" range, so that it covers a strong maser feature or features, which are concentrated within a compact region much smaller than the antenna beam.

6.10 Beam Pattern Measurement

Power pattern of an antenna beam can be derived by measuring received power (or antenna temperature) of a compact strong radio source, in terms of the "grid–mapping" method (Figure 115) or the "drift–scan" method (Figure 116).

In the grid–mapping method, the received power values are measured at centers of a number of grids in the sky, chosen around the source with equal spacing, typically corresponding to a half of the expected beam size, in right ascension and declination. The results are analyzed with standard single–dish imaging softwares, and two–dimensional beam patterns are obtained. An example is shown in the lower panel of Figure 115.

In the drift–scan method, the received power is measured while a source passes in front of the antenna beam due to the diurnal motion, or when the antenna is driven, so that the beam crosses the source in azimuth and elevation directions. In this case, one dimensional beam pattern, as shown in Figure 116, is obtained at each scan.

In the both methods, the antenna pointing must be well calibrated, and radio source positions must be precisely known. Also, the measurements must be done within a short time under the clear sky, in order to avoid atmospheric

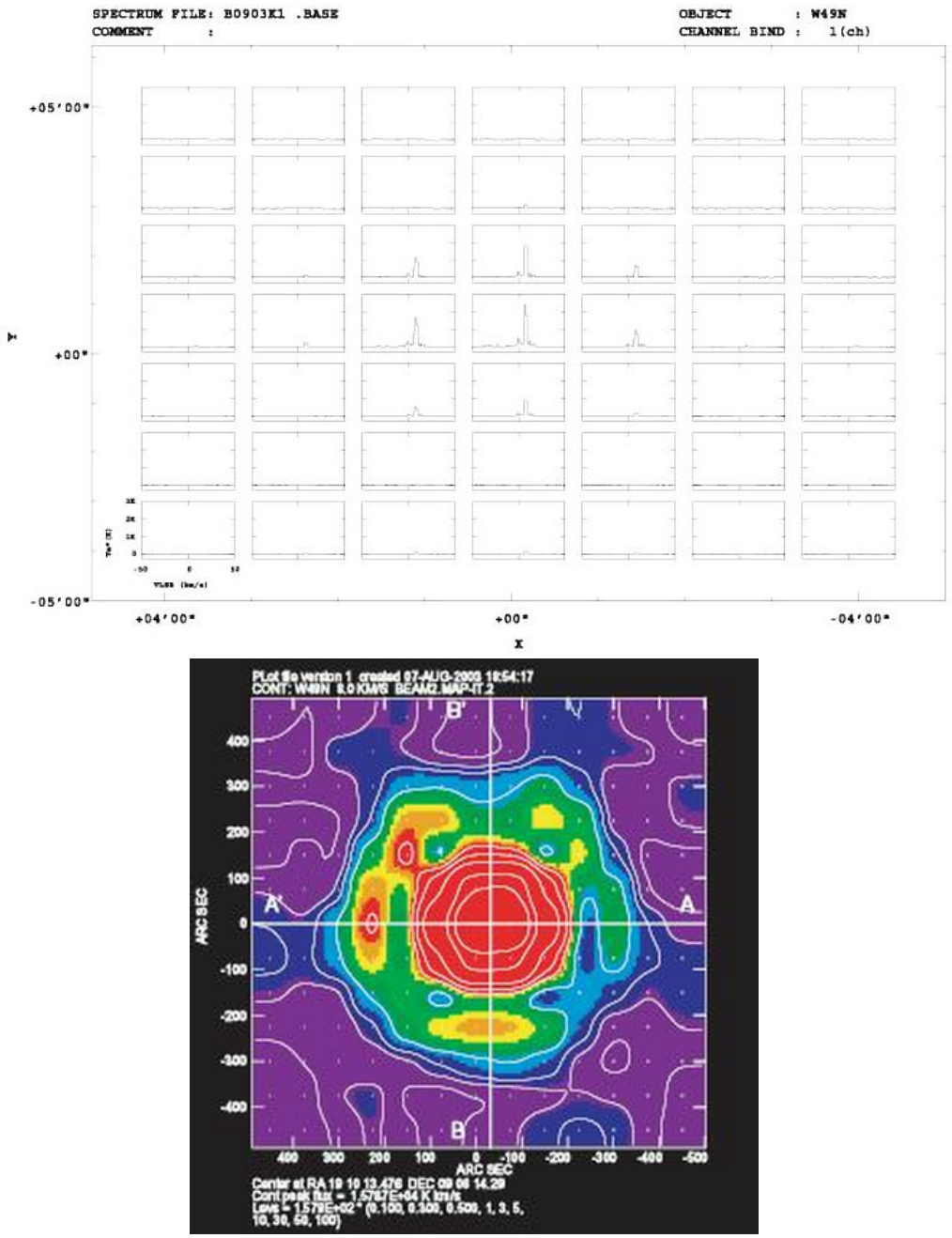

Figure 115: Grid–mapping method for beam pattern measurement (top), and a resultant beam pattern (bottom) (courtesy of the VERA group, 2004).

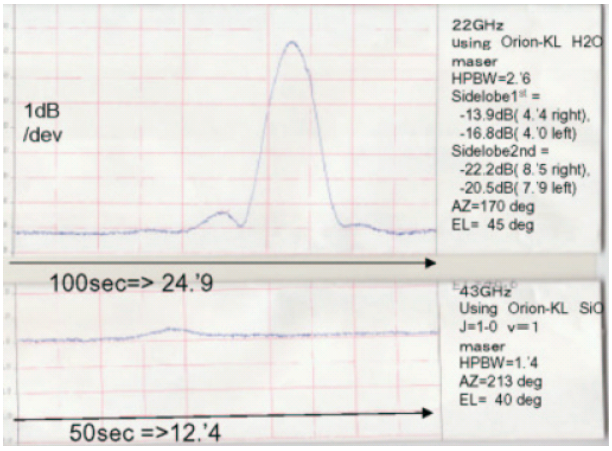

Figure 116: Beam patterns obtained with the drift–scan method at 22 GHz (top) and at 43 GHz (bottom) (courtesy of the VERA group, 2002).

disturbances. It is highly desirable to have automated antenna–operation, data acquisition and analysis softwares for the beam pattern measurements.