1 Overview, Maxwell's equations

- **ECE 329** introduced the **Maxwell's equations** and examined their circuit implications (inductance, capacitance) and TEM plane-wave solutions in homogeneous media and on "two-wire" transmission lines.
- In **ECE 350** we continue our study of the solutions and applications of Maxwell's equations with a focus on:
  1. **Radiation** of *spherical* TEM waves from practical compact **antennas** (e.g., used in cell phones and wireless links).
  2. **Propagation**, **reflection**, and **interference** of TEM waves in 3D geometries.
  3. **Antenna reception** and link budgets in communication applications.
  4. **Dispersion** effects in frequency dependent propagation media.
  5. **Guided waves** in TEM, TE, and TM modes.
  6. Field fluctuations in enclosed **cavities** + **thermal noise** in fields and circuits

ECE 350 completes the introductory description of electromagnetic (EM) effects in our curriculum and prepares the student for specialization courses in EM (ECE 447, 452, 453, 454, 455, 457, 458, etc.) and applications.

***Maxwell's equations*:**

$$\begin{aligned}\nabla\cdot\mathbf{E} &= \frac{\rho}{\epsilon_o}\\ \nabla\cdot\mathbf{B} &= 0\\ \nabla\times\mathbf{E} &= -\frac{\partial\mathbf{B}}{\partial t}\\ \nabla\times\mathbf{B} &= \mu_o\mathbf{J}+\mu_o\epsilon_o\frac{\partial\mathbf{E}}{\partial t}.\end{aligned}$$

such that

$$\mathbf{F}=q(\mathbf{E}+\mathbf{v}\times\mathbf{B}),$$

with

$$\mu_o\equiv 4\pi\times 10^{-7}\,\frac{\mathbf{H}}{\mathbf{m}},$$

and

$$\epsilon_o=\frac{1}{\mu_o c^2}\approx\frac{1}{36\pi\times 10^9}\,\frac{\mathbf{F}}{\mathbf{m}},$$

in mksA units, where

$$c=\frac{1}{\sqrt{\mu_o\epsilon_o}}\approx 3\times 10^8\,\frac{\mathbf{m}}{\mathbf{s}}$$

is the speed of light in free space. (In Gaussian-cgs units $\frac{\mathbf{B}}{c}$ is used in place of B above, while $\epsilon_o=\frac{1}{4\pi}$ and $\mu_o=\frac{1}{\epsilon_o c^2}=\frac{4\pi}{c^2}$.)

Review:

Maxwell's Equations:

$$\begin{aligned}
\nabla \cdot \mathbf{D} &= \rho && \text{Gauss's law} \\
\nabla \cdot \mathbf{B} &= 0 \\
\nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} && \text{Faraday's law} \\
\nabla \times \mathbf{H} &= \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} && \text{Ampere's law}
\end{aligned}$$

where $\Rightarrow$

Microscopic applications:

- ρ and $\mathbf{J}$ describe compact (pointlike) sources,
- $\mathbf{D} = \epsilon_o \mathbf{E}$ and $\mathbf{B} = \mu_o \mathbf{H}$

Macroscopic applications:

- ρ and $\mathbf{J}$ describe smooth sources composed of free charge carriers,
- $\mathbf{D} = \epsilon \mathbf{E}$ and $\mathbf{B} = \mu \mathbf{H}$

  specified in the in frequency domain with ω dependent

  - ***permittivity*** ϵ and
  - ***permeability*** μ.

- Fields $\mathbf{E}$ and $\mathbf{B}$ determine how a "test charge" q with mass m, position $\mathbf{r}$, and velocity $\mathbf{v} \equiv \dot{\mathbf{r}} = \frac{d\mathbf{r}}{dt}$ *accelerates* in accordance with **Lorentz force**

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

and Newton's 2nd law

$$\mathbf{F} = \frac{d}{dt} m\mathbf{v}.$$

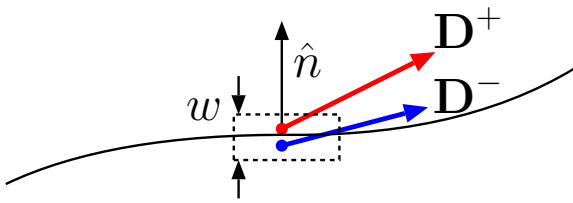
Maxwell's Equations:

$$\begin{aligned}
\nabla \cdot \mathbf{D} &= \rho \\
\nabla \cdot \mathbf{B} &= 0 \\
\nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \\
\nabla \times \mathbf{H} &= \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}
\end{aligned}$$

Boundary Conditions:

$$\begin{aligned}
\hat{n} \cdot [\mathbf{D}^+ - \mathbf{D}^-] &= \rho_s \\
\hat{n} \cdot [\mathbf{B}^+ - \mathbf{B}^-] &= 0 \\
\hat{n} \times [\mathbf{E}^+ - \mathbf{E}^-] &= 0 \\
\hat{n} \times [\mathbf{H}^+ - \mathbf{H}^-] &= \mathbf{J}_s
\end{aligned}$$

where $\hat{n}$ is a unit normal to the boundary surface pointing from $-$ to $+$ side.



Units in mksA system:

- q[=]C=sA,
- ρ[=]C/m^3,
- $\mathbf{J}$[=]A/m^2,
- $\mathbf{E}$[=]N/C=V/m,
- $\mathbf{D}$[=]C/m^2[=]ρ_s,
- $\mathbf{B}$[=]V.s/m^2 =Wb/m^2=T,
- $\mathbf{H}$[=]A/m[=]$\mathbf{J}_s$

where C, N, V, Wb, and T are abbreviations for *Coulombs, Newtons*, *Volts*, *Webers*, and *Teslas*, respectively.

Charge q is quantized in units of $e = 1.602 \times 10^{-19}$ C, a relativistic invariant.

- **Note:** the same units for
  - Displacement $\mathbf{D}$ and surface charge density ρ_s,
  - Magnetic field intensity $\mathbf{H}$ and surface current density $\mathbf{J}_s$.
- In right-handed Cartesian coordinates **div, grad**, and **curl** are produced by applying the del operator

$$\nabla \equiv (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}) = \hat{x}\frac{\partial}{\partial x} + \hat{y}\frac{\partial}{\partial y} + \hat{z}\frac{\partial}{\partial z}$$

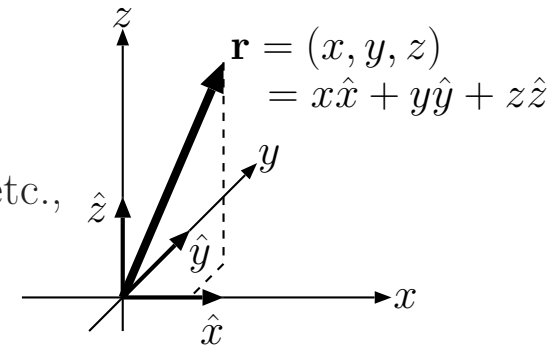
on vector or scalar fields as appropriate.

- Vectors and vector functions can be expressed in terms of mutually ***orthogonal*** **unit vectors** $\hat{x}$, $\hat{y}$, and $\hat{z}$ as in

  $\mathbf{r} = (x, y, z) = x\hat{x}+y\hat{y}+z\hat{z}$ and $\mathbf{E} = (E_x, E_y, E_z) = E_x\hat{x}+E_y\hat{y}+E_z\hat{z}$ etc.,

  where

  - $|\mathbf{r}| \equiv \sqrt{x^2+y^2+z^2}$ and $|\mathbf{E}| \equiv \sqrt{E_x^2+E_y^2+E_z^2}$ etc., are **vector magnitudes**,
  - $\hat{r} \equiv \frac{\mathbf{r}}{|\mathbf{r}|}$ and $\hat{E} \equiv \frac{\mathbf{E}}{|\mathbf{E}|}$ etc., are associated **unit vectors**, with



UNIT VECTORS AND A POSITION VECTOR IN RIGHT-HANDED CARETESIAN COORDINATES

Right handed convention: cross product vector points in the direction indicated by the thumb of your ***right hand*** when you rotate your fingers from vector **A** toward vector **B** through angle θ you decide to use.

Dot products:

- $\hat{r}\cdot\hat{r} = 1$, $\hat{E}\cdot\hat{E} = 1$, $\hat{x}\cdot\hat{x} = 1$, etc., but
- $\hat{x}\cdot\hat{y} = \hat{x}\cdot\hat{z} = \hat{y}\cdot\hat{z} = 0$.

Dot product $\mathbf{A}\cdot\mathbf{B}$ is a **scalar** which is the product of $|\mathbf{A}|$ and $|\mathbf{B}|$ and the cosine of angle θ between **A** and **B**.

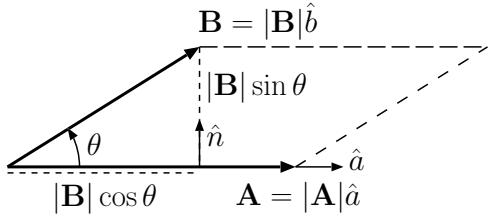
Dot product is zero when angle θ is 90°, as in the case of $\hat{x}$ and $\hat{y}$, etc.

Cross products:

- $\hat{x}\times\hat{y} = \hat{z}$,
  $\hat{y}\times\hat{z} = \hat{x}$,
  $\hat{z}\times\hat{x} = \hat{y}$ in a **right-handed** system.

Cross product $\mathbf{A}\times\mathbf{B}$ is a **vector** with a magnitude the product of $|\mathbf{A}|$ and $|\mathbf{B}|$ and the sine of angle θ between **A** and **B** and a direction orthogonal to **A** and **B** in a **right-handed** sense.

Cross product is zero when the vectors cross multiplied are collinear ($\theta = 0°$) or anti-linear ($\theta = 180°$).



$\mathbf{A}\cdot\mathbf{B} = |\mathbf{A}||\mathbf{B}|\cos\theta$
DOT PRODUCT:product of projected vector lengths

$\mathbf{A}\times\mathbf{B} = |\mathbf{A}||\mathbf{B}|\sin\theta\hat{a}\times\hat{n}$
CROSS PRODUCT: right-handed perpendicular area vector of the parallelogram formed by co-planar vectors

Example 1: A particle with charge $q = 1$ C passing through the origin $\mathbf{r} = (x, y, z) = 0$ of the lab frame is observed to accelerate with forces

$$\mathbf{F}_1 = 2\hat{x}, \quad \mathbf{F}_2 = 2\hat{x} - 6\hat{z}, \quad \mathbf{F}_3 = 2\hat{x} + 9\hat{y}\,\mathrm{N}$$

when the velocity of the particle is

$$\mathbf{v}_1 = 0, \quad \mathbf{v}_2 = 2\hat{y}, \quad \mathbf{v}_3 = 3\hat{z}\,\frac{\mathrm{m}}{\mathrm{s}},$$

in turns. Use the Lorentz force equation

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

to determine the fields $\mathbf{E}$ and $\mathbf{B}$ at the origin.

Solution: Using the Lorentz force formula first with $\mathbf{F} = \mathbf{F}_1$ and $\mathbf{v} = \mathbf{v}_1$, we note that

$$2\hat{x} = (1)(\mathbf{E} + 0 \times \mathbf{B}),$$

which implies that

$$\mathbf{E} = 2\hat{x}\,\frac{\mathrm{N}}{\mathrm{C}} = 2\hat{x}\,\frac{\mathrm{V}}{\mathrm{m}}.$$

Next, we use

$$\mathbf{v} \times \mathbf{B} = \frac{\mathbf{F}}{q} - \mathbf{E} = \frac{\mathbf{F}}{q} - 2\hat{x}$$

with $\mathbf{F}_2 = 2\hat{x} - 6\hat{z}$ and $\mathbf{v}_2 = 2\hat{y}$, as well as $\mathbf{E} = 2\hat{x}$ V/m, to obtain

$$2\hat{y} \times \mathbf{B} = -6\hat{z} \;\;\Rightarrow\;\; \hat{y} \times \mathbf{B} = -3\hat{z};$$

likewise, with $\mathbf{F}_3 = 2\hat{x} + 9\hat{y}$ and $\mathbf{v}_3 = 3\hat{z}$,

$$3\hat{z} \times \mathbf{B} = 9\hat{y} \;\;\Rightarrow\;\; \hat{z} \times \mathbf{B} = 3\hat{y}.$$

Substitute $\mathbf{B} = B_x\hat{x} + B_y\hat{y} + B_z\hat{z}$ in above relations to obtain

$$\hat{y} \times (B_x\hat{x} + B_y\hat{y} + B_z\hat{z}) = -B_x\hat{z} + B_z\hat{x} = -3\hat{z}$$

and

$$\hat{z} \times (B_x\hat{x} + B_y\hat{y} + B_z\hat{z}) = B_x\hat{y} - B_y\hat{x} = 3\hat{y}.$$

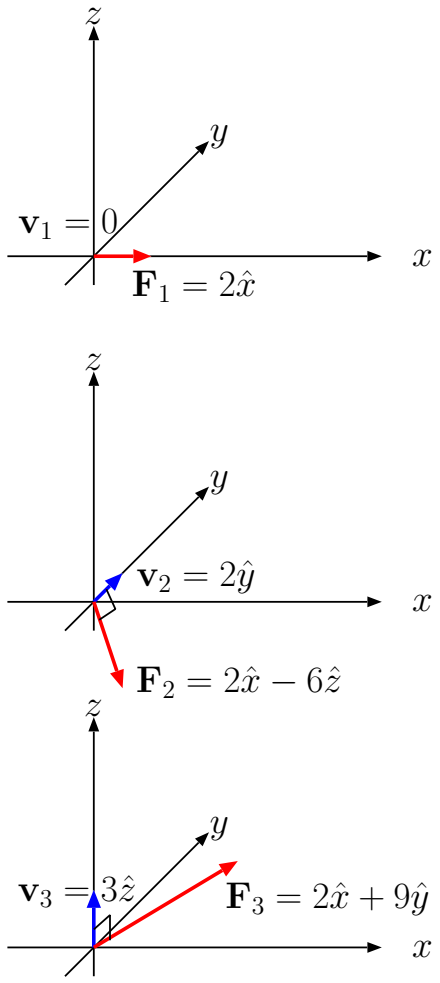
Matching the coefficients of $\hat{x}$, $\hat{y}$, and $\hat{z}$ in each of these relations we find that

$$B_x = 3\,\frac{\mathrm{Wb}}{\mathrm{m}^2}, \;\text{ and }\; B_y = B_z = 0.$$

Hence, vector

$$\mathbf{B} = 3\hat{x}\,\frac{\mathrm{Wb}}{\mathrm{m}^2}.$$

Having three non-collinear force measurements $\mathbf{F}_i$ corresponding to three distinct test particle velocities $\mathbf{v}_i$ is sufficient to determine the fields E and B at any location in space produced by distant sources as illustrated by this example.



Conservation laws

- In HW1 you are asked to derive the **continuity equation**

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0$$

by taking the divergence of Ampere's Law and combining it with Gauss' Law.

  - This equation expresses the *conservation of electrical charge* by putting a constraint on charge density ρ and current density $\mathbf{J}$ as it was first explained in ECE 329 (this is just a review, recall).

- Another conservation law derived in ECE 329 from Maxwell's equations was **Poynting Theorem**, namely

$$\frac{\partial w}{\partial t} + \nabla \cdot \mathbf{S} = -\mathbf{J} \cdot \mathbf{E},$$

where

$$w = \frac{1}{2}\epsilon_o \mathbf{E} \cdot \mathbf{E} + \frac{1}{2}\mu_o \mathbf{H} \cdot \mathbf{H} \quad \text{EM energy density,}$$

$$\mathbf{S} \equiv \mathbf{E} \times \mathbf{H} \quad \text{Poynting vector,}$$

$$-\mathbf{J} \cdot \mathbf{E} \quad \text{power produced per unit volume,}$$

  - expressing the *conservation of electromagnetic energy.*

- All conservation laws found in nature can be expressed mathematically in the forms given above in terms of a time-derivative of the volumetric density of the conserved quantity, the divergence of the flux of the conserved quantity (the so-called transport term), and a production term on the right (zero in case of charge conservation).

- The above conservation laws account for the increase/decrease of the conserved quantity density in terms of *local* transport and production effects. Hence charge conservation, for instance, is a *local* conservation principle.

  - If charge density decreases at a location, it will increase at a neighboring location because of local transport between the locations — charge cannot disappear in one volume and appear simultaneously in another volume (satisfying a so-called *global* conservation principle) without having traveled between the volumes.
  - *All conservation laws* observed in nature are *local* (as opposed to *global*) in the sense just described — the proof for this very broad statement can be based on the principle of relativity[1].

[1]Note that if charge could travel between the volumes with an infinite speed, then "global conservation" as opposed to "local conservation" could have been a viable idea — however no object can travel faster than light according to the principle of relativity and thus conservation laws have to be necessarily local and have mathematical expressions similar to those given in the *continuity equation*. A more general (but simple) proof of the local nature of *all* conservation laws (based on special relativity) is given by *Feynman* (see "The character of physical law", 1965, MIT Press).

2 Static fields and potentials

Static fields

$$\mathbf{E} = \mathbf{E}(\mathbf{r}), \;\; \mathbf{D} = \mathbf{D}(\mathbf{r}), \;\; \mathbf{B} = \mathbf{B}(\mathbf{r}), \;\; \mathbf{H} = \mathbf{H}(\mathbf{r})$$

independent of the time variable t are produced by static source distributions

$$\rho = \rho(\mathbf{r}) \;\; \text{and} \;\; \mathbf{J} = \mathbf{J}(\mathbf{r})$$

which only depend on position vector $\mathbf{r} = (x, y, z)$. In case of static fields Maxwell's equations simplify and decouple as

Time-dependent:

$$\begin{aligned} \nabla \cdot \mathbf{D} &= \rho \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \\ \nabla \times \mathbf{H} &= \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \end{aligned}$$

$\overset{\Rightarrow}{\frac{\partial}{\partial t} = 0}$

Electrostatics: (curl-free)

$$\begin{aligned} \nabla \cdot \mathbf{D} &= \rho \\ \nabla \times \mathbf{E} &= 0 \\ \mathbf{D} &= \epsilon_o \mathbf{E} \end{aligned}$$

Magnetostatics: (divergence-free, solenoidal)

$$\begin{aligned} \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{H} &= \mathbf{J} \\ \mathbf{B} &= \mu_o \mathbf{H} \end{aligned}$$

Important vector identities:

- $\nabla \times (\nabla V) = 0$
- $\nabla \cdot (\nabla \times \mathbf{A}) = 0$
- $\nabla \times \nabla \times \mathbf{A} = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$.

Electrostatics: (curl-free)

$$\begin{aligned} \nabla \cdot \mathbf{D} &= \rho \\ \nabla \times \mathbf{E} &= 0 \\ \mathbf{D} &= \epsilon_o \mathbf{E} \end{aligned}$$

Since all curl-free fields can be expressed in terms of a scalar gradient, we choose

$$\mathbf{E} = -\nabla V,$$

where

$$V = V(x, y, z)$$

is called **electrostatic potential**.

Magnetostatics: (divergence-free)

$$\begin{aligned} \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{H} &= \mathbf{J} \\ \mathbf{B} &= \mu_o \mathbf{H} \end{aligned}$$

Since all divergence-free fields can be expressed in terms of a curl, we choose

$$\mathbf{B} = \nabla \times \mathbf{A}$$

where

$$\mathbf{A} = \mathbf{A}(x, y, z)$$

is called **vector potential**.

Electrostatics: (curl-free)

$$\begin{aligned} \nabla \cdot \mathbf{D} &= \rho \\ \nabla \times \mathbf{E} &= 0 \\ \mathbf{D} &= \epsilon_o \mathbf{E} \end{aligned}$$

such that

$$\mathbf{E} = -\nabla V.$$

Electrostatic potential

$$V = V(x, y, z)$$

signifies the kinetic energy available (i.e., stored potential energy) — total energy being $\frac{1}{2} m\mathbf{v} \cdot \mathbf{v} + qV$ — per unit charge in a static field measured from a convenient reference point (ground).

Magnetostatics: (divergence-free)

$$\begin{aligned} \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{H} &= \mathbf{J} \\ \mathbf{B} &= \mu_o \mathbf{H} \end{aligned}$$

such that

$$\mathbf{B} = \nabla \times \mathbf{A}.$$

If we apply the constraint $\nabla \cdot \mathbf{A} = 0$ — known as **Coulomb gauge** and discussed in more detail next lecture — then the **vector potential**

$$\mathbf{A} = \mathbf{A}(x, y, z)$$

can be interpreted as kinetic momentum $m\mathbf{v}$ available — total (canonical) momentum being $m\mathbf{v} + q\mathbf{A}$ — per unit charge in a static field.

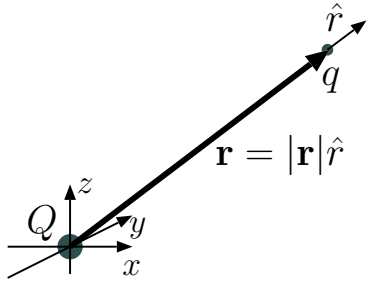
- In general, given V and $\mathbf{A}$, it is easy to compute $\mathbf{E}$ and $\mathbf{B}$.
- **How do we get V and $\mathbf{A}$ from ρ and $\mathbf{J}$?** Before addressing this question in full generality, let's review the electric field $\mathbf{E}$ and the electrostatic potential V of a stationary point charge located at the origin.

Coulomb's law specifies the electric field of a stationary charge Q at the origin as

$$\mathbf{E}(\mathbf{r}) = \frac{Q}{4\pi\epsilon_o|\mathbf{r}|^2}\hat{r}$$

as a function of position vector $\mathbf{r} = (x, y, z)$ with a magnitude

$$|\mathbf{r}| \equiv r = \sqrt{x^2 + y^2 + z^2} \quad \text{and direction unit vector} \quad \hat{r} = \frac{\mathbf{r}}{r}.$$



Force exerted by Q on q:

$$\mathbf{F} = q\mathbf{E}$$

with electric field

$$\mathbf{E} = \frac{Q}{4\pi\epsilon_o|\mathbf{r}|^2}\hat{r}$$

With multiple Q's superpose multiple E's

- This Coulomb field $\mathbf{E}(\mathbf{r})$ will exert a force $\mathbf{F} = q\mathbf{E}(\mathbf{r})$ on any stationary "test charge" q brought within distance r of Q (see margin).

- The associated electrostatic potential is

$$V(\mathbf{r}) = \frac{Q}{4\pi\epsilon_o|\mathbf{r}|}$$

with an implied ground for $|\mathbf{r}| \to \infty$.

Verification: this can be done in two ways,

1. by computing $-\nabla V \equiv \mathbf{E}(\mathbf{r})$, or
2. by computing the line integral $\int_{\mathbf{r}}^{\infty} \mathbf{E} \cdot d\mathbf{l} \equiv V(\mathbf{r})$ along any path.

In HW 1 we will ask you to verify the potential of the point charge using both methods.

Poisson's equations:

Electrostatics: Since

$$\nabla \times \mathbf{E} = 0 \quad \Rightarrow \quad \mathbf{E} = -\nabla V,$$

we have

$$\mathbf{D} = \epsilon_o \mathbf{E} \quad \text{and} \quad \nabla \cdot \mathbf{D} = \rho$$

implying

$$\nabla \cdot (-\epsilon_o \nabla V) = \rho \quad \Rightarrow \nabla^2 V = -\frac{\rho}{\epsilon_o}.$$

Magnetostatics: Since

$$\nabla \cdot \mathbf{B} = 0 \quad \Rightarrow \quad \mathbf{B} = \nabla \times \mathbf{A},$$

we have

$$\mathbf{B} = \mu_o \mathbf{H} \quad \text{and} \quad \nabla \times \mathbf{H} = \mathbf{J}$$

implying

$$\nabla \times (\mu_o^{-1} \nabla \times \mathbf{A}) = \mathbf{J} \quad \Rightarrow \nabla^2 \mathbf{A} = -\mu_o \mathbf{J}$$

after using

$$\nabla \cdot \mathbf{A} = 0 \quad (\textbf{Coulomb gauge})$$

in the expansion of

$$\nabla \times \nabla \times \mathbf{A} = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}.$$

- We can get V and $\mathbf{A}$ from ρ and $\mathbf{J}$ by solving the **Poisson's equations**

$$\nabla^2 V = -\frac{\rho}{\epsilon_o} \quad \text{and} \quad \nabla^2 \mathbf{A} = -\mu_o \mathbf{J}$$

where

$$\nabla^2 \equiv \nabla \cdot \nabla = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \text{ is } \textbf{Laplacian} \text{ operator.}$$

The solution of **electrostatic Poisson's equation**

$$\nabla^2 V = -\frac{\rho}{\epsilon_o}$$

with an arbitrary $\rho(\mathbf{r})$ existing over any finite region in space can be obtained as

$$V(\mathbf{r}) = \int \frac{\rho(\mathbf{r}')}{4\pi\epsilon_o|\mathbf{r}-\mathbf{r}'|} d^3\mathbf{r}'$$

where $d^3\mathbf{r}' \equiv dx'dy'dz'$ and the 3D volume integral on the right over the primed coordinates is performed over the entire region where the charge density is non-zero (see margin).

- **Verification:** The solution above can be verified by combining a number of results we have seen earlier on:

  1. Electric potential $V(\mathbf{r})$ of a point charge Q at the origin is

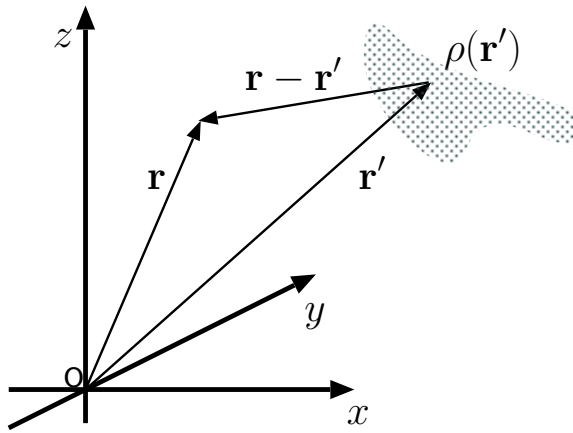
  $$V(\mathbf{r}) = \frac{Q}{4\pi\epsilon_o|\mathbf{r}|}.$$

  Clearly, this *singular* result is a solution of Poisson's equation above for a charge density input of

  $$\rho(\mathbf{r}) = Q\delta(\mathbf{r}).$$

  (a) Using ECE 210-like terminology and notation, the above result can be represented as

  $$\delta(\mathbf{r}) \rightarrow \boxed{\text{Poisson's Eqn}} \rightarrow \frac{1}{4\pi\epsilon_o|\mathbf{r}|}$$



The general solution

$$V(x,y,z)$$

is obtained by performing a 3D volume integral of

$$\frac{\rho(x',y',z')}{4\pi\epsilon_o|(x,y,z)-(x',y',z')|}$$

over the primed coordinates. In abbreviated notation

$$d^3\mathbf{r}' \equiv dx'dy'dz'$$

denotes an infinitesimal volume of the primed coordinate system.

identifying the output on the right as a 3D "impulse response" of the **linear** and **shift-invariant** (LSI) system represented by the Poisson's equation.

(b) Because of shift-invariance, we have

$$\delta(\mathbf{r}-\mathbf{r}') \rightarrow \boxed{\text{Poisson's Eqn}} \rightarrow \frac{1}{4\pi\epsilon_o|\mathbf{r}-\mathbf{r}'|},$$

meaning that a shifted impulse causes a shifted impulse response.

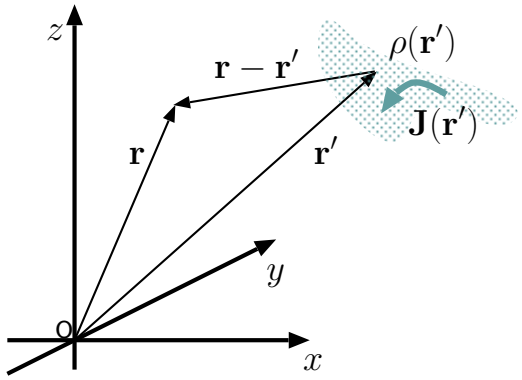
The shifted impulse response is usually called "Green's function" $G(\mathbf{r},\mathbf{r}')$ in EM theory.

(c) Because of linearity, we are allowed to use superpositioning arguments like

$$\int \rho(\mathbf{r}')\delta(\mathbf{r}-\mathbf{r}')d^3\mathbf{r}' = \rho(\mathbf{r}) \rightarrow \boxed{\text{Poisson's Eqn}} \rightarrow \int \rho(\mathbf{r}')\frac{1}{4\pi\epsilon_o|\mathbf{r}-\mathbf{r}'|}d^3\mathbf{r}' = V(\mathbf{r}),$$

which concludes our verification. Note how we made use of the *sifting property* of the impulse (from ECE 210) in above calculation.

Solutions of Poisson's equations:



Electrostatics:

$$\nabla^2 V = -\frac{\rho}{\epsilon_o}$$

implies a general solution

$$V(\mathbf{r}) = \int \frac{\rho(\mathbf{r}')}{4\pi\epsilon_o|\mathbf{r}-\mathbf{r}'|} d^3\mathbf{r}'.$$

Magnetostatics:

$$\nabla^2 \mathbf{A} = -\mu_o \mathbf{J}$$

implies a general solution

$$\mathbf{A}(\mathbf{r}) = \int \frac{\mu_o \mathbf{J}(\mathbf{r}')}{4\pi|\mathbf{r}-\mathbf{r}'|} d^3\mathbf{r}'.$$

These results indicate that potentials

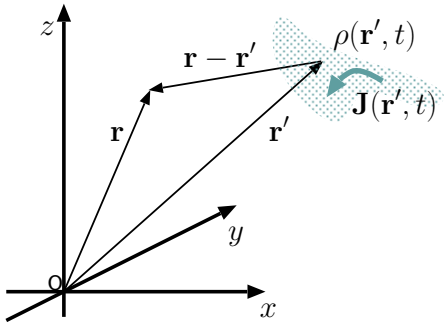
$$V(\mathbf{r}) \quad \text{and} \quad \mathbf{A}(\mathbf{r})$$

are appropriately weighted sums of

$$\rho(\mathbf{r}) \quad \text{and} \quad \mathbf{J}(\mathbf{r})$$

in convolution-like 3D space integrals.

Quasi-static approximation:



Electro-quasi-statics:

$$\mathbf{E} \approx -\nabla V$$

with

$$V(\mathbf{r},t) \approx \int \frac{\rho(\mathbf{r}',t)}{4\pi\epsilon_o|\mathbf{r}-\mathbf{r}'|}d^3\mathbf{r}'$$

for slowly varying $\rho(\mathbf{r},t)$.

Magneto-quasi-statics:

$$\mathbf{B} \approx \nabla \times \mathbf{A}$$

with

$$\mathbf{A}(\mathbf{r},t) \approx \int \frac{\mu_o\mathbf{J}(\mathbf{r}',t)}{4\pi|\mathbf{r}-\mathbf{r}'|}d^3\mathbf{r}'$$

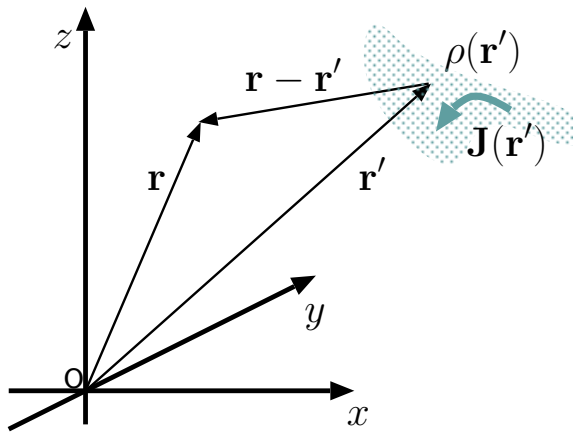
for slowly varying $\mathbf{J}(\mathbf{r},t)$.

Validity of these quasi-static results requires that

$$T \gg \frac{L}{c}$$

where T is the period of the highest frequency in source functions $\rho(\mathbf{r},t)$ and $\mathbf{J}(\mathbf{r},t)$, while L is the size of the region around the source region where quasi-static approximation is acceptable. This condition cannot be satisfied as $L \to \infty$, in which case the required "fix" is to replace the potential functions above by their "retarded potential" counterparts — see next lecture.

3 Lorenz gauge and inhomogeneous wave equation



Last lecture we found out that given the static sources

$$\rho = \rho(\mathbf{r}) \quad \text{and} \quad \mathbf{J} = \mathbf{J}(\mathbf{r}),$$

static fields

$$\mathbf{E} = -\nabla V \quad \text{and} \quad \mathbf{B} = \nabla \times \mathbf{A}$$

satisfying

Electrostatics: (curl-free)

$$\begin{aligned} \nabla \cdot \mathbf{D} &= \rho \\ \nabla \times \mathbf{E} &= 0 \\ \mathbf{D} &= \epsilon_o \mathbf{E} \end{aligned}$$

Magnetostatics: (divergence-free)

$$\begin{aligned} \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{H} &= \mathbf{J} \\ \mathbf{B} &= \mu_o \mathbf{H} \end{aligned}$$

can be computed using the potentials

$$V(\mathbf{r}) = \int \frac{\rho(\mathbf{r}')}{4\pi\epsilon_o |\mathbf{r}-\mathbf{r}'|} d^3\mathbf{r}',$$

the solution of

$$\nabla^2 V = -\frac{\rho}{\epsilon_o}.$$

$$\mathbf{A}(\mathbf{r}) = \int \frac{\mu_o \mathbf{J}(\mathbf{r}')}{4\pi |\mathbf{r}-\mathbf{r}'|} d^3\mathbf{r}',$$

the solution of

$$\nabla^2 \mathbf{A} = -\mu_o \mathbf{J}.$$

- Over the next two lectures we will explain why in case of time-varying sources

$$\rho = \rho(\mathbf{r}, t) \quad \text{and} \quad \mathbf{J} = \mathbf{J}(\mathbf{r}, t),$$

the full set of Maxwell's equations (see margin) can be satisfied by

$$\mathbf{E} = -\nabla V - \frac{\partial \mathbf{A}}{\partial t} \quad \text{and} \quad \mathbf{B} = \nabla \times \mathbf{A}$$

in terms of delayed or **retarded potentials** specified as

$$\begin{aligned} \nabla \cdot \mathbf{D} &= \rho \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \\ \nabla \times \mathbf{H} &= \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \end{aligned}$$

$$V(\mathbf{r}, t) = \int \frac{\rho(\mathbf{r}', t - \frac{|\mathbf{r}-\mathbf{r}'|}{c})}{4\pi\epsilon_o |\mathbf{r} - \mathbf{r}'|} d^3\mathbf{r}',$$

the solution of inhomogeneous wave equation

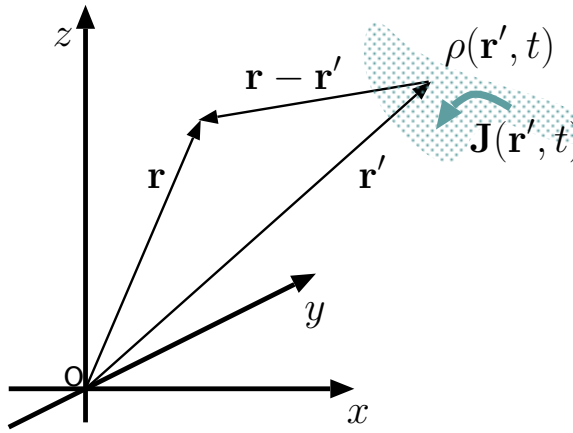
$$\nabla^2 V - \mu_o \epsilon_o \frac{\partial^2 V}{\partial t^2} = -\frac{\rho}{\epsilon_o}$$

$$\mathbf{A}(\mathbf{r}, t) = \int \frac{\mu_o \mathbf{J}(\mathbf{r}', t - \frac{|\mathbf{r}-\mathbf{r}'|}{c})}{4\pi |\mathbf{r} - \mathbf{r}'|} d^3\mathbf{r}',$$

the solution of inhomogeneous wave equation

$$\nabla^2 \mathbf{A} - \mu_o \epsilon_o \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu_o \mathbf{J}$$

where $c \equiv \dfrac{1}{\sqrt{\mu_o \epsilon_o}}$ is the speed of light in free space.



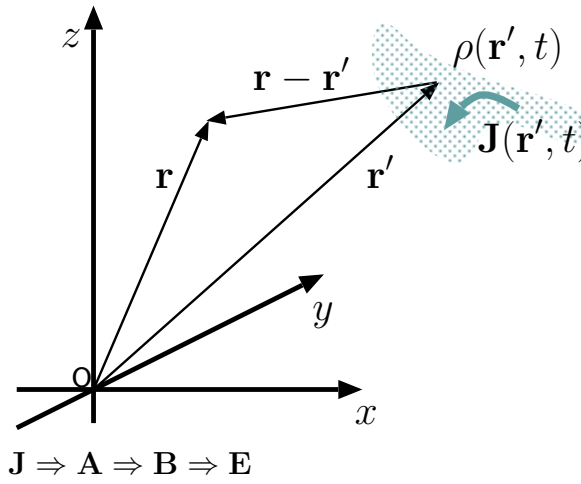
- Note that *retarded* potentials

$$V(\mathbf{r}, t) \quad \text{and} \quad \mathbf{A}(\mathbf{r}, t)$$

are essentially weighted and *delayed* sums of charge and current densities

$$\rho(\mathbf{r},t) \quad \text{and} \quad \mathbf{J}(\mathbf{r},t),$$

while the fields $\mathbf{E}$ and $\mathbf{B}$ are obtained by spatial and temporal derivatives of the potentials.



$\mathbf{J} \Rightarrow \mathbf{A} \Rightarrow \mathbf{B} \Rightarrow \mathbf{E}$

- Alternatively, we can first use

$$\mathbf{A}(\mathbf{r},t) = \int \frac{\mu_o \mathbf{J}(\mathbf{r}', t - \frac{|\mathbf{r}-\mathbf{r}'|}{c})}{4\pi|\mathbf{r}-\mathbf{r}'|} d^3\mathbf{r}' \quad \Rightarrow \quad \mathbf{B} = \nabla \times \mathbf{A}$$

and then find the anti-derivative of Ampere's law

$$\nabla \times \frac{\mathbf{B}}{\mu_o} = \epsilon_o \frac{\partial \mathbf{E}}{\partial t}$$

to determine $\mathbf{E}$ outside the region where $\mathbf{J}$ is non-zero, bypassing the use of scalar retarded potential $V(\mathbf{r},t)$ — that is the most common approach used in **radiation studies**.

We will next verify the procedure outlined above and the start discussing its applications in radiation studies.

$$\begin{aligned}
\nabla \cdot \mathbf{D} &= \rho \\
\nabla \cdot \mathbf{B} &= 0 \\
\nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \\
\nabla \times \mathbf{H} &= \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}
\end{aligned}$$

- The full set of Maxwell's equations is repeated in the margin for convenience. Divergence-free nature of $\mathbf{B}$ compels us to define a vector potential $\mathbf{A}$ via

$$\mathbf{B} = \nabla \times \mathbf{A}$$

just as before. Inserting this in Faraday's law we get

$$\nabla \times \mathbf{E} = -\frac{\partial}{\partial t} \nabla \times \mathbf{A} \quad \Rightarrow \quad \nabla \times (\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t}) = 0.$$

Evidently

$$\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} \text{ is curl free, so it must be true that } \mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} = -\nabla V,$$

or

$$\mathbf{E} = -\nabla V - \frac{\partial \mathbf{A}}{\partial t}$$

in terms of some scalar potential V.

Main difference from statics appears to be the need for *two* potentials, instead of one, to represent the electric field E under time-varying conditions. We continue ....

- Now substitute

$$\mathbf{B} = \nabla \times \mathbf{A} \text{ and } \mathbf{E} = -\nabla V - \frac{\partial \mathbf{A}}{\partial t}$$

in the remaining two Maxwell's equations — Gauss's and Ampere's laws

$$\nabla \cdot (\epsilon_o \mathbf{E}) = \rho \quad \text{and} \quad \nabla \times (\mu_o^{-1} \mathbf{B}) = \mathbf{J} + \frac{\partial}{\partial t}(\epsilon_o \mathbf{E}),$$

that we have not touched yet. Upon substitutions we get

$$\epsilon_o \nabla \cdot (-\nabla V - \frac{\partial \mathbf{A}}{\partial t}) \;=\; \rho \quad \text{and} \quad \underbrace{\nabla \times \nabla \times \mathbf{A}}_{\nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}} = \mu_o \mathbf{J} + \mu_o \epsilon_o \frac{\partial}{\partial t}(-\nabla V - \frac{\partial \mathbf{A}}{\partial t}),$$

which looks like a big mess.

- But if we specify

$$\nabla \cdot \mathbf{A} = -\mu_o \epsilon_o \frac{\partial V}{\partial t} \qquad (\textbf{Lorenz gauge})$$

these messy equations simplify as

$$\nabla^2 V - \mu_o \epsilon_o \frac{\partial^2 V}{\partial t^2} = -\frac{\rho}{\epsilon_o} \quad \text{and} \quad \nabla^2 \mathbf{A} - \mu_o \epsilon_o \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu_o \mathbf{J}$$

which we recognize as the inhomogeneous or "forced" wave equations for V and $\mathbf{A}$ stated earlier on.

  - The derivation of the decoupled wave equations above hinged upon our use of **Lorenz gauge** which reduces to the **Coulomb gauge**, $\nabla \cdot \mathbf{A} = 0$, in static situations.
  - Note also that the forced wave equations reduce to Poisson's equations under time-static conditions.

- Since we know how to solve the unforced wave equation from ECE 329, and since we know how to solve the Poisson's equation, it is now a matter of combining those methods to solve the forced wave equations obtained above.

Just a few additional comments on ***gauge selection*** before we go on (next lecture):

- Gauge selection amounts to deciding what to assign to $\nabla \cdot \mathbf{A}$.
- We can make any assignment that pleases us. This is like choosing the ground node in a circuit problem. Whatever simplifies the problem the most is the best gauge to use.

  - Lorenz gauge is clearly a good one since it led to decoupled wave equations which are very convenient to work with.

  We can attack the decoupled equations for V and $\mathbf{A}$ *one at a time*.

4 Time harmonic sources and *retarded* potentials

- The solution of forced wave equation

$$\nabla^2 V - \mu_o \epsilon_o \frac{\partial^2 V}{\partial t^2} = -\frac{\rho}{\epsilon_o}$$

for scalar potential V is most conveniently obtained in the frequency domain:

Consider a time-harmonic forcing function ρ and a time-harmonic response V expressed as

$$\rho(\mathbf{r}, t) = \mathrm{Re}\{\tilde{\rho}(\mathbf{r}) e^{j\omega t}\} \quad \text{and} \quad V(\mathbf{r}, t) = \mathrm{Re}\{\tilde{V}(\mathbf{r}) e^{j\omega t}\}$$

in terms of phasors

$$\tilde{\rho}(\mathbf{r}) \quad \text{and} \quad \tilde{V}(\mathbf{r}).$$

Then, the above wave equation transforms — upon replacing $\frac{\partial}{\partial t}$ by $j\omega$ — into phasor form as

$$\nabla^2 \tilde{V} + \mu_o \epsilon_o \omega^2 \tilde{V} = -\frac{\tilde{\rho}}{\epsilon_o}.$$

- For $\omega = 0$ the above equation reduces to Poisson's equation, which we know has, with an impulse forcing (for 1 C point charge at the origin)

$$\tilde{\rho}(\mathbf{r}) = \delta(\mathbf{r}), \quad \text{an impulse response solution } \tilde{V}(\mathbf{r}) = \frac{1}{4\pi\epsilon_o |\mathbf{r}|} \equiv \frac{1}{4\pi\epsilon_o r}$$

where $r \equiv |\mathbf{r}|$ denotes the distance of the observing point $\mathbf{r}$ from the impulse point located at the origin.

- Note that this impulse response $\propto 1/r$ is symmetric with respect to the origin just like the impulse input $\delta(\mathbf{r})$.

We now *postulate* and subsequently prove that for $\omega \geq 0$, the impulse response solution of the forced wave equation — i.e., with forcing function $\tilde{\rho}(\mathbf{r}) = \delta(\mathbf{r})$ — is

$$\tilde{V}(\mathbf{r}) = \frac{e^{-jk|\mathbf{r}|}}{4\pi\epsilon_o|\mathbf{r}|} \quad \text{with} \quad k \equiv \omega\sqrt{\mu_o\epsilon_o} = \frac{\omega}{c}.$$

Proof: For $\tilde{\rho}(\mathbf{r}) = \delta(\mathbf{r})$ the source of the forced wave equation (for an arbitrary ω) is *symmetric with respect to the origin*, implying that the corresponding solution $\tilde{V}(\mathbf{r})$ should also have the same type of symmetry. Then, with *no loss of generality*, we can claim a solution for the case $\tilde{\rho}(\mathbf{r}) = \delta(\mathbf{r})$ of the form

$$\tilde{V}(\mathbf{r}) = \frac{f(r)}{r}$$

where

- $f(r) = \frac{1}{4\pi\epsilon_o}$ for $\omega = 0$, and
- $f(r)$ is *to be determined* for an arbitrary ω as follows:

Note that by substituting the source function $\delta(\mathbf{r})$ and response function $\frac{1}{4\pi\epsilon_o r}$ back into the Poisson's equation we obtain an equality

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

which is a useful vector identity.

– Substituting $f(r)/r$ for $\tilde{V}(r)$ and $\delta(\mathbf{r})$ for $\rho(\mathbf{r})$ in the forced wave equation (see margin), we obtain

$$\nabla^2(\frac{f(r)}{r}) + k^2\frac{f(r)}{r} = -\frac{\delta(\mathbf{r})}{\epsilon_o}$$

which reduces, for $r \neq 0$, to

$$\nabla^2(\frac{f(r)}{r}) + k^2\frac{f(r)}{r} = 0.$$

– Since (as shown in HW) we have, by using spherical coordinates (reviewed next lecture),

$$\nabla^2(\frac{f(r)}{r}) = \frac{1}{r}\frac{\partial^2 f}{\partial r^2},$$

it follows that we have, for $r \neq 0$,

$$\frac{1}{r}(\frac{\partial^2 f}{\partial r^2} + k^2 f) = 0,$$

which is in turn satisfied by

$$f(r) = ge^{\mp jkr} = ge^{\mp j\omega r/c}$$

with an arbitrary constant g.

– Finally, the constraint that $f(r) = 1/4\pi\epsilon_o$ for $\omega = 0$ indicates that

$$g = \frac{1}{4\pi\epsilon_o},$$

Forced wave eqn (phasor form):

$$\nabla^2\tilde{V} + k^2\tilde{V} = -\frac{\tilde{\rho}}{\epsilon_o}$$

with

$$k = \omega\sqrt{\mu_o\epsilon_o} = \frac{\omega}{c}.$$

and thus

$$f(r) = \frac{e^{\mp jkr}}{4\pi\epsilon_o}$$

in the solutions $f(r)/r$ of the wave equation with $\tilde{\rho}(\mathbf{r}) = \delta(\mathbf{r})$.

This concludes our proof of the postulated solution

$$\tilde{V}(\mathbf{r}) = \frac{e^{-jk|\mathbf{r}|}}{4\pi\epsilon_o|\mathbf{r}|} \quad \text{with} \quad k \equiv \omega\sqrt{\mu_o\epsilon_o} = \frac{\omega}{c}$$

where the sign choice in the exponent favors the physically relevant *causal* solution as opposed to the acausal alternative (see discussion below).

The choice $-jkr$ leads to so-called *retarded* solution of the wave equation. The alternative choice $+jkr$ is not used because it leads to an *advanced* solution that depends on future values of the charge distribution not available in practice (this *causality* constraint is further discussed later in this lecture).

Note that k is another symbol for wavenumber β. In this and higher level courses in EM and signal processing k is favored over β (for a good number of reasons which will become apparent as we learn more).

For the record, by scaling the result above:

For $\tilde{\rho}(\mathbf{r}) = Q\delta(\mathbf{r})$, the causal solution of the forced wave equation

$$\nabla^2\tilde{V} + k^2\tilde{V} = -\frac{\tilde{\rho}}{\epsilon_o},$$

where $k \equiv \omega\sqrt{\mu_o\epsilon_o}$ is the phasor

$$\tilde{V}(\mathbf{r}) = \frac{Q}{4\pi\epsilon_o}\frac{e^{-jkr}}{r}.$$

Likewise, for $\tilde{J}_z(\mathbf{r}) = P\delta(\mathbf{r})$, the causal solution of the forced wave equation

$$\nabla^2\tilde{A}_z + k^2\tilde{A}_z = -\mu_o\tilde{J}_z,$$

where $k \equiv \omega\sqrt{\mu_o\epsilon_o}$ must be the phasor

$$\tilde{A}_z(\mathbf{r}) = \frac{\mu_o P}{4\pi}\frac{e^{-jkr}}{r},$$

which describes, with $P = I\Delta z$, the vector potential of the *Hertzian dipole* defined in Lecture 6.

- We can next argue as follows:

$$\delta(\mathbf{r}) \rightarrow \boxed{\text{Forced Wave Eqn}} \rightarrow \frac{e^{-jk|\mathbf{r}|}}{4\pi\epsilon_o|\mathbf{r}|}$$

and

$$\delta(\mathbf{r}-\mathbf{r}') \rightarrow \boxed{\text{Forced Wave Eqn}} \rightarrow \frac{e^{-jk|\mathbf{r}-\mathbf{r}'|}}{4\pi\epsilon_o|\mathbf{r}-\mathbf{r}'|}$$

imply that

$$\int \tilde{\rho}(\mathbf{r}')\delta(\mathbf{r}-\mathbf{r}')d^3\mathbf{r}' = \tilde{\rho}(\mathbf{r}) \rightarrow \boxed{\text{Forced Wave Eqn}} \rightarrow \int \frac{\tilde{\rho}(\mathbf{r}')e^{-jk|\mathbf{r}-\mathbf{r}'|}}{4\pi\epsilon_o|\mathbf{r}-\mathbf{r}'|}d^3\mathbf{r}' = \tilde{V}(\mathbf{r}),$$

giving us, on the right-hand side, the **retarded potential solution** in the *frequency domain*.

- Finally, inverse Fourier transforming the above result back to time domain, we obtain

$$V(\mathbf{r},t) = \int \frac{\rho(\mathbf{r}', t-\frac{|\mathbf{r}-\mathbf{r}'|}{c})}{4\pi\epsilon_o|\mathbf{r}-\mathbf{r}'|}d^3\mathbf{r}',$$

where we made an explicit use of the time-shift property of the Fourier transform as in

$$\rho(\mathbf{r}', t-\frac{|\mathbf{r}-\mathbf{r}'|}{c}) \leftrightarrow R(\mathbf{r}',\omega)e^{-j\omega|\mathbf{r}-\mathbf{r}'|/c} \equiv \tilde{\rho}(\mathbf{r}')e^{-jk|\mathbf{r}-\mathbf{r}'|}.$$

  - Note that $V(\mathbf{r},t)$ is a weighted superposition of the *past* values of charge density $\rho(\mathbf{r},t)$ (as opposed to future values) because of our

use of the causal solution[1] (as opposed to acausal solution) of the forced wave equation discussed above.

It is useful to stress at this point the relationship between a phasor (of a time harmonic function) and a Fourier transform (of a time domain function) as follows:

- A phasor, say, $\tilde{V}(\mathbf{r})$ is a sample of a Fourier transform function $V(\mathbf{r}, \omega)$ at the frequency ω of a time-harmonic function that the phasor represents.

- Conversely, a Fourier transform $V(\mathbf{r}, \omega)$ represents a continuous collection of phasors $\tilde{V}(\mathbf{r})$ representing time-harmonic functions of all possible ω.

  Based on the above correspondence principle we feel free to switch between phasor and Fourier transform concepts as convenient.

Question: is *causality* an additional postulate on top of Maxwell's equations that needs to be invoked to understand radiation?

Answer: no, not really, we need to invoke causality at this stage to pick the relevant root of the solution for the forced wave equation simply because we took a shortcut of using a steady-state solution based on Fourier transforms (phasors). Had we solved the same problem as an initial value problem (using the Laplace transform), only the retarded potential solution would have figured in our answer naturally without having to invoke a separate causality postulate — see *J. L. Anderson*, "Why we use retarded potentials", *Am.J. Phys.*, 60, 465, 1992.

[1] This choice is also referred to as Sommerfeld's *radiation condition* after Arnold Sommerfeld who also developed an asymptotic formula that retains the causal solution and rejects the acausal one.

Having finished the derivation of the retarded potential solution of the forced wave equation for scalar potential, we can re-state our result, and by *analogy* the result for the retarded vector potential as:

$$V(\mathbf{r},t) = \int \frac{\rho(\mathbf{r}', t - \frac{|\mathbf{r}-\mathbf{r}'|}{c})}{4\pi\epsilon_o|\mathbf{r}-\mathbf{r}'|} d^3\mathbf{r}',$$

the solution of inhomogeneous wave equation

$$\nabla^2 V - \mu_o\epsilon_o \frac{\partial^2 V}{\partial t^2} = -\frac{\rho}{\epsilon_o}$$

$$\mathbf{A}(\mathbf{r},t) = \int \frac{\mu_o \mathbf{J}(\mathbf{r}', t - \frac{|\mathbf{r}-\mathbf{r}'|}{c})}{4\pi|\mathbf{r}-\mathbf{r}'|} d^3\mathbf{r}',$$

the solution of inhomogeneous wave equation

$$\nabla^2 \mathbf{A} - \mu_o\epsilon_o \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu_o \mathbf{J}$$

where

$$c \equiv \frac{1}{\sqrt{\mu_o\epsilon_o}} \text{ is the speed of light in free space.}$$

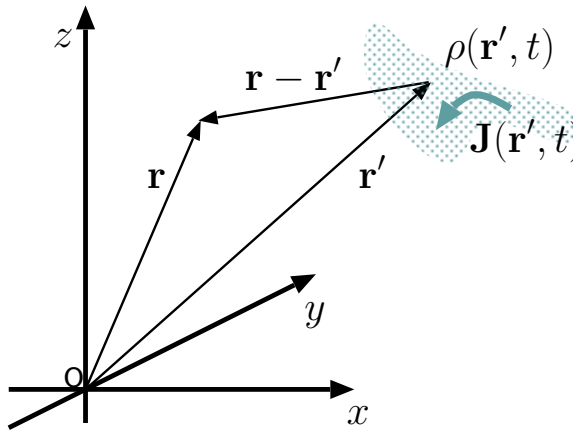
These results indicate that *retarded* potentials

$$V(\mathbf{r},t) \quad \text{and} \quad \mathbf{A}(\mathbf{r},t)$$

are appropriately weighted and *delayed* sums of

$$\rho(\mathbf{r},t) \quad \text{and} \quad \mathbf{J}(\mathbf{r},t)$$

in convolution-like 3D space integrals.



Next turning our attention to retarded **vector potential** solutions, we note that the results stated in time and frequency domains are as follows:

Time-domain:

$$\mathbf{A}(\mathbf{r},t)=\int\frac{\mu_o\mathbf{J}(\mathbf{r}',t-\frac{|\mathbf{r}-\mathbf{r}'|}{c})}{4\pi|\mathbf{r}-\mathbf{r}'|}d^3\mathbf{r}',$$

the solution of the inhomogeneous wave equation

$$\nabla^2\mathbf{A}-\frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2}=-\mu_o\mathbf{J}.$$

Frequency-domain:

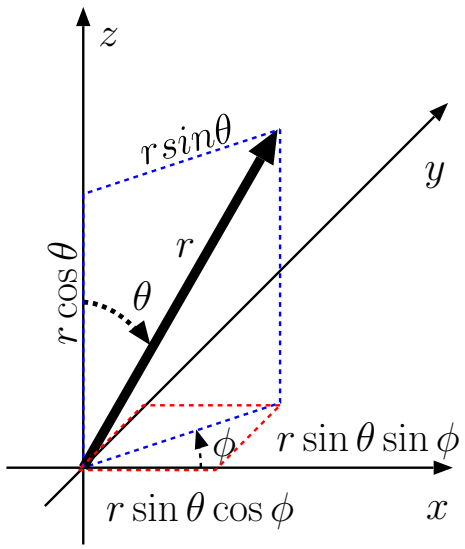
$$\tilde{\mathbf{A}}(\mathbf{r})=\int\frac{\mu_o\tilde{\mathbf{J}}(\mathbf{r}')e^{-jk|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}d^3\mathbf{r}',$$

the solution of the inhomogeneous wave equation

$$\nabla^2\tilde{\mathbf{A}}+\frac{\omega^2}{c^2}\tilde{\mathbf{A}}=-\mu_o\tilde{\mathbf{J}}.$$

In the next lecture we will learn how to perform vector calculus operations in spherical coordinates and then apply the frequency-domain result obtained above to the calculation of radiation from short current elements.

5 Vector calculus in spherical coordinates



In studies of radiation from compact antennas it is more convenient to use **spherical coordinates** instead of the Cartesian coordinates that we are familiar with. In this lecture we will learn

1. how to represent vectors and vector fields in spherical coordinates,
2. how to perform div, grad, curl, and Laplacian operations in spherical coordinates.

- A 3D position vector

$$\mathbf{r} = (x, y, z)$$

with *Cartesian coordinates* (x, y, z) is said to have *spherical coordinates* (r, θ, ϕ) where

$$\begin{aligned} \text{length } r &\equiv |\mathbf{r}| = \sqrt{x^2 + y^2 + z^2} \\ \text{zenith angle } \theta &= \tan^{-1} \frac{\sqrt{x^2 + y^2}}{z} \\ \text{azimuth angle } \phi &= \tan^{-1} \frac{y}{x} = \text{atan2}(y, x) \end{aligned}$$

In terms of spherical coordinates, Cartesian coordinates can be expressed as

$$\begin{aligned} x &= r \sin\theta \cos\phi \\ y &= r \sin\theta \sin\phi \\ z &= r \cos\theta. \end{aligned}$$

Ratios $x/r = \sin\theta\cos\phi$, $y/r = \sin\theta\sin\phi$, and $z/r = \cos\theta$ are referred to as **direction cosines** $\cos\theta_x$, $\cos\theta_y$, and $\cos\theta_z$, respectively, as they represent the *cosine* of the angle between vector $\mathbf{r} = (x, y, z)$ and the x-, y-, and z-**directions**, respectively, namely θ_x, θ_y, $\theta_z = \theta$.

- In Cartesian coordinates we have mutually orthogonal unit vectors

$$\hat{x},\ \hat{y},\ \hat{z}$$

pointing in the direction of increasing Cartesian coordinates x, y, z, respectively.

- Likewise, in spherical coordinates we have mutually orthogonal unit vectors

$$\hat{r},\ \hat{\theta},\ \hat{\phi}$$

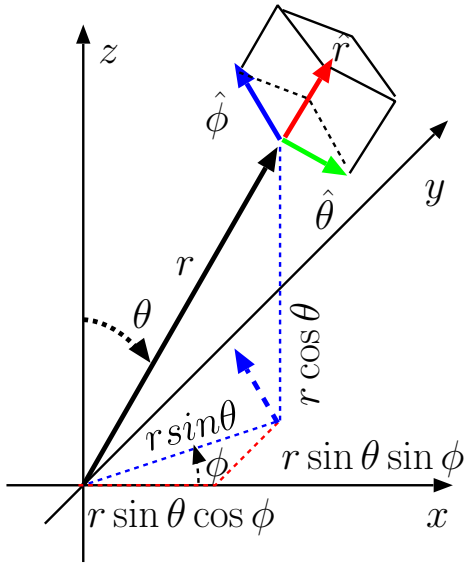
pointing in the direction of increasing coordinates r, θ, ϕ, respectively.

- However, unlike $\hat{x}$, $\hat{y}$, $\hat{z}$, the unit vectors $\hat{r}$, $\hat{\theta}$, $\hat{\phi}$ are **not global** — rather they are **local** in the sense that their directions depend on the local coordinates.

  – The local nature of $\hat{r}$, $\hat{\theta}$, $\hat{\phi}$ becomes clear when they are expressed in terms of the global unit vectors $\hat{x}$, $\hat{y}$, $\hat{z}$ as follows:

$$\begin{aligned}
\hat{r} &= \frac{\mathbf{r}}{r} = \frac{(x,y,z)}{r} = \hat{x}\sin\theta\cos\phi + \hat{y}\sin\theta\sin\phi + \hat{z}\cos\theta \\
\hat{\phi} &= \frac{(-y,x,0)}{\sqrt{x^2+y^2}} = -\hat{x}\sin\phi + \hat{y}\cos\phi \\
\hat{\theta} &= \hat{\phi}\times\hat{r} = \hat{x}\cos\theta\cos\phi + \hat{y}\cos\theta\sin\phi - \hat{z}\sin\theta
\end{aligned}$$

  Make sure you understand each of the terms above with reference to the figure shown in the margin.



Unit-vectors **$\hat{r}$, $\hat{\theta}$,and $\hat{\phi}$ shown in red, green, and blue point in mutually orthogonal directions of increasing spherical coordinates r, θ, and ϕ, respectively, such that $\hat{\theta}\times\hat{\phi} = \hat{r}$. Note that $\hat{r}$, $\hat{\theta}$,and $\hat{\phi}$ are local unit vectors (i.e., coordinate dependent) unlike the global unit vectors $\hat{x}$, $\hat{y}$, and $\hat{z}$ of the Cartesian coordinate system.**

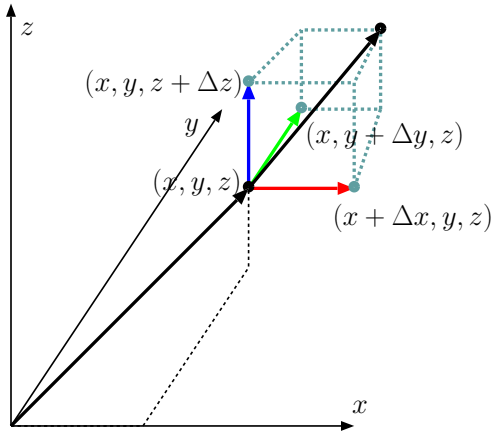
- In Cartesian coordinates we have an infinitesimal volume element

$$dV = dxdydz$$

which is used in 3D volume integrals and often denoted as "$d^3\mathbf{r}$".

  - Note that dV is the volume of a rectangular box formed by the intersection of **constant coordinate surfaces** of two infinitesimally close points having a separation vector

$$d\mathbf{r} = \hat{x}dx + \hat{y}dy + \hat{z}dz.$$



Any vector

$$\mathbf{A}(\mathbf{r}) = A_x\hat{x} + A_y\hat{y} + A_z\hat{z},$$

where A_x, A_y, and A_z are the projections of $\mathbf{A}(\mathbf{r})$ on red, green, and blue arrows aligned with $\hat{x}$,$\hat{y}$, $\hat{z}$, respectively.

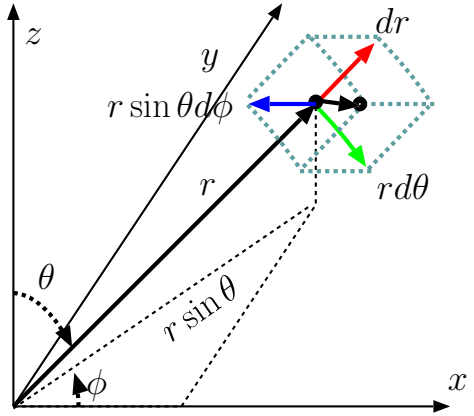
- Infinitesimal volume element $d^3\mathbf{r}$ expressed in terms of spherical coordinates and their increments is

$$dV = (dr)\,(rd\theta)\,(r\sin\theta d\phi) = dr\,r^2\underbrace{\sin\theta\,d\theta\,d\phi}_{\text{solid angle increment } = d\Omega}.$$

  - Once again dV is the volume of a rectangular box formed by the intersection of **constant coordinate surfaces** of two infinitesimally close points having a separation vector

$$d\mathbf{r} = \hat{r}dr + \hat{\theta}rd\theta + \hat{\phi}r\sin\theta d\phi.$$

  - Note that in this case constant coordinate surfaces are no longer planar globally, but over infinitesimal dimensions of dV the surfaces will appear locally planar.



Any vector

$$\mathbf{A}(\mathbf{r}) = A_r\hat{r} + A_\theta\hat{\theta} + A_\phi\hat{\phi},$$

where A_r, A_θ, and A_ϕ are the projections of $\mathbf{A}(\mathbf{r})$ on red, green, and blue arrows aligned with $\hat{r}$,$\hat{\theta}$, $\hat{\phi}$, respectively.

In Cartesian coordinates div, curl, and grad

$$\nabla \cdot \mathbf{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}$$

$$\nabla \times \mathbf{A} = \begin{vmatrix} \hat{x} & \hat{y} & \hat{z} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ A_x & A_y & A_z \end{vmatrix}$$

$$\nabla V = \frac{\partial V}{\partial x}\hat{x} + \frac{\partial V}{\partial y}\hat{y} + \frac{\partial V}{\partial z}\hat{z}$$

are obtained by applying the del operator

$$\nabla \equiv (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$$

"algebraically" to vectors

$$\mathbf{A} = A_x\hat{x} + A_y\hat{y} + A_z\hat{z}$$

and scalars

$$V(x, y, z)$$

as indicated above.

In spherical coordinates div, curl, and grad

$$\nabla \cdot \mathbf{A} = \frac{1}{r^2}\frac{\partial(r^2 A_r)}{\partial r} + \frac{1}{r\sin\theta}\frac{\partial(\sin\theta A_\theta)}{\partial\theta} + \frac{1}{r\sin\theta}\frac{\partial A_\phi}{\partial\phi}$$

$$\nabla \times \mathbf{A} = \begin{vmatrix} \frac{\hat{r}}{r^2\sin\theta} & \frac{\hat{\theta}}{r\sin\theta} & \frac{\hat{\phi}}{r} \\ \frac{\partial}{\partial r} & \frac{\partial}{\partial\theta} & \frac{\partial}{\partial\phi} \\ A_r & rA_\theta & r\sin\theta A_\phi \end{vmatrix}$$

$$\nabla V = \frac{\partial V}{\partial r}\hat{r} + \frac{1}{r}\frac{\partial V}{\partial\theta}\hat{\theta} + \frac{1}{r\sin\theta}\frac{\partial V}{\partial\phi}\hat{\phi}$$

are obtained for vectors

$$\mathbf{A} = A_r\hat{r} + A_\theta\hat{\theta} + A_\phi\hat{\phi}$$

and scalars

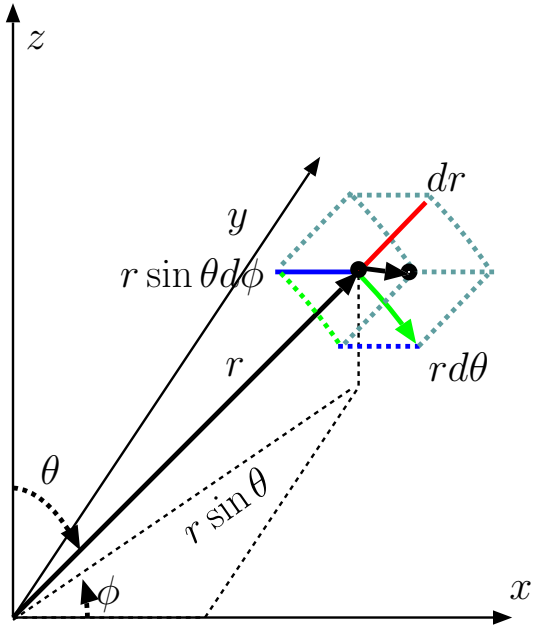
$$V(r, \theta, \phi)$$

as indicated above. **Note that there is no del operator that "works algebraically" in spherical coordinates**.

Example 1: Verify the $\hat{r}$ component of $\nabla \times \mathbf{A}$ formula in spherical coordinates by showing that it corresponds to

$$\lim_{A_C \to 0} \frac{\oint_C \mathbf{A} \cdot d\mathbf{l}}{A_C}$$

where A_C is the enclosed area of contour C orthogonal to $\hat{r}$ marked in the margin by blue and green edges.



Solution: In spherical coordinates

$$\nabla \times \mathbf{A} = \begin{vmatrix} \frac{\hat{r}}{r^2 \sin\theta} & \frac{\hat{\theta}}{r\sin\theta} & \frac{\hat{\phi}}{r} \\ \frac{\partial}{\partial r} & \frac{\partial}{\partial \theta} & \frac{\partial}{\partial \phi} \\ A_r & rA_\theta & r\sin\theta A_\phi \end{vmatrix}$$

and, therefore, $\hat{r}$ component of $\nabla \times \mathbf{A}$ is

$$(\nabla \times \mathbf{A}) \cdot \hat{r} \;=\; \frac{1}{r^2 \sin\theta}\left(\frac{\partial}{\partial\theta} r\sin\theta A_\phi - \frac{\partial}{\partial\phi} rA_\theta\right) = \frac{1}{r\sin\theta}\left(\frac{\partial}{\partial\theta}\sin\theta A_\phi - \frac{\partial}{\partial\phi} A_\theta\right).$$

To show that this expression corresponds (as it should by definition) to

$$\lim_{A_C \to 0} \frac{\oint_C \mathbf{A} \cdot d\mathbf{l}}{A_C}$$

where circulation path C and enclosed area A_C are as described in the question statement, we first note that

$$A_C \approx (r\sin\theta d\phi)(rd\theta)$$

to second order in increments $d\theta$ and $d\phi$. Also,

$$\begin{aligned}\oint_C \mathbf{A}\cdot d\mathbf{l} &= A_\theta(r,\theta,\phi)rd\theta + A_\phi(r,\theta+d\theta,\phi)r\sin(\theta+d\theta)d\phi \\ &\quad -A_\theta(r,\theta,\phi+d\phi)rd\theta - A_\phi(r,\theta,\phi)r\sin\theta d\phi\end{aligned}$$

starting on the green edge. Thus

$$\begin{aligned}\frac{\oint_C \mathbf{A}\cdot d\mathbf{l}}{A_C} &= \frac{A_\theta(r,\theta,\phi)-A_\theta(r,\theta,\phi+d\phi)}{r\sin\theta d\phi} \\ &\quad +\frac{A_\phi(r,\theta+d\theta,\phi)\sin(\theta+d\theta)-A_\phi(r,\theta,\phi)\sin\theta}{r\sin\theta d\theta}\end{aligned}$$

which yields in the limit of vanishing $d\theta$ and $d\phi$

$$-\frac{1}{r\sin\theta}\frac{\partial}{\partial\phi}A_\theta + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}\sin\theta A_\phi \equiv (\nabla\times\mathbf{A})\cdot\hat{r}$$

as requested.

See Appendix A and B in *Rao* for a complete coverage of the derivation of div, grad, curl in spherical coordinates.

Example 2: Verify the gradient procedure

$$\nabla V = \frac{\partial V}{\partial r}\hat{r} + \frac{1}{r}\frac{\partial V}{\partial \theta}\hat{\theta} + \frac{1}{r\sin\theta}\frac{\partial V}{\partial \phi}\hat{\phi}$$

in spherical coordinates.

Solution: Independent of the coordinate employed, the *total differential* dV and the *gradient* ∇V of a scalar field $V(\mathbf{r})$ are related by

$$dV = \nabla V \cdot d\mathbf{r}.$$

In the Cartesian coordinate system where $V = V(x, y, z)$, this relation expands as

$$dV = \frac{\partial V}{\partial x}dx + \frac{\partial V}{\partial y}dy + \frac{\partial V}{\partial z}dz = \nabla V \cdot (\hat{x}dx + \hat{y}dy + \hat{z}dz)$$

and implies

$$\nabla V = \frac{\partial V}{\partial x}\hat{x} + \frac{\partial V}{\partial y}\hat{y} + \frac{\partial V}{\partial z}\hat{z}.$$

Likewise, for spherical coordinates where $V = V(r, \theta, \phi)$, we have

$$dV = \frac{\partial V}{\partial r}dr + \frac{\partial V}{\partial \theta}d\theta + \frac{\partial V}{\partial \phi}d\phi = \nabla V \cdot (\hat{r}dr + \hat{\theta}rd\theta + \hat{\phi}r\sin\theta d\phi)$$

implying that

$$\nabla V = \frac{\partial V}{\partial r}\hat{r} + \frac{1}{r}\frac{\partial V}{\partial \theta}\hat{\theta} + \frac{1}{r\sin\theta}\frac{\partial V}{\partial \phi}\hat{\phi}.$$

Example 3: Show that the Laplacian of a scalar field $V(r,\theta,\phi)$ is specified as

$$\nabla^2 V = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2\frac{\partial V}{\partial r}) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta\frac{\partial V}{\partial\theta}) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 V}{\partial\phi^2}.$$

Solution: Since the Laplacian is the divergence of a gradient, we start by noting that

$$\nabla V = \frac{\partial V}{\partial r}\hat{r} + \frac{1}{r}\frac{\partial V}{\partial\theta}\hat{\theta} + \frac{1}{r\sin\theta}\frac{\partial V}{\partial\phi}\hat{\phi}.$$

Applying to this vector the divergence formula

$$\begin{aligned}\nabla\cdot\nabla V &= \frac{1}{r^2}\frac{\partial(r^2(\nabla V)_r)}{\partial r} + \frac{1}{r\sin\theta}\frac{\partial(\sin\theta(\nabla V)_\theta)}{\partial\theta} + \frac{1}{r\sin\theta}\frac{\partial(\nabla V)_\phi}{\partial\phi}\\ &= \frac{1}{r^2}\frac{\partial(r^2\frac{\partial V}{\partial r})}{\partial r} + \frac{1}{r\sin\theta}\frac{\partial(\sin\theta\frac{1}{r}\frac{\partial V}{\partial\theta})}{\partial\theta} + \frac{1}{r\sin\theta}\frac{\partial(\frac{1}{r\sin\theta}\frac{\partial V}{\partial\phi})}{\partial\phi}\end{aligned}$$

the above result for the Laplacian is readily obtained.

6 Spherical waves

In this lecture we will find out that short-filaments of oscillatory currents produce *uniform* **spherical waves** of the vector potential propagating away from the filament. The relationship between spherical waves of the vector potential and the corresponding electromagnetic wave fields will be examined in the next lecture.

We recall that time-varying solutions of Maxwell's equations can be obtained via

$$\mathbf{B} = \nabla \times \mathbf{A},$$

where the vector potential $\mathbf{A}(\mathbf{r}, t)$ is related to time-varying current density $\mathbf{J}(\mathbf{r}, t)$ via

Time-domain:

$$\mathbf{A}(\mathbf{r}, t) = \int \frac{\mu_o \mathbf{J}(\mathbf{r}', t - \frac{|\mathbf{r}-\mathbf{r}'|}{c})}{4\pi|\mathbf{r}-\mathbf{r}'|} d^3\mathbf{r}'.$$

Frequency-domain:

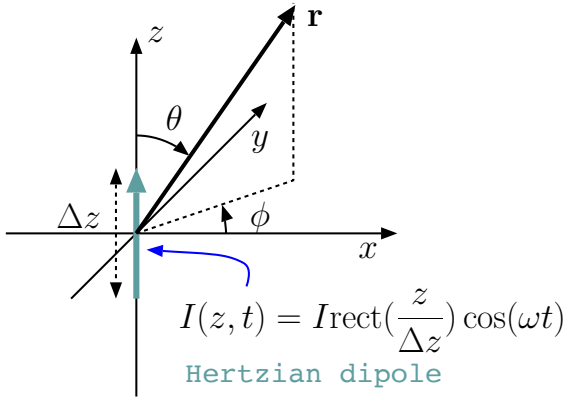
$$\tilde{\mathbf{A}}(\mathbf{r}) = \int \frac{\mu_o \tilde{\mathbf{J}}(\mathbf{r}') e^{-jk|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} d^3\mathbf{r}',$$

where

$$k = \omega\sqrt{\mu_o \epsilon_o}.$$

- We will next examine the implications of the above results from Lecture 4 for an $\hat{z}$ directed **infinitesimal current filament** defined as

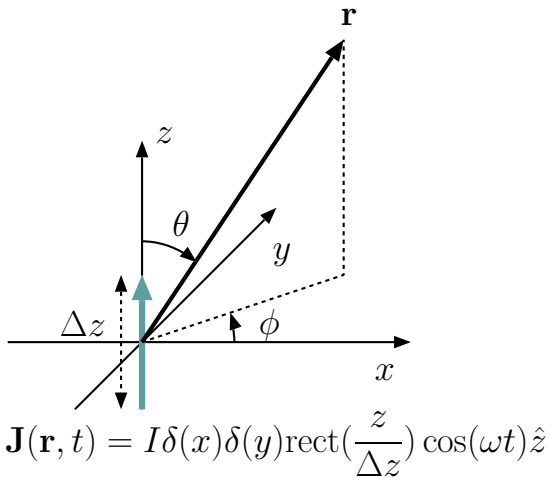
$$I(\mathbf{r}, t) = \begin{cases} I\cos(\omega t), & \text{for } x = 0,\ y = 0,\ -\frac{\Delta z}{2} < z < \frac{\Delta z}{2} \\ 0, & \text{otherwise.} \end{cases}$$



where constant I is specified in units of amperes (A). We can associate with this infinitesimal current the following current density function

$$\begin{aligned}\mathbf{J}(\mathbf{r},t) &= \begin{cases} I\delta(x)\delta(y)\cos(\omega t)\hat{z}, & \text{for } -\frac{\Delta z}{2} < z < \frac{\Delta z}{2} \\ 0, & \text{otherwise.} \end{cases} \\ &= I\delta(x)\delta(y)\text{rect}(\frac{z}{\Delta z})\cos(\omega t)\hat{z}\,\frac{\text{A}}{\text{m}^2}\end{aligned}$$

recalling that the dimension of an impulse $\delta(x)$ is m^{-1}.



- The oscillatory and $\hat{z}$ directed infinitesimal current filament of a length Δz can in turn can be represented in terms of a *phasor*

$$\tilde{\mathbf{J}}(\mathbf{r}) = I\delta(x)\delta(y)\text{rect}(\frac{z}{\Delta z})\hat{z}\,\frac{\text{A}}{\text{m}^2}.$$

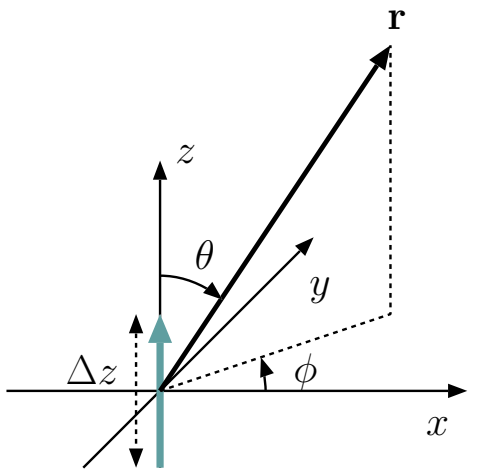
We can also re-write this as

$$\tilde{\mathbf{J}}(\mathbf{r}) = I\delta(x)\delta(y)\Delta z\,\frac{\text{rect}(\frac{z}{\Delta z})}{\Delta z}\,\hat{z}\,\frac{\text{A}}{\text{m}^2}$$

in which the ratio with the rectangle in the numerator can be treated as the *impulse* "$\delta(z)$" provided that the width, Δz, of the rectangle is considered an *infinitesimal* so that the ratio

$$\frac{\text{rect}(\frac{z}{\Delta z})}{\Delta z}$$

represents *in effect* an infinitely thin and tall function centered about $z = 0$ having a unity area underneath[1].



$$\tilde{\mathbf{J}}(\mathbf{r}) = I\Delta z\delta(x)\delta(y)\delta(z)\hat{z}$$

- Next we substitute this current density phasor $\tilde{\mathbf{J}}(\mathbf{r})$ (with Δz considered an infinitesimal) into the phasor formula for the retarded vector potential to obtain

$$\tilde{\mathbf{A}}(\mathbf{r}) = \int \frac{\mu_o \tilde{\mathbf{J}}(\mathbf{r}')e^{-jk|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}d^3\mathbf{r}'$$

$$= \int\int\int \frac{\mu_o \overbrace{I\delta(x')\delta(y')\Delta z\delta(z')\hat{z}}^{\tilde{\mathbf{J}}(\mathbf{r}')} e^{-jk|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}dx'dy'dz'$$

  where the integrations are to be carried over x', y', and z' in the range $-\infty$ to $+\infty$.

- These are very easy integrals to take because of $\delta(x')$, $\delta(y')$, and $\delta(z')$ factors in the integrand, and lead to (after replacing all x', y', and z' elsewhere in the integrand by 0)

$$\tilde{\mathbf{A}}(\mathbf{r}) = \frac{\mu_o}{4\pi} I\Delta z \frac{e^{-jkr}}{r}\hat{z},$$

[1]This treatment of Δz as an infinitesimal is permissible in $\tilde{\mathbf{A}}(\mathbf{r})$ calculation provided that $\Delta z \ll \lambda/2\pi = k^{-1}$ and $r = |\mathbf{r}|$. This can be seen easily by noting that $e^{-jk|\mathbf{r}-\mathbf{r}'|} \approx e^{-jkr}e^{\mp jk\Delta z\cos\theta/2}$ for $\mathbf{r}' = \pm\frac{\Delta z}{2}\hat{z}$ and $\Delta z/2 \ll r$, which is in turn $\approx e^{-jkr}$ provided that $k\Delta z \ll 1$. This amounts to having Δz small compared to all remaining lengths, ramely λ and r, in the problem!

where $r = |\mathbf{r}|$ as usual. Converting this result into time domain by multiplying it with $e^{j\omega t}$ and taking the real part of the product we obtain

$$\mathbf{A}(\mathbf{r},t) = \frac{\mu_o}{4\pi} I\Delta z \frac{\cos(\omega t - kr)}{r}\hat{z}.$$

We have just finished deriving the retarded vector potential solution of an oscillatory infinitesimal current filament known as the **Hertzian dipole**.

Our results indicate that for a Hertzian dipole oriented in $\hat{z}$ direction, the vector potential solution

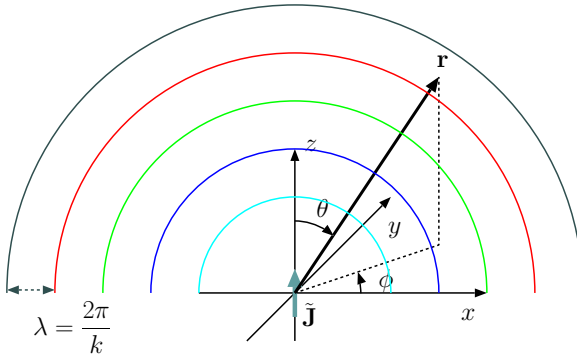
Frequency-domain:

$$\tilde{\mathbf{A}}(\mathbf{r}) = \frac{\mu_o}{4\pi} I\Delta z \frac{e^{-jkr}}{r}\hat{z}$$

Time-domain:

$$\mathbf{A}(\mathbf{r},t) = \frac{\mu_o}{4\pi} I\Delta z \frac{\cos(\omega t - kr)}{r}\hat{z}$$

is also oriented in the $\hat{z}$ direction and oscillate in time at the frequency ω of the oscillating dipole. Note that:



1. These vector potential solutions describe a **spherical wave** (as opposed to a plane wave) characterized by **spherical surfaces of constant phase** associated with

   $$e^{-jkr} \quad \text{and} \quad \cos(\omega t - kr)$$

   variations in frequency and time domains.

2. Spherical wave solution is *uniform* in the sense that the vector potential phasor $\tilde{\mathbf{A}}$ is constant (in direction and magnitude) on spherical surfaces

of constant phase (in analogy to uniform TEM plane waves of electric and magnetic fields studies in ECE 329).

3. Clearly, the propagation speed of the spherical wave is

$$v_p = \frac{\omega}{k} = \frac{\omega}{\omega\sqrt{\mu_o\epsilon_o}} = c.$$

4. The spherical wave is also characterized by an oscillation amplitude that varies as $\frac{1}{r}$ away from the radiating source

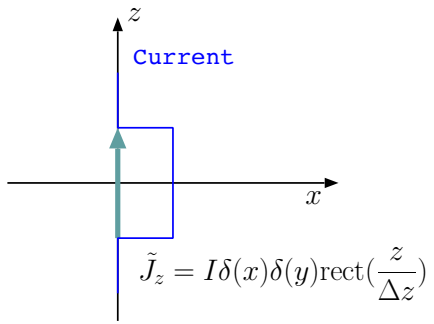
- In the next lecture we will take the curl of this result (using spherical coordinates operators) to obtain spherical (but non-uniform) waves of **B** that accompany the **A**-waves, and then derive the accompanying spherical (but non-uniform) **E**-waves using Ampere's law.

  - We will find out **E**- and **B**-waves derived from **A**-waves are in general non-uniform and form "beams" of directions along which field magnitudes $|\tilde{\mathbf{E}}|$ and $|\tilde{\mathbf{B}}|$ maximize over spherical planes of constant phase.
  - The mathematical description of these *beams* is provided by the "gain function" and the "solid angle" of the radiating system to be defined and explored in Lecture 10.

- In deriving **E**- and **B**-waves from **A** we will not explicitly worry about $V(\mathbf{r}, t)$ and $\rho(\mathbf{r}, t)$ that accompanies the Hertzian dipole behavior (since **J** contains all information included in ρ variations).

- For completeness sake, however, let us examine what kind of $\rho(\mathbf{r}, t)$ variation should be expected for the Hertzian dipole.

The Hertzian dipole is a *hypothetical* radiation element defined and introduced above. Its main utility is that it has the simplest radiation properties that one could imagine and use as a building block to represent more complicated (and practical rather than hypothetical) radiation elements.



- A Hertzian dipole was defined as a filament of an infinitesimal length Δz which is carrying a constant (z-independent) current at each instant of time t.

  – Since outside the filament the current vanishes, charge conservation and the continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0$$

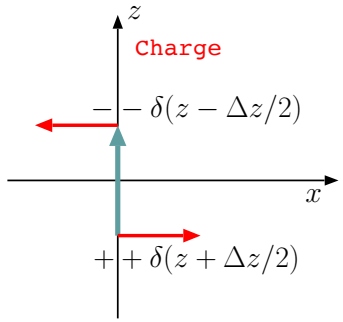
  demand that there has to be a time-varying charge accumulation at the two ends of the filament.

  Since for a $\hat{z}$ directed Hertzian dipole, $\mathbf{J} = \hat{z} J_z$, we can write the phasor domain form of the continuity equation as

$$j\omega\tilde{\rho} + \frac{\partial \tilde{J}_z}{\partial z} = 0.$$

  Thus, with

$$\tilde{J}_z = I\,\delta(x)\delta(y)\,\mathrm{rect}(\frac{z}{\Delta z})$$



$$\tilde{\rho} = j\frac{I}{\omega}\delta(x)\delta(y)$$
$$[\delta(z + \Delta z/2) - \delta(z - \Delta z/2)]$$

Depicted charge density (red)leads the depicted current density (blue) profile by a quarter period because of j term in charge density.

Positive reservoir of charge at z<0 end of the dipole discharges into the negative reservoir at the other end causing half a cycle of z-directed current across the filament.

By the end of half-cycle the top end is positively charged and the bottom end negatively, so a new half-cycle with motions in the opposite direction starts.

and

$$\frac{\partial \tilde{J}_z}{\partial z} = I\,\delta(x)\delta(y)[\delta(z+\frac{\Delta z}{2}) - \delta(z-\frac{\Delta z}{2})],$$

we get

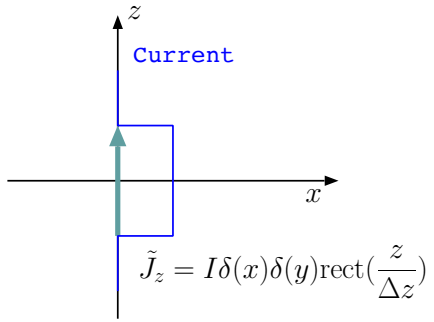
$$\tilde{\rho} = -\frac{1}{j\omega}\frac{\partial \tilde{J}_z}{\partial z} = j\frac{I}{\omega}\delta(x)\delta(y)[\delta(z+\frac{\Delta z}{2}) - \delta(z-\frac{\Delta z}{2})].$$

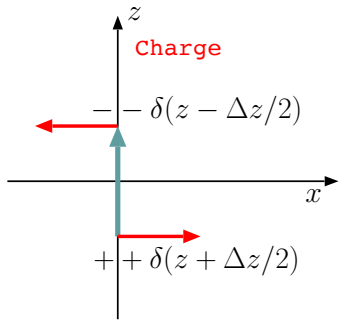
In time-domain this corresponds to

$$\rho(\mathbf{r},t) = \frac{I}{\omega}\delta(x)\delta(y)[\delta(z-\frac{\Delta z}{2}) - \delta(z+\frac{\Delta z}{2})]\sin(\omega t)\ \frac{\mathrm{C}}{\mathrm{m}^3}$$

accompanying the current density variation

$$\mathbf{J}(\mathbf{r},t) \ = \ I\delta(x)\delta(y)\mathrm{rect}(\frac{z}{\Delta z})\cos(\omega t)\hat{z}\,\frac{\mathrm{A}}{\mathrm{m}^2}.$$





$\tilde{\rho} = j\frac{I}{\omega}\delta(x)\delta(y)$
$[\delta(z+\Delta z/2) - \delta(z-\Delta z/2)]$

Depicted charge density (red)leads the depicted current density (blue) profile by a quarter period because of j term in charge density.

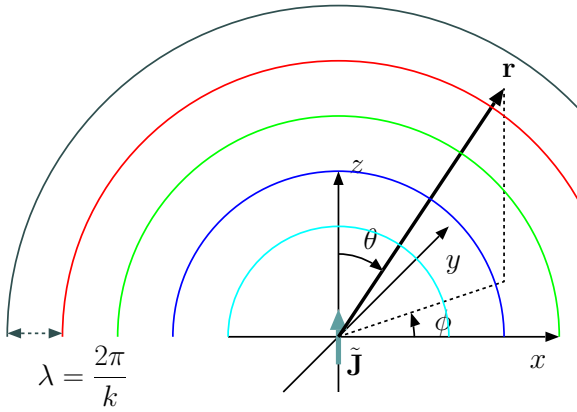
Positive reservoir of charge at z<0 end of the dipole discharges into the negative reservoir at the other end causing half a cycle of z-directed current across the filament.

By the end of half-cycle the top end is positively charged and the bottom end negatively, so a new half-cycle with motions in the opposite direction starts.

- Clearly, the result above shows that the "ends" of a Hertzian dipole element located at $z = \pm\frac{\Delta z}{2}$ serve as point-charge reservoirs (of opposite polarities) sustaining the current variations of the element.

  - Radiated fields of the Hertzian dipole should be attributed to both the time-varying ρ and the time-varying $\mathbf{J}$ even though considerations of $\mathbf{J}$ will be sufficient to determine the radiated fields owing to the dependence of ρ on $\mathbf{J}$ that is built-in within Maxwell's equations.

7 Hertzian dipole fields



- We concluded the last lecture with the retarded potential solutions

Frequency-domain:

$$\tilde{\mathbf{A}}(\mathbf{r}) = \frac{\mu_o}{4\pi} I \Delta z \frac{e^{-jkr}}{r} \hat{z}$$

Time-domain:

$$\mathbf{A}(\mathbf{r}, t) = \frac{\mu_o}{4\pi} I \Delta z \frac{\cos(\omega t - kr)}{r} \hat{z}$$

of a $\hat{z}$ directed Hertzian dipole.

- We noted that these oscillatory solutions describe spherical waves by virtue of the e^{-jkr} dependence of the potential phasor on r:

  - the variable r measures distance in *all* directions away from the origin, as opposed to, say, x measuring distance only along one coordinate axis labelled as x.

  Thus, while the phasor variation e^{-jkx} describes a plane wave, the phasor e^{-jkr} describes a spherical wave (see margin).

We will next determine the magnetic and electric fields produced by a Hertzian dipole.

- To calculate the magnetic field phasor $\tilde{\mathbf{B}}$ we will make use of

$$\tilde{\mathbf{B}} = \nabla \times \tilde{\mathbf{A}} \quad \text{and} \quad \nabla \times \tilde{\mathbf{A}} = \begin{vmatrix} \frac{\hat{r}}{r^2 \sin\theta} & \frac{\hat{\theta}}{r \sin\theta} & \frac{\hat{\phi}}{r} \\ \frac{\partial}{\partial r} & \frac{\partial}{\partial \theta} & \frac{\partial}{\partial \phi} \\ \tilde{A}_r & r\tilde{A}_\theta & r\sin\theta \tilde{A}_\phi \end{vmatrix} \text{ in spherical coordinates.}$$

- Given that

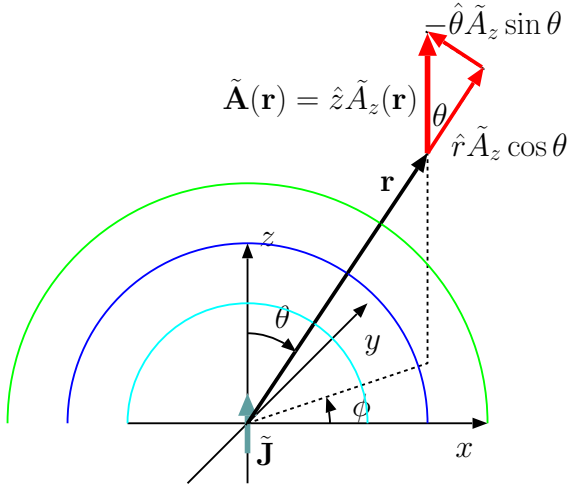
$$\tilde{\mathbf{A}}(\mathbf{r}) = \underbrace{\frac{\mu_o}{4\pi} I \Delta z \frac{e^{-jkr}}{r}}_{\tilde{A}_z(\mathbf{r})} \hat{z}$$

and

$$\hat{z} \cdot \hat{r} = \cos\theta, \quad \hat{z} \cdot \hat{\theta} = -\sin\theta, \quad \hat{z} \cdot \hat{\phi} = 0,$$

it follows that

$$\begin{aligned} \tilde{A}_r &= \tilde{\mathbf{A}}(\mathbf{r}) \cdot \hat{r} = \tilde{A}_z(\mathbf{r}) \cos\theta, \\ \tilde{A}_\theta &= \tilde{\mathbf{A}}(\mathbf{r}) \cdot \hat{\theta} = -\tilde{A}_z(\mathbf{r}) \sin\theta, \\ \tilde{A}_\phi &= \tilde{\mathbf{A}}(\mathbf{r}) \cdot \hat{\phi} = 0. \end{aligned}$$



$\tilde{A}_r(\mathbf{r}) = \tilde{A}_z(\mathbf{r}) \cos\theta$

$\tilde{A}_\theta(\mathbf{r}) = -\tilde{A}_z(\mathbf{r}) \sin\theta$

Substituting $\tilde{A}_r$, $\tilde{A}_\theta$, $\tilde{A}_\phi$ into the curl formula, we proceed as

$$\nabla \times \tilde{\mathbf{A}} = \frac{\mu_o}{4\pi} I \Delta z \begin{vmatrix} \frac{\hat{r}}{r^2 \sin\theta} & \frac{\hat{\theta}}{r \sin\theta} & \frac{\hat{\phi}}{r} \\ \frac{\partial}{\partial r} & \frac{\partial}{\partial \theta} & \frac{\partial}{\partial \phi} \\ \frac{e^{-jkr}}{r}\cos\theta & -r\frac{e^{-jkr}}{r}\sin\theta & 0 \end{vmatrix}.$$

Expanding the determinant, we obtain

$$\begin{aligned}\tilde{\mathbf{B}} = \nabla \times \tilde{\mathbf{A}} &= \frac{\mu_o}{4\pi} I\Delta z \frac{\hat{\phi}}{r}\{-\frac{\partial}{\partial r}e^{-jkr}\sin\theta - \frac{\partial}{\partial\theta}\frac{e^{-jkr}}{r}\cos\theta\} \\ &= \frac{\mu_o}{4\pi} I\Delta z(jk+\frac{1}{r})\sin\theta\frac{e^{-jkr}}{r}\hat{\phi}.\end{aligned}$$

Consequently,

$$\tilde{\mathbf{H}} = jkI\Delta z\sin\theta\frac{e^{-jkr}}{4\pi r}\hat{\phi}(1+\frac{1}{jkr}).$$

Notice, the wave field

$$\tilde{\mathbf{H}}(\mathbf{r}) = \hat{\phi}\tilde{H}_\phi(\mathbf{r})$$

of the Hertzian dipole is purely "azimuthal" — this is the direction the *right-hand-rule* would give if the right-hand-thumb were directed in the direction of dipole current.

- To obtain the accompanying electric field phasor we will next employ Ampere's law

$$\nabla \times \tilde{\mathbf{H}} = \tilde{\mathbf{J}} + j\omega\epsilon_o\tilde{\mathbf{E}},$$

with $\tilde{\mathbf{J}} = 0$, which is true at all locations outside the Hertzian dipole. In that case

$$\begin{aligned}\tilde{\mathbf{E}} &= \frac{\nabla\times\tilde{\mathbf{H}}}{j\omega\epsilon_o} = \\ &= \frac{1}{j\omega\epsilon_o}\begin{vmatrix} \frac{\hat{r}}{r^2\sin\theta} & \frac{\hat{\theta}}{r\sin\theta} & \frac{\hat{\phi}}{r} \\ \frac{\partial}{\partial r} & \frac{\partial}{\partial\theta} & \frac{\partial}{\partial\phi} \\ \tilde{H}_r & r\tilde{H}_\theta & r\sin\theta\tilde{H}_\phi \end{vmatrix} = \frac{1}{j\omega\epsilon_o}\begin{vmatrix} \frac{\hat{r}}{r^2\sin\theta} & \frac{\hat{\theta}}{r\sin\theta} & \frac{\hat{\phi}}{r} \\ \frac{\partial}{\partial r} & \frac{\partial}{\partial\theta} & \frac{\partial}{\partial\phi} \\ 0 & 0 & r\sin\theta\tilde{H}_\phi \end{vmatrix} \\ &= \frac{1}{j\omega\epsilon_o}\{\frac{\hat{r}}{r^2\sin\theta}\frac{\partial}{\partial\theta}r\sin\theta\tilde{H}_\phi - \frac{\hat{\theta}}{r\sin\theta}\frac{\partial}{\partial r}r\sin\theta\tilde{H}_\phi\}\end{aligned}$$

$$= \frac{1}{j\omega\epsilon_o}\{\frac{\hat{r}}{r\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\tilde{H}_\phi - \frac{\hat{\theta}}{r}\frac{\partial}{\partial r}r\tilde{H}_\phi\}.$$

Substituting

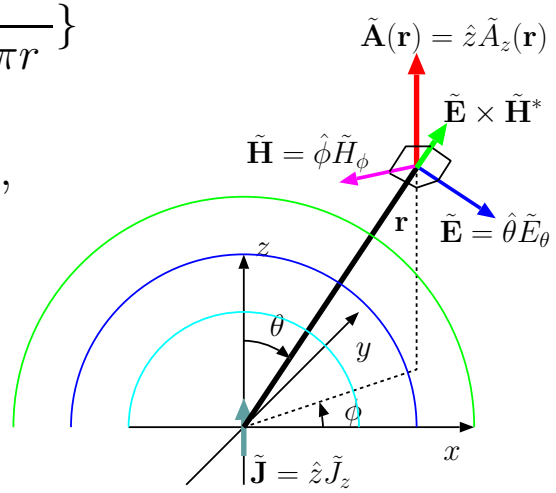
$$\tilde{H}_\phi = jkI\Delta z(1+\frac{1}{jkr})\sin\theta\frac{e^{-jkr}}{4\pi r}$$

from above, and simplifying, we have

$$\begin{aligned}
\tilde{\mathbf{E}} &= \frac{jkI\Delta z}{j\omega\epsilon_o}\{\frac{\hat{r}}{r\sin\theta}\frac{\partial}{\partial\theta}\sin\theta(1+\frac{1}{jkr})\sin\theta\frac{e^{-jkr}}{4\pi r} - \frac{\hat{\theta}}{r}\frac{\partial}{\partial r}r(1+\frac{1}{jkr})\sin\theta\frac{e^{-jkr}}{4\pi r}\} \\
&= \frac{kI\Delta z}{\omega\epsilon_o}\{\frac{\hat{r}}{r\sin\theta}(1+\frac{1}{jkr})\frac{e^{-jkr}}{4\pi r}\frac{\partial}{\partial\theta}\sin^2\theta - \frac{\hat{\theta}}{r}\sin\theta\frac{\partial}{\partial r}(1+\frac{1}{jkr})\frac{e^{-jkr}}{4\pi}\} \\
&= \sqrt{\frac{\mu_o}{\epsilon_o}}I\Delta z\{\frac{\hat{r}}{r}(1+\frac{1}{jkr})\frac{e^{-jkr}}{4\pi r}2\cos\theta + \hat{\theta}\sin\theta[jk(1+\frac{1}{jkr}) + \frac{1}{jkr^2}]\frac{e^{-jkr}}{4\pi r}\} \\
&= jk\eta_o I\Delta z\frac{e^{-jkr}}{4\pi r}\{\sin\theta\hat{\theta}[1+\frac{1}{jkr}+(\frac{1}{jkr})^2] + 2\cos\theta\hat{r}[\frac{1}{jkr}+(\frac{1}{jkr})^2]\},
\end{aligned}$$

where

$$\eta_o \equiv \sqrt{\frac{\mu_o}{\epsilon_o}}.$$



- This is a very complicated looking result.
  - Fortunately, many of the terms above are important only at very small values of r!

- If were to drop all of the terms in $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{H}}$ above except for those varying as $\frac{1}{r}$, we would be left with

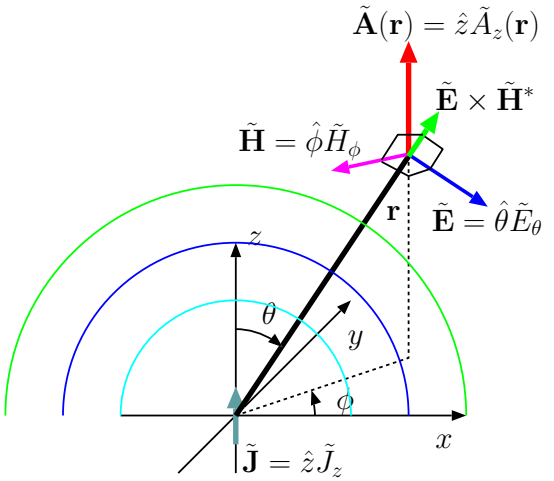
$$\tilde{\mathbf{E}} = j\eta_o Ik\Delta z \sin\theta \frac{e^{-jkr}}{4\pi r}\hat{\theta} \quad \text{and} \quad \tilde{\mathbf{H}} = jIk\Delta z \sin\theta \frac{e^{-jkr}}{4\pi r}\hat{\phi},$$

which are the only terms of the fields of the Hertzian dipole that matter at large distances (of interest for communication and remote sensing purposes).

  - They are called the **radiation fields** of the Hertzian dipole, and the remainder (the terms which have been dropped) are called the **storage fields**.
  - The reasoning behind this terminology is as follows:
    The average Poynting vector

$$\langle \mathbf{E} \times \mathbf{H} \rangle = \frac{1}{2}\mathrm{Re}\{\tilde{\mathbf{E}} \times \tilde{\mathbf{H}}^*\}$$

    computed with the full expressions for $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{H}}$ gives the same result as that computed with only the simplified *radiation* fields.
  - What that means is the remaining parts of $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{H}}$ (*storage* fields) do not contribute to the transport of energy away from the dipole.
  - They only represent a local energy exchange (and storage) between inductive and capacitive attributes of the dipole — recall that the



Radiation fields:

$$\tilde{\mathbf{E}} = j\eta_o Ik\Delta z \sin\theta \frac{e^{-jkr}}{4\pi r}\hat{\theta}$$

and

$$\tilde{\mathbf{H}} = jIk\Delta z \sin\theta \frac{e^{-jkr}}{4\pi r}\hat{\phi}.$$

dipole is both a filament having some inductance and a capacitor with two reservoirs for charge storage.

In many applications of radiation theory we only need to focus on the radiation fields.

Fortunately, the expressions for radiation fields are simple and have features resembling the plane TEM waves that we are already familiar with. Let's see what these features are:

1. The phasors are **orthogonal** and

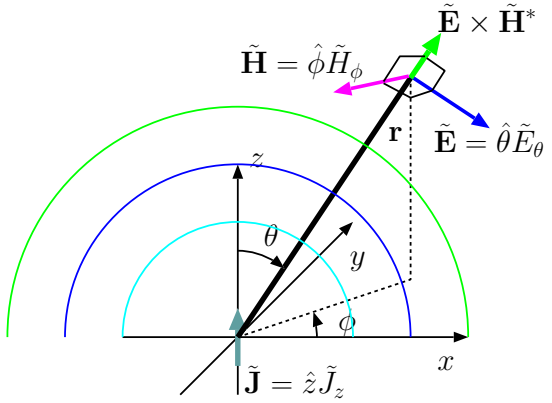
$$\tilde{\mathbf{E}} \times \tilde{\mathbf{H}}^* \propto \hat{\theta} \times \hat{\phi} = \hat{r}$$

   points in the radial direction $\hat{r}$ of the spherical wave propagation just as in plane TEM waves.

2. The magnitude of $\tilde{\mathbf{H}}$ can be obtained by dividing the magnitude of $\tilde{\mathbf{E}}$ by the intrinsic impedance η_o just as for plane TEM waves.

3. Conversely, the magnitude of $\tilde{\mathbf{E}}$ can be obtained by multiplying the magnitude of $\tilde{\mathbf{H}}$ by the intrinsic impedance η_o just as for plane TEM waves.

4. The direction of $\tilde{\mathbf{H}}$ can be deduced from the direction of $\tilde{\mathbf{E}}$ (and vice versa) by a 90° rotation and enforcing the right-hand-rule of having $\tilde{\mathbf{E}} \times \tilde{\mathbf{H}}^*$ point in $\hat{r}$ direction.



Radiation fields:

$$\tilde{\mathbf{E}} = j\eta_o I k \Delta z \sin\theta \frac{e^{-jkr}}{4\pi r}\hat{\theta}$$

and

$$\tilde{\mathbf{H}} = jIk\Delta z \sin\theta \frac{e^{-jkr}}{4\pi r}\hat{\phi}.$$

On the other hand, these spherical TEM waves radiated by the Hertzian dipole differ from *uniform* plane TEM waves by the facts that:

1. Field amplitude is not constant in the propagation direction because of $\frac{1}{r}$ dependence.

2. Field amplitude is not constant in the direction orthogonal to the propagation direction because of $\sin\theta$ dependence.

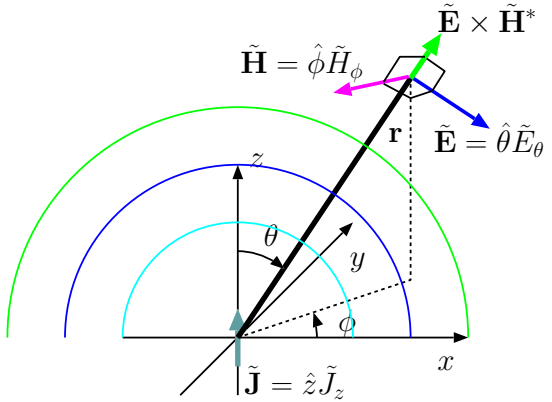
   As such, a Hertzian dipole radiates TEM waves which are *non-uniform* as well as spherical (non-planar).

   As such, Hertzian dipole radiation is said to be **anisotropic**!

- Radiation is strong — forms a "beam", so to speak — in the **broadside direction** of $\theta = 90°$ (with respect to the dipole axis),

- Radiation vanishes for $\theta = 0°, 180°$ along the **dipole axis**.

  – In short, radiation strength scales with $\Delta z \sin\theta$, a foreshortened version of length Δz of the dipole "seen" from an angle θ with respect to the dipole axis. More on this later on...



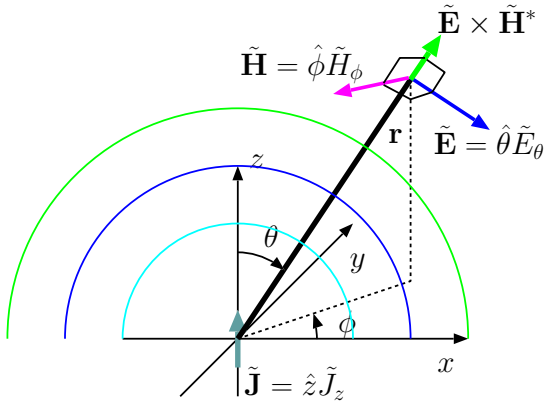
Radiation fields:

$$\tilde{\mathbf{E}} = j\eta_o I k \Delta z \sin\theta \frac{e^{-jkr}}{4\pi r}\hat{\theta}$$

and

$$\tilde{\mathbf{H}} = jIk\Delta z \sin\theta \frac{e^{-jkr}}{4\pi r}\hat{\phi}.$$

8 Radiation fields of dipole antennas



- Radiation fields of a $\hat{z}$-directed Hertzian dipole from the last lecture are repeated in the margin.

- In this lecture we will first obtain the radiation fields of **short dipole** antennas by superposing the Hertzian dipole fields.

- A "short dipole" is a practical antenna — as opposed to a hypothetical Hertzian dipole — consisting of a pair of thin straight conducting wires of equal lengths $\frac{L}{2}$ placed along a common axis leaving a short gap between them (see margin).

  - A short dipole is typically used by connecting a "source" across the gap that constitutes the "input port" of the dipole antenna.
  - Let's assume that the source is an independent current source

$$I(t) = I_o \cos \omega t \, \text{A} \quad \Leftrightarrow \quad \tilde{I} = I_o \angle 0 = I_o \, \text{A}$$

    and that the gap is an infinitesimal Δz so that the dipole and its input port occupy the region $-\frac{L}{2} < z < \frac{L}{2}$ in total.
  - We can then envision the entire dipole, including its input port, to be a stack of Hertzian dipoles of lengths Δz, with each Hertzian dipole centered about position z (in the interval $-\frac{L}{2} < z < \frac{L}{2}$) carrying a current $\tilde{I}(z)$, subject to boundary conditions

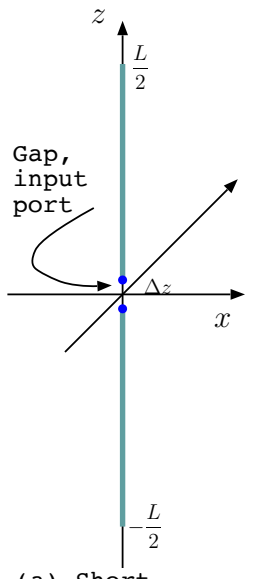
$$\tilde{I}(0) = I_o \angle 0 \, \text{A} \quad \text{and} \quad \tilde{I}(\pm\frac{L}{2}) = 0.$$

Radiation fields:

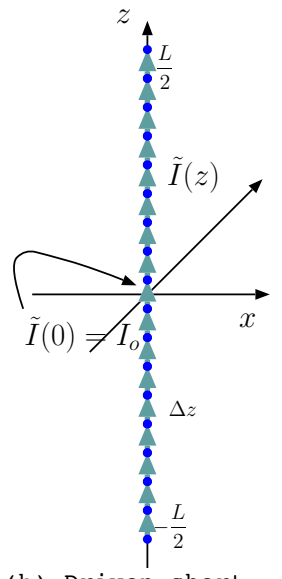
$$\tilde{\mathbf{E}} = j\eta_o I k \Delta z \sin\theta \frac{e^{-jkr}}{4\pi r}\hat{\theta}$$

and

$$\tilde{\mathbf{H}} = jIk\Delta z \sin\theta \frac{e^{-jkr}}{4\pi r}\hat{\phi}.$$



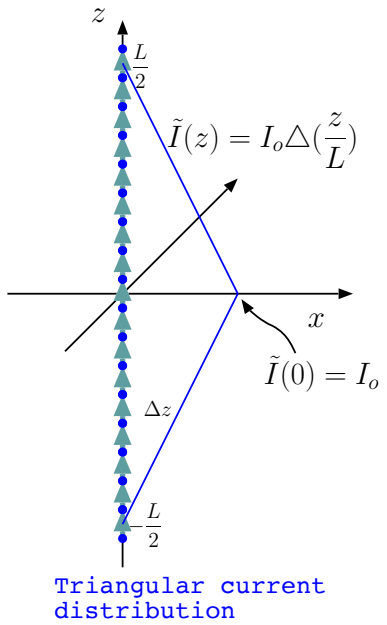
(a) Short dipole--- undriven



(b) Driven short dipole modeled as a stack of Hertzian dipoles

In conformity with these boundary conditions we will assume that $\tilde{I}(z)$ is a *triangular* current distribution

$$\tilde{I}(z) = I_o \triangle(\frac{z}{L})\,\text{A}.$$



- What are the *radiation fields* of the short dipole antenna described above?

  We can answer this question in several different ways:

  1. We could turn the specified $\tilde{I}(z)$ into a corresponding $\tilde{\mathbf{J}}(z)$, and then, in succession, calculate the retarded potential $\tilde{\mathbf{A}}$, the magnetic field $\tilde{\mathbf{B}} = \nabla \times \tilde{\mathbf{A}}$, and then obtain $\tilde{\mathbf{E}}$ from $\tilde{\mathbf{B}}$ using Ampere's law as we did for the Hertzian dipole. Finally, the storage fields decaying faster with distance than $\frac{1}{r}$ would be dropped from $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{B}}$ to obtain the radiation fields exclusively.
  2. A variant of (1), but with the radiation field $\tilde{\mathbf{H}}$ immediately deduced from $\tilde{\mathbf{B}}$, and then $\tilde{\mathbf{E}}$ is obtained by multiplying $\tilde{\mathbf{H}}$ by η_o and rotating it by 90° so that $\tilde{\mathbf{E}} \times \tilde{\mathbf{H}}^*$ points in $\hat{r}$ direction.
  3. Superpose shifted and scaled versions of the radiation fields of the Hertzian dipole (as we will do shortly).

  All these options enumerated above will work because Maxwell's equations and radiation process have linearity properties.

- The radiation electric field

$$\tilde{\mathbf{E}} = j\eta_o I k \Delta z \sin\theta \frac{e^{-jkr}}{4\pi r}\hat{\theta}$$

of a $\hat{z}$-directed Hertzian dipole

$$\tilde{\mathbf{J}} = \hat{z} I \Delta z \delta(x)\delta(y)\delta(z)$$

implies the following linear relationships for a $\hat{z}$-polarized radiation process, where the input function shown on the left represents the current distribution of the radiator:

$$\delta(z) \rightarrow \boxed{\hat{z}\text{-pol radiator}} \rightarrow j\eta_o k \sin\theta \frac{e^{-jk|\mathbf{r}|}}{4\pi|\mathbf{r}|}\hat{\theta}$$

and

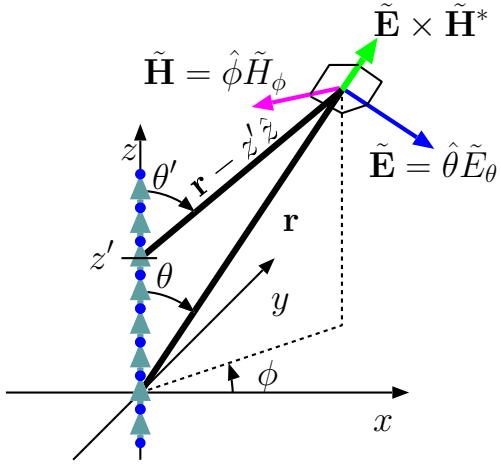
$$\delta(z-z') \rightarrow \boxed{\hat{z}\text{-pol radiator}} \rightarrow j\eta_o k \sin\theta' \frac{e^{-jk|\mathbf{r}-z'\hat{z}|}}{4\pi|\mathbf{r}-z'\hat{z}|}\hat{\theta}',$$

where (see margin)

$$\cos\theta' = \hat{z} \cdot \frac{\mathbf{r}-z'\hat{z}}{|\mathbf{r}-z'\hat{z}|};$$

the implication is then

$$\int \tilde{I}(z')\delta(z-z')dz' = \tilde{I}(z) \rightarrow \boxed{\hat{z}\text{-pol radiator}} \rightarrow \int j\eta_o \tilde{I}(z') k \sin\theta' \frac{e^{-jk|\mathbf{r}-z'\hat{z}|}}{4\pi|\mathbf{r}-z'\hat{z}|}\hat{\theta}' dz' = \tilde{E}(\mathbf{r})$$



Field due to dipole at the origin

$$\tilde{\mathbf{E}} = j\eta_o \tilde{I}(0) k \Delta z \sin\theta \frac{e^{-jkr}}{4\pi r}\hat{\theta}$$

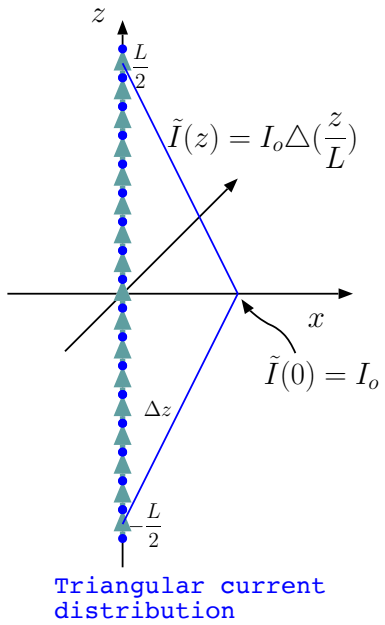
Field due to displaced dipole

$$\tilde{\mathbf{E}} = j\eta_o \tilde{I}(z') k \Delta z \sin\theta' \frac{e^{-jk|\mathbf{r}-z'\hat{z}|}}{4\pi|\mathbf{r}-z'\hat{z}|}\hat{\theta}'$$

- The final result, the expression

$$\tilde{E}(\mathbf{r}) = \int j\eta_o \tilde{I}(z') k \sin\theta' \frac{e^{-jk|\mathbf{r}-z'\hat{z}|}}{4\pi|\mathbf{r}-z'\hat{z}|}\hat{\theta}' dz'$$

for the radiation electric field phasor is in fact very general, and applicable to dipole antennas of *all lengths* provided that the current distribution $\tilde{I}(z)$ on the dipole is known.



- In practice, the triangular current distribution

$$\tilde{I}(z) = I_o\triangle(\frac{z}{L})\,\mathrm{A}$$

we have described earlier turns out to be applicable only when the dipole length

$$L \ll \lambda = \frac{c}{f}$$

at the operation frequency $f = \frac{\omega}{2\pi}$. For such dipoles

$$\begin{aligned}
\tilde{E}(\mathbf{r}) &= \int j\eta_o I_o \triangle(\frac{z'}{L}) k \sin\theta' \frac{e^{-jk|\mathbf{r}-z'\hat{z}|}}{4\pi|\mathbf{r}-z'\hat{z}|}\hat{\theta}' dz' \\
&= j\eta_o I_o k \int \triangle(\frac{z'}{L}) \sin\theta' \frac{e^{-jk|\mathbf{r}-z'\hat{z}|}}{4\pi|\mathbf{r}-z'\hat{z}|}\hat{\theta}' dz' \\
&\approx j\eta_o I_o k \underbrace{\{\int \triangle(\frac{z'}{L}) dz'\}}_{L/2,\ \text{area of a triangle with height 1 and base } L} \sin\theta \frac{e^{-jk|\mathbf{r}|}}{4\pi|\mathbf{r}|}\hat{\theta} = j\eta_o I_o k \frac{L}{2}\sin\theta \frac{e^{-jkr}}{4\pi r}\hat{\theta},
\end{aligned}$$

where the condition $L \ll \lambda$ is used to justify the replacement of $|\mathbf{r} - z'\hat{z}|$ by $|\mathbf{r}| = r$.

- Notice that the result

$$\tilde{E}(\mathbf{r}) = j\eta_o I_o k \frac{L}{2} \sin\theta \frac{e^{-jkr}}{4\pi r} \hat{\theta}$$

is identical with the radiation field of the Hertzian dipole except that

  - infinitesimal length Δz has been replaced by a finite length $\frac{L}{2}$ corresponding to the **dipole half-length**.

- The corresponding radiation magnetic field of the short dipole is

$$\tilde{H}(\mathbf{r}) = jI_o k \frac{L}{2} \sin\theta \frac{e^{-jkr}}{4\pi r} \hat{\phi}.$$

  - Dipole *half-length* $\frac{L}{2}$ is also known as **effective length** of the short dipole antenna.

- The term **effective length** is used more broadly to denote

$$\ell(\theta) \equiv \int \frac{\tilde{I}(z)}{I_o} e^{jkz\cos\theta} dz$$

defined for any length dipole antenna having a phasor current distribution $\tilde{I}(z)$ and a phasor current I_o at the input port (or input terminals).

Radiation fields of the short dipole:

$$\tilde{\mathbf{E}} = j\eta_o I_o k \frac{L}{2} \sin\theta \frac{e^{-jkr}}{4\pi r} \hat{\theta}$$

and

$$\tilde{\mathbf{H}} = jI_o k \frac{L}{2} \sin\theta \frac{e^{-jkr}}{4\pi r} \hat{\phi}.$$

– For a **short dipole** with the current distribution

$$\tilde{I}(z) = I_o \triangle(\frac{z}{L}),$$

where $L \ll \lambda$, this definition yields

$$\ell(\theta) = \frac{L}{2}.$$

– For a **half-wave dipole** with the current distribution

$$\tilde{I}(z) = I_o \text{rect}(\frac{z}{L}) \cos(kz),$$

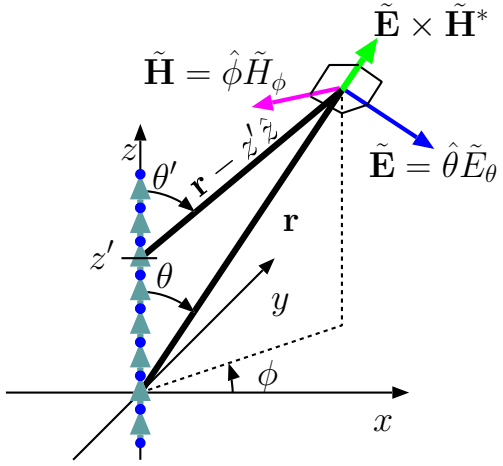
where $L = \frac{\lambda}{2}$, this definition yields

$$\ell(\theta) = \frac{\lambda}{\pi} \frac{\cos(\frac{\pi}{2}\cos\theta)}{\sin^2\theta},$$

as will be shown in ECE 454.

Effective length

$$\ell(\theta) \equiv \int \frac{\tilde{I}(z)}{I_o} e^{jkz\cos\theta} dz.$$



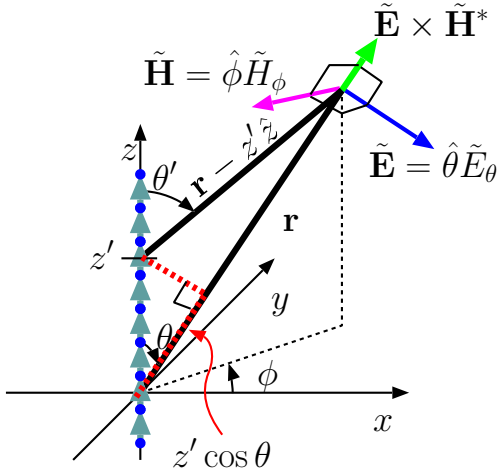
Radiation fields of all linearly polarized antennas can be obtained from those of the Hertzian dipole by replacing "Δz" with an appropriate effective length "$\ell(\theta)$" as illustrated above.

- The justification of this general rule is as follows:

Replacing $\triangle(\frac{z'}{L})$ with an arbitrary $\frac{\tilde{I}(z')}{I_o}$ in the second line of the above expression for $\tilde{E}(\mathbf{r})$, we have

$$\begin{aligned}\tilde{E}(\mathbf{r}) &= j\eta_o I_o k \int \frac{\tilde{I}(z')}{I_o} \sin\theta' \frac{e^{-jk|\mathbf{r}-z'\hat{z}|}}{4\pi|\mathbf{r}-z'\hat{z}|}\hat{\theta}' dz' \\ &\approx j\eta_o I_o k \{\int \frac{\tilde{I}(z')}{I_o} e^{-jk|\mathbf{r}-z'\hat{z}|} dz'\} \sin\theta \frac{1}{4\pi|\mathbf{r}|}\hat{\theta},\end{aligned}$$

where in the second line we have replaced all occurrences of θ' by θ and $|\mathbf{r}-z'\hat{z}|$ by $|\mathbf{r}| = r$, except for in the complex exponential which is highly sensitive to z'.



- Replacements outside the exponential are easily justified with any finite length L (large or small) for $r \gg L$.

- Treating $k|\mathbf{r}-z'\hat{z}|$ as $k|\mathbf{r}| = kr$ in the exponent cannot be permitted when k is large,

  $$\text{but approximating } \; k|\mathbf{r}-z'\hat{z}| \; \text{ as } \; kr - kz'\cos\theta$$

  can be tolerated for sufficiently large r in the so-called **paraxial approximation**[1] to obtain

  $$e^{-jk|\mathbf{r}-z'\hat{z}|} \approx e^{-jk(r-z'\cos\theta)} = e^{-jkr}e^{jkz'\cos\theta},$$

[1] ... also known as *small angle approximation* in reference to the angle between the vectors $\mathbf{r}$ and $\mathbf{r}-z'\hat{z}$ — see the margin plot. In this approximation we pretend that the vectors $\mathbf{r}$ and $\mathbf{r}-z'\hat{z}$ are parallel to one another!

leading in to

$$\tilde{E}(\mathbf{r}) \;\approx\; j\eta_o I_o k \underbrace{\int \frac{\tilde{I}(z')}{I_o} e^{jkz'\cos\theta} dz'}_{\equiv \ell(\theta)} \sin\theta \frac{e^{-jkr}}{4\pi r}\hat{\theta},$$

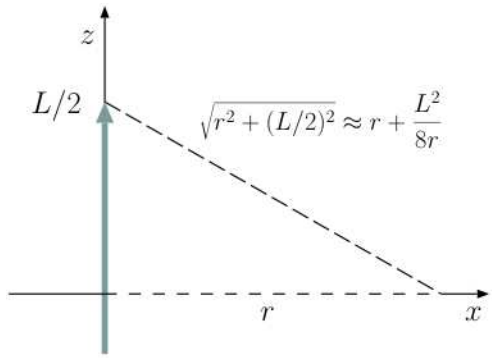
a general radiation field expression formulated in terms of effective length $\ell(\theta)$.

- This result is certainly valid for all $r \gg L$ where it makes sense to consider $\mathbf{r} - z'\hat{z}$ and $\mathbf{r}$ to be parallel vectors.
- The validity limit of paraxial approximation can be investigated more carefully by expanding $k|\mathbf{r} - z'\hat{z}|$ to a higher order, and finding under what condition high-order correction factors are really unnecessary — that exercise (see the margin) shows that paraxial approximation is well justified for

  $$r \gtrsim \frac{2L^2}{\lambda},$$

  where the "threshold distance" $2L^2/\lambda$ is known as **Rayleigh distance**.

- Even though we have developed a general representation for the radiation fields of arbitrary dipoles in this lecture, our discussions over the next few lectures will focus mainly on short dipoles as our basic radiation elements.



In paraxial approximation the dashed lines are treated as having equal lengths (which is only accurate for r going to infinity), leading to a phase error of

$$\frac{kL^2}{8r} = \frac{\pi}{8}\frac{2L^2}{\lambda r} \text{ rad.}$$

The phase error is less than

$\frac{\pi}{8}$ rad and tolerable if $r > \frac{2L^2}{\lambda}$.

Alternatively,

$$\begin{aligned} k|\mathbf{r} - z'\hat{z}| &= k\sqrt{x^2 + y^2 + (z - z')^2} \\ &= kr\sqrt{1 - 2zz'/r^2 + z'^2/r^2}, \end{aligned}$$

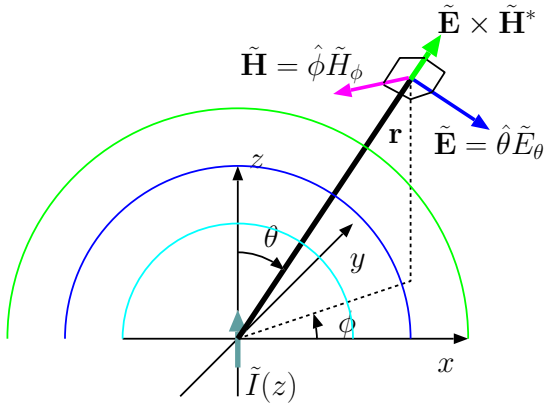
where $r = \sqrt{x^2 + y^2 + z^2}$. Hence

$$\begin{aligned} k|\mathbf{r} - z'\hat{z}| &\approx kr(1 - zz'/r^2 + z'^2/r^2/2) \\ &= kr - k\cos\theta z' + kz'^2/(2r) \end{aligned}$$

provided that $z' \ll r$.

Finally the term $kz'^2/(2r)$ can be neglected above, even with maximal values at $z' = \pm L/2$, leading to the *paraxial approximation* result $|\mathbf{r} - z'\hat{z}| \approx r - z'\cos\theta$, if $k(L/2)^2/(2r) < \pi/8$, the same as $r > 2L^2/\lambda$.

9 Poynting vector, radiated power, radiation resistance



Consider the radiation fields of a $\hat{z}$-polarized short-dipole antenna shown in the margin:

- How much average power is radiated by the short-dipole antenna to sustain these fields, and
- how can we determine this amount, P_{rad}, by electrical measurements which can be performed at the antenna input port — the small gap at the dipole center where the dipole is connected to the source circuit (typically via some transmission line network)?

To answer these questions we will calculate in this lecture the **average Poynting vector** of radiation fields of the dipole antenna and the "flux" of the same vector computed over a sphere imagined to surround the dipole.

- Recall once again that Poynting vector

$$\mathbf{S} \equiv \mathbf{E} \times \mathbf{H}$$

denotes the energy transported by electromagnetic fields per unit time and per unit area normal to the vector itself. With time-harmonic fields the average value of Poynting vector can be denoted and computed as

$$\langle \mathbf{E} \times \mathbf{H} \rangle = \frac{1}{2}\mathrm{Re}\{\tilde{\mathbf{E}} \times \tilde{\mathbf{H}}^*\}$$

in terms of field phasors $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{H}}$.

Radiation fields of short dipole:

$$\tilde{\mathbf{E}} = \tilde{E}_\theta \hat{\theta}$$

and

$$\tilde{\mathbf{H}} = \frac{\tilde{E}_\theta}{\eta_o}\hat{\phi}$$

where

$$\tilde{E}_\theta = j\eta_o I_o k\ell \sin\theta \frac{e^{-jkr}}{4\pi r}$$

and $\ell = L/2$**.**

- It is this quantity

$$\langle \mathbf{S} \rangle = \langle \mathbf{E} \times \mathbf{H} \rangle$$

which is independent of the storage fields of dipole antennas and only depend on their radiation fields.

- Using (see margin once again)

$$\tilde{\mathbf{E}} = \hat{\theta}\tilde{E}_\theta \quad \text{and} \quad \tilde{\mathbf{H}} = \hat{\phi}\frac{\tilde{E}_\theta}{\eta_o} \quad \text{with} \quad \tilde{E}_\theta = j\eta_o I k\ell \sin\theta \frac{e^{-jkr}}{4\pi r},$$

we find

$$\tilde{\mathbf{E}} \times \tilde{\mathbf{H}}^* = \hat{\theta}\tilde{E}_\theta \times (\hat{\phi}\frac{\tilde{E}_\theta}{\eta_o})^* = \hat{\theta} \times \hat{\phi}\frac{|\tilde{E}_\theta|^2}{\eta_o} = \hat{r}\frac{|\tilde{\mathbf{E}}|^2}{\eta_o}$$

and

$$\langle \mathbf{E} \times \mathbf{H} \rangle = \frac{1}{2}\mathrm{Re}\{\tilde{\mathbf{E}} \times \tilde{\mathbf{H}}^*\} = \hat{r}\frac{|\tilde{\mathbf{E}}|^2}{2\eta_o}.$$

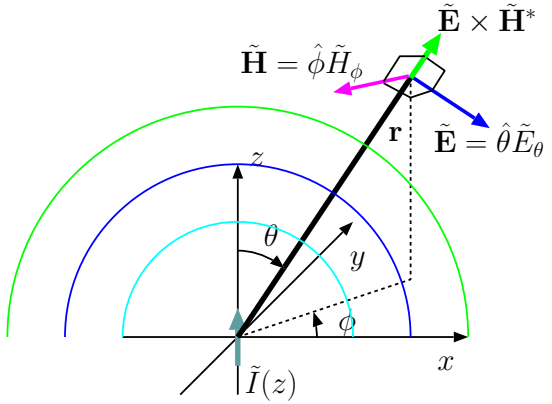
- Since

$$|\tilde{\mathbf{E}}|^2 = |\tilde{E}_\theta|^2 = \frac{\eta_o^2 |I_o|^2 k^2 |\ell|^2 \sin^2\theta}{(4\pi r)^2} = \frac{\eta_o^2 |I_o|^2 |\ell|^2 \sin^2\theta}{4(\lambda r)^2},$$

we have

$$\langle \mathbf{E} \times \mathbf{H} \rangle = \hat{r}\frac{|\tilde{\mathbf{E}}|^2}{2\eta_o} = \frac{\eta_o}{8}|I_o|^2 \frac{|\ell|^2}{(\lambda r)^2}\sin^2\theta \hat{r}.$$

- The expression above is the **energy flux density** or **transmitted power density** of the dipole antenna as a function of distance r from the dipole and angle θ of viewing direction off the dipole axis.



Radiation fields of short dipole:

$$\tilde{\mathbf{E}} = \tilde{E}_\theta \hat{\theta}$$

and

$$\tilde{\mathbf{H}} = \frac{\tilde{E}_\theta}{\eta_o}\hat{\phi}$$

where

$$\tilde{E}_\theta = j\eta_o I_o k\ell \sin\theta \frac{e^{-jkr}}{4\pi r}$$

and $\ell = L/2$**.**

- The average power output of the dipole — radiated power P_{rad} — can next be obtained by computing the flux of $\langle \mathbf{E} \times \mathbf{H} \rangle$ over any closed surface surrounding the dipole.

  – This calculation is most easily carried out over a spherical surface of radius r having infinitesimal surface elements

  $$d\mathbf{S} = \hat{r}(r\sin\theta d\phi)(rd\theta) = \hat{r}r^2\sin\theta d\theta d\phi \equiv \hat{r}r^2 d\Omega,$$

  where

  $$d\Omega \equiv \sin\theta d\theta d\phi$$

  (introduced to maintain a compact notation) is called a **solid angle** increment.

We then note that

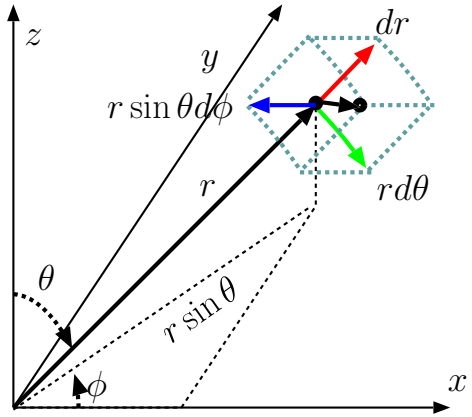
$$\langle \mathbf{E} \times \mathbf{H} \rangle \cdot d\mathbf{S} = \frac{\eta_o}{8\lambda^2}|I_o|^2|\ell|^2\sin^2\theta d\Omega$$

and the flux of $\langle \mathbf{E} \times \mathbf{H} \rangle$ is

$$\underbrace{\oint \langle \mathbf{E} \times \mathbf{H} \rangle \cdot d\mathbf{S}}_{P_{rad}} = \frac{\eta_o}{8\lambda^2}|I_o|^2 \int d\Omega |\ell|^2 \sin^2\theta$$

where it is implied that

$$\int d\Omega = \int_{\phi=0}^{2\pi} d\phi \int_{\theta=0}^{\pi} d\theta \sin\theta.$$



Infinitesimal area on a constant r surface is

$$\begin{aligned} dS &= (rd\theta)(r\sin\theta d\phi) \\ &= r^2 d\Omega \end{aligned}$$

where

$$d\Omega \equiv \sin\theta d\theta d\phi$$

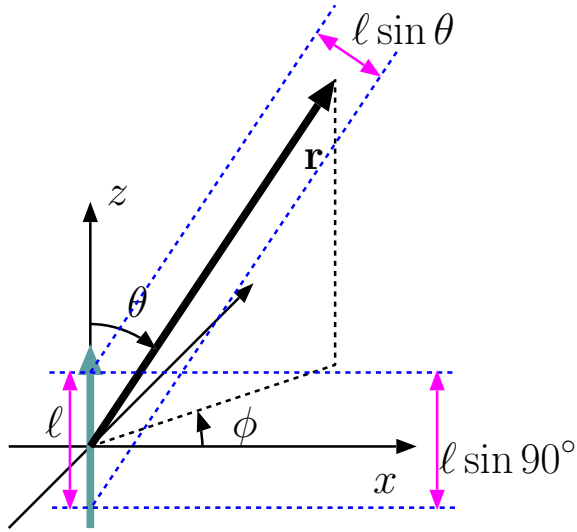
is called infinitesimal solid angle.

– This result can be cast as

$$P_{rad} = \frac{1}{2} R_{rad} |I_o|^2 \quad \text{where} \quad R_{rad} = \frac{\eta_o}{4\lambda^2} \int d\Omega |\ell \sin\theta|^2$$

is known as **radiation resistance**.

- If ℓ is the *effective length* of a dipole — distinct from its physical length L because of current weighting — then $\ell \sin\theta$ is "how long the effective length looks" when one sees it (the dipole) at an angle (see margin).



Note: recall that ℓ may be a function of θ itself!

Solid angle integral of the square of this "foreshortened" effective length, namely

$$\int d\Omega |\ell \sin\theta|^2,$$

determines the radiation resistance of the dipole antenna.

Since for a short dipole $\ell = \frac{L}{2}$ is independent of angle θ (unlike for half-wave dipole), we have

$$\begin{aligned} \int d\Omega |\ell \sin\theta|^2 &= (\frac{L}{2})^2 \int d\Omega |\sin\theta|^2 = (\frac{L}{2})^2 \int_0^{2\pi} d\phi \int_0^{\pi} d\theta \sin\theta |\sin\theta|^2 \\ &= (\frac{L}{2})^2 2\pi \underbrace{\int_0^{\pi} d\theta \sin\theta |\sin\theta|^2}_{4/3} = \frac{2\pi L^2}{3}. \end{aligned}$$

Hence the radiation resistance of the short dipole is

$$R_{rad} = \frac{\eta_o}{4\lambda^2} \int d\Omega |\ell \sin\theta|^2 = \frac{\eta_o}{4\lambda^2} \frac{2\pi L^2}{3} = 20\pi^2 (\frac{L}{\lambda})^2 \, \Omega.$$

- Since a short dipole is constrained to have $\frac{L}{\lambda} \ll 1$, say $\frac{1}{10}$ or smaller, R_{rad} will be equal to or less than about $2\,\Omega$.

- Thus, a short dipole with an input current of $\tilde{I}(0) = I_o = 1$ A will have at best an average power output $\frac{1}{2} I_o^2 R_{rad}$ of about 1 W.

  This is not quite at the level of 100s of W's of power that typical radio stations transmit!

  Using antennas with higher R_{rad} than a short dipole[1] — e.g., a half-wave dipole for which $R_{rad} \approx 73\,\Omega$ — is the best way of addressing this difficulty since the alternate solution of increasing I_o (as needed) is not recommended because of *antenna losses*:

  – In practice, antennas appear as a circuit element with input resistance

  $$R_o = R_{rad} + R_{loss}$$

  where R_{loss} represents ohmic losses (heating of antenna wires) — an antenna consumes an average power of

  $$\frac{1}{2} I_o^2 (R_{rad} + R_{loss})$$

  out of which only

  $$\frac{1}{2} I_o^2 R_{rad}$$

[1] Short dipoles are typically employed as receiving antennas rather than transmitting antennas because of this. Receiving properties of antennas are closely related to their transmission properties, but figures of merit of antennas pertinent in transmission and reception are somewhat different as we will learn later on in the course.

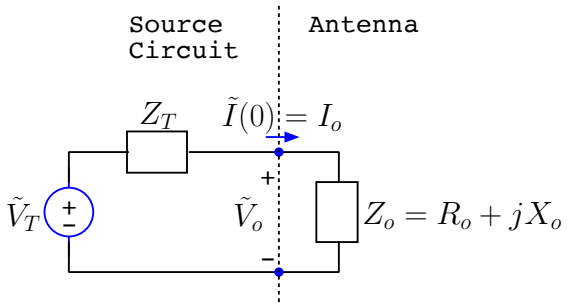
is the useful radiated power.

Typically $R_{loss} \propto L$, whereas $R_{rad} \propto L^2$ for small L, so going to longer dipoles (and learning more about them in ECE 454) really helps.

- A source circuit connected to the antenna terminals "sees" the antenna (and the radiation volume with which the antenna interacts) as a two-terminal element having some impedance

$$Z_o \equiv \frac{\tilde{V}_o}{\tilde{I}(0)} = R_o + jX_o$$

known as **antenna impedance**.



  - We have already discussed the **resistive component** R_o above.
  - Modeling the **reactive component** X_o requires working with storage fields of the antenna as well, matching components of total fields to proper boundary conditions imposed by the actual surfaces of antenna wires (i.e., antenna geometry needs to be specified in detail before X_o can be determined).

    Antenna reactance will be examined in some detail in ECE 454 (along with methods of calculating $\tilde{I}(z)$).
  - We will not need to calculate antenna reactances in this course. However, it is worth mentioning that
    1. antennas with $X_o = 0$ are known as **resonant antennas**, and
    2. half-wave dipole is a resonant antenna (see margin note).

Antenna reactance:

Short dipoles **have** ***capacitive reactances*****, just like the line-impedance at a small distance away from an open termination on a transmission line.**

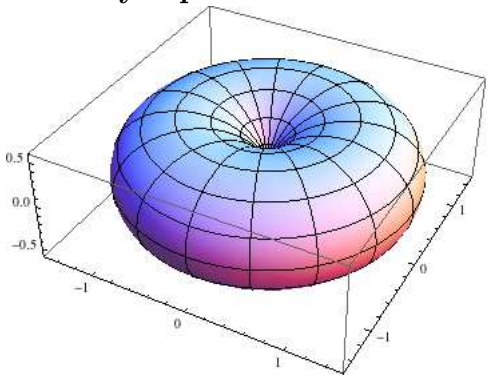
Capacitive reactance switches to an inductive one when the dipole length is about $\lambda/2$, just like the line-impedance at a distance $\lambda/4$ away from an open termination.

Thus the ***half-wave dipole*** **is** ***resonant*****, having a zero input reactance — $Z_o = R_{rad} + j0$ for an ideal half-wave dipole.**

In practice, resonant half-wave dipole with a length L and wire radius a has $L + 2a = \lambda/2$ to a good approximation.

10 Antenna gain, beam pattern, directivity

Beam pattern plot of low directivity dipole antenna:



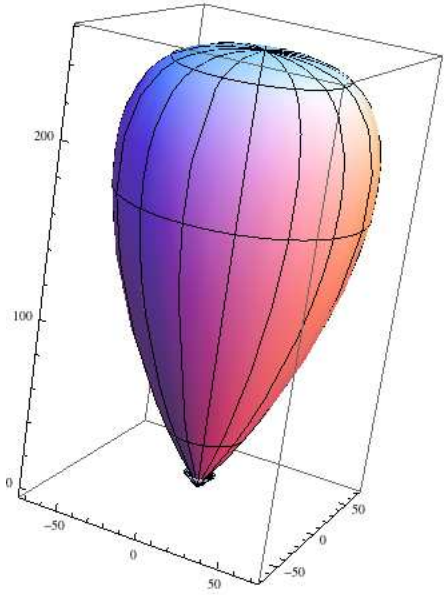
- A dipole antenna (or a closely related monopole to be studied in Lecture 18) is a "near perfect" radiator for purposes of "broadcasting" — that is, sending waves of equal amplitudes in all directions to reach out multiple targets or receivers.

- However dipole is a poor choice when the objective is to radiate the power P_{rad} in a specific direction (i.e., towards a specific receiver), as in

  - **communication** with deep space probes or orbiting satellites, or with
  - **radar beacons** where the objective is to determine the direction of a moving target.

  In such applications we need **high-gain** and **directive** antennas, as opposed to low-gain and non-directive antennas such as a single dipole.

A higher directivity beam pattern of an array antenna



- Qualitatively speaking, gain and directivity of an antenna measures its ability to confine its radiated wave fields within a narrow field of view called the **antenna beam** or **beam pattern**.

  - when a narrow antenna beam is achieved, and all the radiated power P_{rad} of the antenna is conveyed through this beam, the power density of the waves is naturally high within the beam.

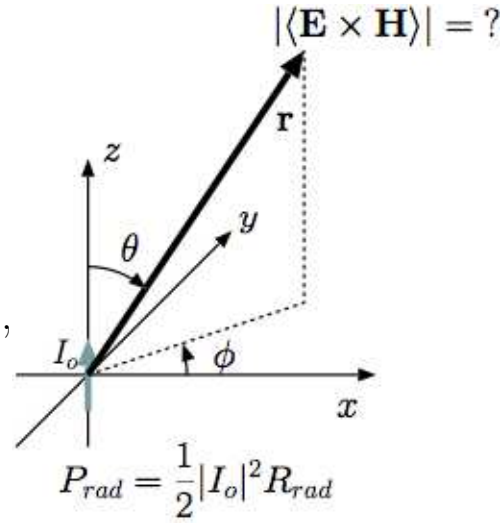
Arrays **of dipoles can serve as high-gain antennas needed in beaming applications as we will learn in the next lecture.**

In this lecture we will focus on the definition of antenna gain and directivity as well as the related concept of beam solid angle.

- Consider an antenna located at the origin with an input current of I_o, radiation resistance R_{rad}, and a radiated power

$$P_{rad} = \frac{1}{2}|I_o|^2 R_{rad}\,\text{Watts}.$$

What would be the time-average Poynting magnitude $|\langle \mathbf{E}(\mathbf{r},t)\times\mathbf{H}(\mathbf{r},t)\rangle|$, i.e., the power density in Watts/m^2 of the radiation fields at a location $\mathbf{r} = (r,\theta,\phi)$ a distance r away from the antenna?



Power density of a radiating antenna in the far field

Answer:

– If the antenna were an **isotropic radiator** then we would have a **power density** of

$$|\langle \mathbf{E}(\mathbf{r},t)\times\mathbf{H}(\mathbf{r},t)\rangle| = \frac{P_{rad}}{4\pi r^2};$$

however, no real antenna is an isotropic radiator, and thus the correct answer can be formally cast as

$$|\langle \mathbf{E}(\mathbf{r},t)\times\mathbf{H}(\mathbf{r},t)\rangle| = \frac{P_{rad}}{4\pi r^2}G(\theta,\phi)$$

in terms of an **antenna gain** (over isotropic radiator)

$$G(\theta,\phi) \equiv \frac{|\langle \mathbf{E}(\mathbf{r},t)\times\mathbf{H}(\mathbf{r},t)\rangle|}{\frac{P_{rad}}{4\pi r^2}},$$

to be determined.

Clearly, gain $G(\theta,\phi)$ is the ratio of the radiated average power density of an antenna to that of an **isotropic radiator** (hypothetical perfect broadcasting antenna) radiating the same average power P_{rad}.

– According to this definition, the solid angle integral of gain $G(\theta,\phi)$ is

$$\begin{aligned}\int d\Omega G(\theta,\phi) &\equiv \frac{4\pi\int d\Omega r^2|\langle \mathbf{E}(\mathbf{r},t)\times\mathbf{H}(\mathbf{r},t)\rangle|}{P_{rad}} \\ &= \frac{4\pi\oint\langle \mathbf{E}(\mathbf{r},t)\times\mathbf{H}(\mathbf{r},t)\rangle\cdot d\mathbf{S}}{P_{rad}} = 4\pi, \text{ a fixed value.}\end{aligned}$$

Since

$$G(\theta,\phi)\propto|\langle\mathbf{E}\times\mathbf{H}\rangle|\propto|\ell\sin\theta|^2,$$

we can write

$$G(\theta,\phi)=K|\ell\sin\theta|^2$$

in terms of a proportionality constant K, which is subsequently identified as

$$K=\frac{4\pi}{\int d\Omega|\ell\sin\theta|^2} \text{ after applying the constraint } \int d\Omega G(\theta,\phi)=4\pi.$$

Thus we obtain a general gain formula

$$G(\theta,\phi) = \frac{4\pi|\ell \sin\theta|^2}{\int d\Omega |\ell \sin\theta|^2}$$

applicable to all antennas for which the foreshortened effective length $\ell \sin\theta$ is known.

- For an arbitrary antenna, gain calculation can be complicated because of the solid angle integral in the denominator in $G(\theta,\phi)$ formula.

  However, for a short dipole with $\ell = L/2$ the calculation is simple and leads to (in case of $\hat{z}$-polarization)

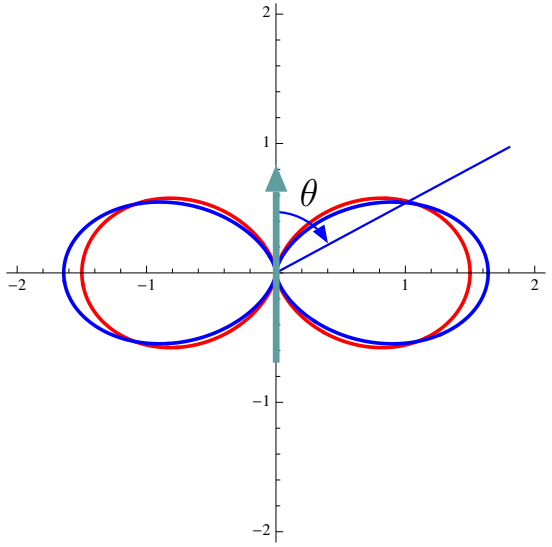
$$G(\theta,\phi) = \frac{4\pi|\sin\theta|^2}{\int d\Omega |\sin\theta|^2} = \frac{4\pi|\sin\theta|^2}{2\pi\frac{4}{3}} = \frac{3}{2}\sin^2\theta.$$

  For a half-wave dipole it works out that

$$G(\theta,\phi) \approx 1.64\,\frac{\cos^2(\frac{\pi}{2}\cos\theta)}{\sin^2\theta}$$

  with a maximum value of 1.64 at $\theta = 90°$.

  – Having maximum gains of 1.5 and 1.64, respectively, short- and half-wave-dipoles are considered to be **low-directivity** antennas.



Gain functions $G(\theta,\phi)$ depicted on a constant ϕ plane for
(a) short-dipole (red curve), and
(b) half-wave dipole (blue curve).

Directivity D of any antenna is defined to be the maximum value of its gain $G(\theta, \phi)$, i.e.,

$$D = G(\theta, \phi)_{max}.$$

While the solid angle integral of $G(\theta, \phi)$ is constrained to have a fixed value of 4π, there is no constraint on the maximum value of $G(\theta, \phi)$; therefore, it is possible to design antennas with arbitrarily large directivities D by making the antenna beam shape arbitrarily narrow.

- Note that the constraint

$$\int d\Omega G(\theta, \phi) = 4\pi$$

implies

$$D \int d\Omega \frac{G(\theta, \phi)}{G(\theta, \phi)_{max}} = 4\pi,$$

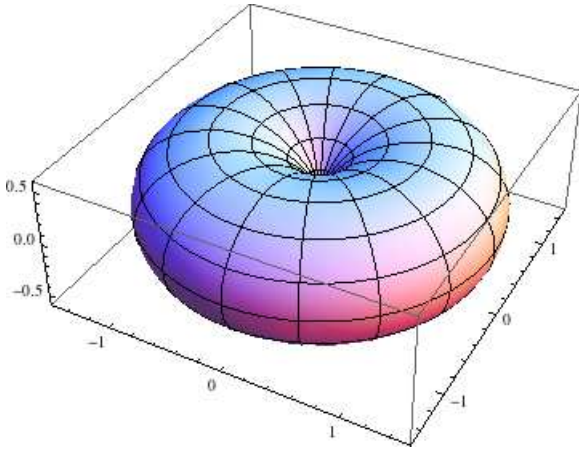
which can also be written as

$$D\Omega_o = 4\pi.$$

in terms of **beam solid angle**

$$\Omega_o \equiv \int d\Omega \frac{G(\theta, \phi)}{G(\theta, \phi)_{max}}$$

to be discussed further in this lecture.



Gain function

$$G(\theta, \phi) = \frac{3}{2} \sin^2 \theta$$

of *short-dipole* depicted as a 3D polar plot — gain in any direction (θ, ϕ) is proportional to the radius vector from the origin to the depicted surface.

A short-dipole has a low directivity of

$$D = 1.5$$

because it radiates with a broad beam that is isotropic in azimuth.

Antennas with high-directivity have narrow and pointy beam shapes.

- **Important result:** the product of antenna directivity D and the beam solid angle Ω_o is fixed, specifically

$$D\Omega_o = 4\pi,$$

which implies that if D is large then Ω_o is small and vice versa.

- A useful method to determine the antenna directivity is to use

$$D = \frac{4\pi}{\Omega_o},$$

where the solid angle

$$\Omega_o = \int d\Omega \frac{G(\theta,\phi)}{G(\theta,\phi)_{max}} = \int d\Omega \frac{|\ell \sin\theta|^2}{|\ell \sin\theta|^2_{max}}$$

can be calculated once the antenna effective length is known.

Example 1: For a short dipole with $\ell = L/2$, we have

$$\Omega_o = \int d\Omega \frac{|\sin\theta|^2}{|\sin\theta|^2_{max}} = \int d\Omega |\sin\theta|^2 = 2\pi\frac{4}{3} = \frac{8\pi}{3}.$$

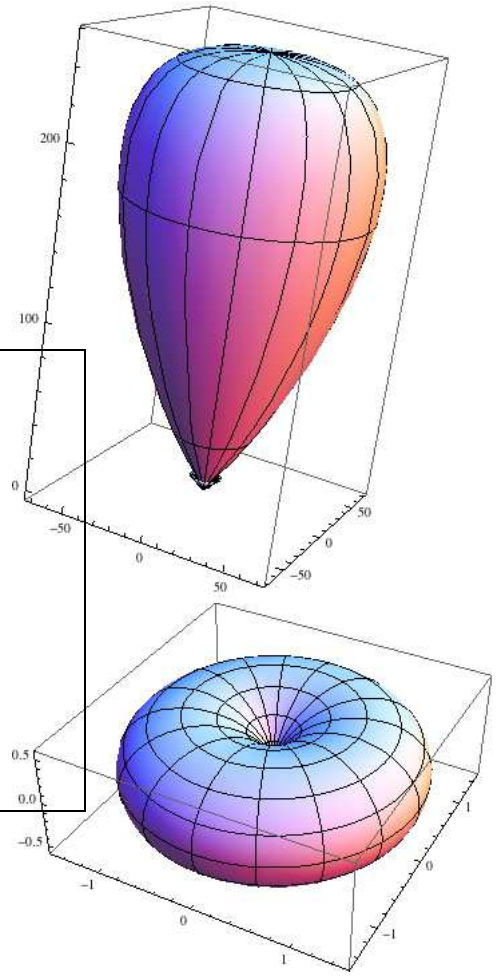
Consequently,

$$D = \frac{4\pi}{\Omega_o} = \frac{4\pi}{8\pi/3} = 1.5$$

consistent with what we learned above.

  - This method of finding D from Ω_o is very useful because there are geometrical methods for estimating Ω_o in terms of the physical antenna size (as we will learn later on).

Antennas with high-directivity have narrow and pointy beam shapes.

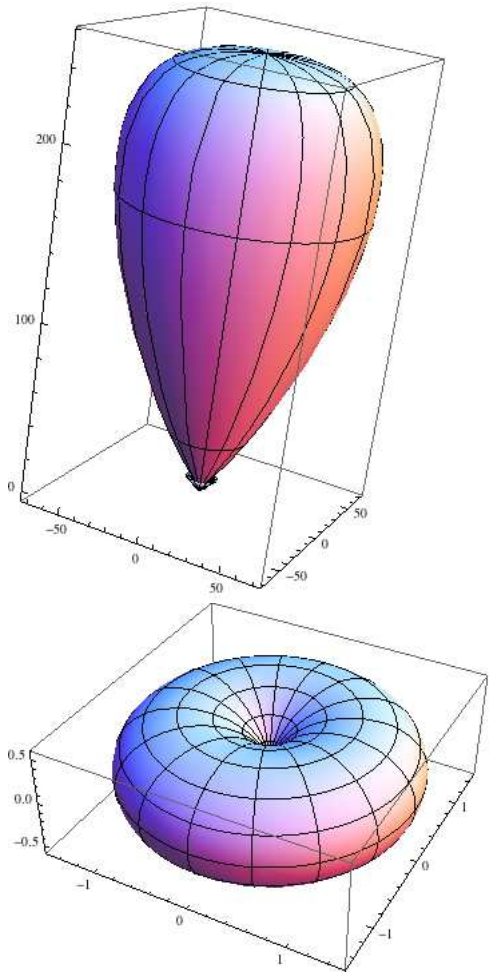


Once D is determined, the gain of the antenna can be written as

$$G(\theta,\phi) = D\frac{|\ell \sin\theta|^2}{|\ell \sin\theta|^2_{max}}$$

without the need to perform a solid angle integral in practice.

Antennas with high-directivity have narrow and pointy beam shapes.



- The beam solid angle

$$\Omega_o = \int d\Omega \frac{G(\theta,\phi)}{G(\theta,\phi)_{max}} = \int d\Omega \frac{|\ell \sin\theta|^2}{|\ell \sin\theta|^2_{max}}$$

extends the concept of "angle" from 2D to 3D to describe the angular width of the antenna beam pattern. Let us examine this parameter more closely.

- Ordinary **angles** ranging from 0 to 2π **radians** (with degree equivalents ranging from 0 to 360) correspond to **arc lengths** measured on unit-radius circles drawn on 2D planar surfaces.

- **Solid angles** ranging from 0 to 4π **steradians** correspond to **areas of patches** or **spots** specified on unit-radius spheres defined in 3D space.

  – An **antenna-beam solid angle**

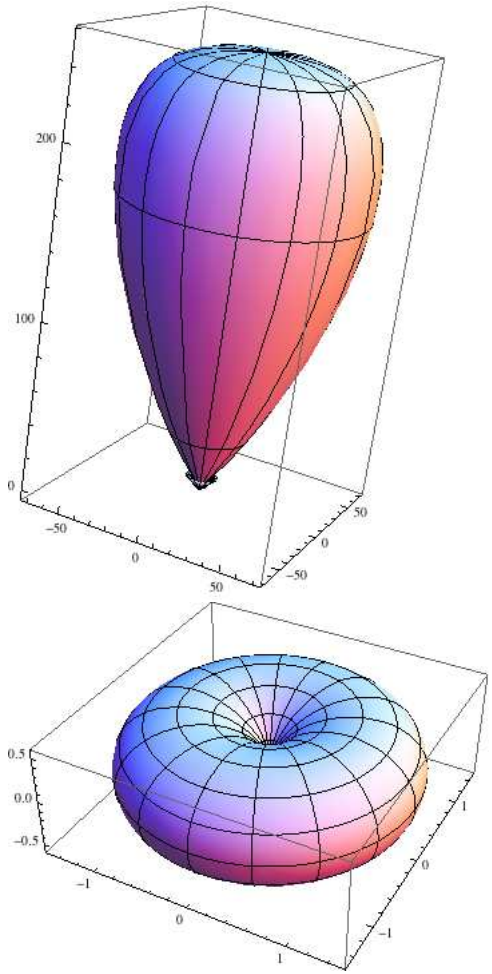
$$\Omega_o = \int d\Omega \frac{G(\theta,\phi)}{G(\theta,\phi)_{max}} = \int d\Omega \frac{|\langle \mathbf{E}\times\mathbf{H}\rangle|}{|\langle \mathbf{E}\times\mathbf{H}\rangle|_{max}}$$

is an **equivalent area** of a spot or a patch (centered about the direction of $|\langle \mathbf{E}\times\mathbf{H}\rangle|_{max}$) specified on a unit sphere surrounding

the antenna, having the property that the entire power output P_{rad} of the antenna would flood this area with an equal flux density of $|\langle \mathbf{E} \times \mathbf{H} \rangle|_{max}$ if the beam were reformed into a conical shape.

– Beam shapes of high-directivity antennas with small Ω_o can be well represented by equivalent conical beams, but such a representation is not appropriate to dipole-like broadcast antennas (see margin).

Antennas with high-directivity have narrow and pointy beam shapes.



Example 2: For a short dipole with

$$G(\theta, \phi) = \frac{3}{2} \sin^2 \theta$$

we have

$$D = \frac{3}{2} \quad \text{and} \quad \Omega_o = \int d\Omega \frac{|\sin\theta|^2}{|\sin\theta|^2_{max}} = 2\pi \frac{4}{3} = \frac{8\pi}{3}$$

as we already established in Example 1.

Consequently,

$$D = \frac{4\pi}{\Omega_o} = \frac{4\pi}{8\pi/3} = 1.5$$

consistent with what we learned above.

Example 3: An antenna designer comes up with a model that has a gain function specified as

$$G(\theta,\phi)=\begin{cases} D\sin^2\theta, & 0<\theta<\frac{\pi}{2} \\ 0, & \text{othewise,} \end{cases}$$

where D is the antenna directivity. Determine both D and the beam solid angle Ω_o.

Solution: Since the solid angle integral of $G(\theta,\phi)$ has to equal 4π, it must be true that

$$\int d\Omega G(\theta,\phi)=\int_{\phi=0}^{2\pi}d\phi\int_{\theta=0}^{\pi/2}d\theta\sin\theta D\sin^2\theta=4\pi.$$

It follows that

$$2\pi D\int_{\theta=0}^{\pi/2}d\theta\sin\theta\sin^2\theta=4\pi \;\;\Rightarrow\;\; -D\int_{\theta=0}^{\pi/2}(d\cos\theta)(1-\cos^2\theta)=2$$

from which we get

$$D=\frac{2}{\int_{\theta=\pi/2}^{0}d\cos\theta(1-\cos^2\theta)}=\frac{2}{(\cos\theta-\frac{\cos^3\theta}{3}|_{\pi/2}^{0}}=\frac{2}{1-\frac{1}{3}}=3.$$

This is twice the directivity of the short-dipole (which makes sense because half the gain function of the short dipole is missing from the gain of this antenna).

As for the beam solid angle, it is

$$\Omega_o=\frac{4\pi}{D}=\frac{4\pi}{3},$$

which is half the solid angle of a short dipole (again for the same reason).

11 Beam pattern, wave interference

In this lecture we will see how antenna beams can be "patterned" by using interference effects of fields radiated by multiple dipoles or dipole-like elements.

- Let's recall that the **antenna beam** is the shape of the antenna gain function $G(\theta, \phi)$ that can be depicted as a surface plot in 3D.

  Also

  $$D = G(\theta, \phi)_{max} = \frac{4\pi}{\Omega_o} \quad \text{and} \quad \Omega_o = \int d\Omega \frac{G(\theta, \phi)}{G(\theta, \phi)_{max}} = \int d\Omega \frac{|\langle \mathbf{E} \times \mathbf{H} \rangle|}{|\langle \mathbf{E} \times \mathbf{H} \rangle|_{max}}$$

  as well as

  $$|\langle \mathbf{E} \times \mathbf{H} \rangle| = \frac{|\tilde{E}_\theta|^2}{2\eta_o} \quad \text{with} \quad \tilde{E}_\theta = j\eta_o I_o k\ell \sin\theta \frac{e^{-jkr}}{4\pi r}$$

  for $\hat{z}$-polarized antennas and elements.

  - With $\ell = L/2$ the above equations would represent a short dipole.
  - An antenna system constructed by an *array* of such dipoles would also be represented by the same equations, but with a different $\ell = \ell(\theta, \phi)$ (to be determined).

- The design and analysis of multi-element or multi-dipole arrays are facilitated by the **linearity** of wave solutions of Maxwell's equations:

  If

  $$\tilde{\mathbf{J}}_1 \rightarrow \boxed{\textbf{ME}} \rightarrow \tilde{\mathbf{E}}_1$$

  and

  $$\tilde{\mathbf{J}}_2 \rightarrow \boxed{\textbf{ME}} \rightarrow \tilde{\mathbf{E}}_2$$

  then

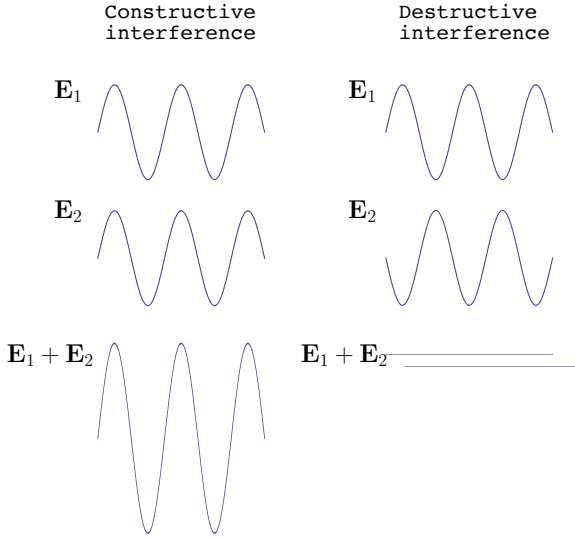
  $$\alpha\tilde{\mathbf{J}}_1 + \beta\tilde{\mathbf{J}}_2 \rightarrow \boxed{\textbf{ME}} \rightarrow \alpha\tilde{\mathbf{E}}_1 + \beta\tilde{\mathbf{E}}_2$$

  - If radiators $\tilde{\mathbf{J}}_1$ and $\tilde{\mathbf{J}}_2$ produce radiated wave solutions $\tilde{\mathbf{E}}_1$ and $\tilde{\mathbf{E}}_2$, respectively, then a radiator $\alpha\tilde{\mathbf{J}}_1 + \beta\tilde{\mathbf{J}}_2$ would produce a wave solution $\alpha\tilde{\mathbf{E}}_1 + \beta\tilde{\mathbf{E}}_2$ with arbitrary (complex) weights α and β.
  - By induction, the above principle of superposition can be extended to n elements.

  Note that this superposition principle applies at the level of fields rather than power. This is similar to superposition principle applying at the level of voltage and currents in circuit analysis.

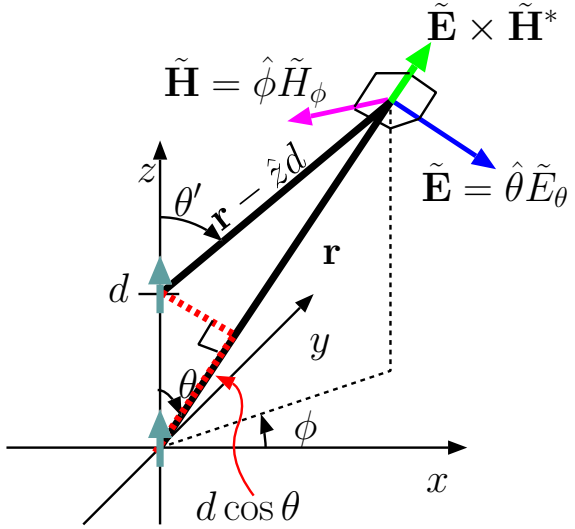
  Superposition of wave fields can produce resultant wave fields with enhanced or reduced wave amplitudes as a consequence of **interference** effects.

  - A **constructive interference** occurs at locations where the waves being superposed are "in phase", meaning that the phasors representing the wave fields are complex numbers having the same angle — i.e., $\angle\tilde{\mathbf{E}}_2 = \angle\tilde{\mathbf{E}}_1$.
  - A **destructive interference** occurs where the waves being superposed are "out of phase", meaning that the phasors representing the wave fields are complex numbers having an angle difference of $\pm 180°$ — i.e., $\angle\tilde{\mathbf{E}}_2 = \angle\tilde{\mathbf{E}}_1 \pm 180°$.



With a judicious choice of the locations and relative amplitudes of the radiators $\tilde{\mathbf{J}}_1$ and $\tilde{\mathbf{J}}_2$, it is possible to arrange for $\alpha\tilde{\mathbf{E}}_1 + \beta\tilde{\mathbf{E}}_2$ to exhibit constructive interference in desired beam directions — that is the essence of antenna beam design and designing high directivity antenna systems.

– One final detail before showing some examples: the calculation of the superposed wave fields is considerably simplified at distances r to the source elements that far exceed the largest distance separating the source elements.



Example 1: Two $\hat{z}$ polarized dipole antennas with equal input currents I_o are located at $(0,0,0)$ and $(0,0,d)$. Find the phasor expression $\tilde{\mathbf{E}}(\mathbf{r})$ representing the superposition of the fields radiated by each dipole individually. What are the maximum and minimum values of the field intensity $|\tilde{\mathbf{E}}(\mathbf{r})|$ as compared to intensity $|\tilde{\mathbf{E}}_1(\mathbf{r})|$ of the field due to the dipole at the origin?

Solution: First, the dipole at $(x,y,z) = (0,0,0)$ has a wave field phasor

$$\tilde{\mathbf{E}}_1(\mathbf{r}) = j\eta_o I_o k\ell \sin\theta \frac{e^{-jkr}}{4\pi r}\hat{\theta} = j\eta_o I_o k\ell \sin\theta \frac{e^{-jk|\mathbf{r}|}}{4\pi|\mathbf{r}|}\hat{\theta}.$$

The field phasor of the second dipole at $(x,y,z) = (0,0,d)$ is a shifted counterpart of $\tilde{\mathbf{E}}_1$, namely

$$\tilde{\mathbf{E}}_2(\mathbf{r}) = \tilde{\mathbf{E}}_1(\mathbf{r} - \hat{z}d) = j\eta_o I_o k\ell \sin\theta' \frac{e^{-jk|\mathbf{r}-\hat{z}d|}}{4\pi|\mathbf{r} - \hat{z}d|}\hat{\theta}',$$

where the angle θ' is the angle between vectors $\hat{z}$ and $\mathbf{r} - \hat{z}d$ (see margin) such that

$$\hat{z}.\frac{(\mathbf{r} - \hat{z}d)}{|\mathbf{r} - \hat{z}d|} = \cos\theta'.$$

When both dipoles are "on", the total electric field phasor is

$$\tilde{\mathbf{E}}(\mathbf{r}) = \tilde{\mathbf{E}}_1(\mathbf{r}) + \tilde{\mathbf{E}}_2(\mathbf{r}) = j\eta_o I_o k\ell[\sin\theta\frac{e^{-jk|\mathbf{r}|}}{4\pi|\mathbf{r}|}\hat{\theta} + \sin\theta'\frac{e^{-jk|\mathbf{r}-\hat{z}d|}}{4\pi|\mathbf{r} - \hat{z}d|}\hat{\theta}'].$$

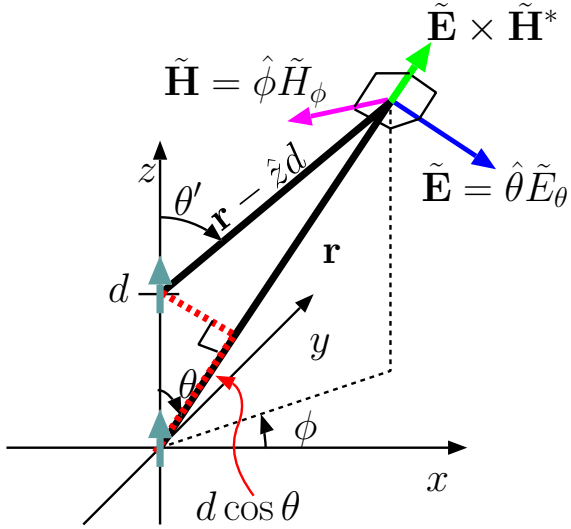
This superposition field phasor can also be expressed more compactly as

$$\tilde{\mathbf{E}}(\mathbf{r}) = j\eta_o I_o k\ell\sin\theta\frac{e^{-jk|\mathbf{r}|}}{4\pi|\mathbf{r}|}[\hat{\theta} + \frac{\sin\theta'}{\sin\theta}\frac{|\mathbf{r}|}{|\mathbf{r} - \hat{z}d|}\frac{e^{-jk|\mathbf{r}-\hat{z}d|}}{e^{-jk|\mathbf{r}|}}\hat{\theta}'],$$

from which it follows that

$$|\tilde{\mathbf{E}}(\mathbf{r})| = |\tilde{\mathbf{E}}_1(\mathbf{r})||\hat{\theta} + \frac{\sin\theta'}{\sin\theta}\frac{|\mathbf{r}|}{|\mathbf{r} - \hat{z}d|}\frac{e^{-jk|\mathbf{r}-\hat{z}d|}}{e^{-jk|\mathbf{r}|}}\hat{\theta}'|.$$

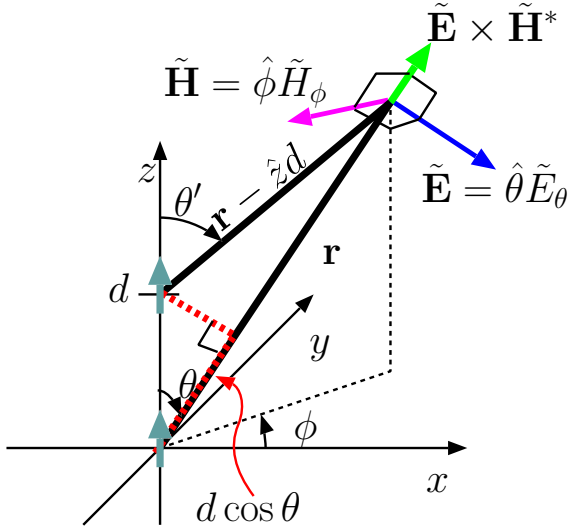
From this result it is evident that $|\tilde{\mathbf{E}}(\mathbf{r})|$ can be at most twice $|\tilde{\mathbf{E}}_1(\mathbf{r})|$ when the primed term on the right approaches $\hat{\theta}$ (constructive interference), but it can also vanish when the primed term on the right approaches $-\hat{\theta}$ (destructive interference).



Example 2: Simplify the superposition field

$$\tilde{\mathbf{E}}(\mathbf{r}) = \tilde{\mathbf{E}}_1(\mathbf{r}) + \tilde{\mathbf{E}}_2(\mathbf{r}) = j\eta_o I_o k\ell \sin\theta \frac{e^{-jk|\mathbf{r}|}}{4\pi|\mathbf{r}|}[\hat{\theta} + \frac{\sin\theta'}{\sin\theta}\frac{|\mathbf{r}|}{|\mathbf{r}-\hat{z}d|}\frac{e^{-jk|\mathbf{r}-\hat{z}d|}}{e^{-jk|\mathbf{r}|}}\hat{\theta}']$$

from Example 1 by making **paraxial approximation** in the expansion of $|\mathbf{r}-\hat{z}d|$ in relation to $|\mathbf{r}|$. From the simplified expression, find the *effective length* ℓ_{eff} of the two element antenna array of short dipoles by forcing $\tilde{\mathbf{E}}(\mathbf{r})$ to have the standard form of a $\hat{z}$-polarized radiation field.



Solution: Making **paraxial approximation** in the expansion of $|\mathbf{r} - \hat{z}d|$ in relation to $|\mathbf{r}|$ amounts to having $|\mathbf{r}| = r \gg d$ so that vectors $\mathbf{r}$ and $\mathbf{r} - \hat{z}d$ can be regarded as being parallel — under that condition we can use $\theta' = \theta$, $\hat{\theta}' = \hat{\theta}$, and

$$|\mathbf{r} - \hat{z}d| = |\mathbf{r}| - d\cos\theta.$$

Then, the total field phasor simplifies as

$$\begin{aligned}\tilde{\mathbf{E}}(\mathbf{r}) &= j\eta_o I_o k\ell \sin\theta \frac{e^{-jk|\mathbf{r}|}}{4\pi|\mathbf{r}|}\hat{\theta}[1 + \frac{|\mathbf{r}|}{|\mathbf{r}| - d\cos\theta}\frac{e^{-jk(|\mathbf{r}|-d\cos\theta)|}}{e^{-jk|\mathbf{r}|}}] \\ &\approx \tilde{\mathbf{E}}_1(\mathbf{r})[1 + e^{jkd\cos\theta}].\end{aligned}$$

Alternatively,

$$\tilde{\mathbf{E}}(\mathbf{r}) = j\eta_o I_o k \underbrace{\ell[1 + e^{jkd\cos\theta}]}_{\ell_{eff}} \sin\theta \frac{e^{-jk|\mathbf{r}|}}{4\pi|\mathbf{r}|}\hat{\theta},$$

from which we have

$$\ell_{eff} = \ell[1 + e^{jkd\cos\theta}]$$

for the effective length of the array in terms of the effective length $\ell = \frac{L}{2}$ of the short-dipole array element.

Example 3: For the two-element antenna array of short dipoles examined in Examples 1 and 2 with field phasor

$$\tilde{\mathbf{E}}(\mathbf{r}) = j\eta_o I_o k \underbrace{\ell[1+e^{jkd\cos\theta}]}_{\ell_{eff}} \sin\theta \frac{e^{-jk|\mathbf{r}|}}{4\pi|\mathbf{r}|}\hat{\theta},$$

and effective length

$$\ell_{eff} = \ell[1+e^{jkd\cos\theta}],$$

determine the gain function in terms of array directivity D.

Solution: For any linear polarized antenna we can write

$$G(\theta,\phi) = Df(\theta,\phi)$$

where function $f(\theta,\phi)$ has a maximum value of 1 and is proportional to $|\ell_{eff}\sin\theta|^2$ where θ is the angle measured from the element axis. For our two-element array described above we have

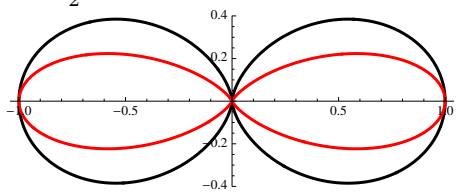
$$\begin{aligned} f(\theta,\phi) &\propto |\ell_{eff}\sin\theta|^2 \propto |1+e^{jkd\cos\theta}|^2\sin^2\theta \\ &= |e^{j\frac{1}{2}kd\cos\theta}(e^{j\frac{1}{2}kd\cos\theta}+e^{-j\frac{1}{2}kd\cos\theta})|^2\sin^2\theta \\ &= |e^{j\frac{1}{2}kd\cos\theta}|^2|e^{j\frac{1}{2}kd\cos\theta}+e^{-j\frac{1}{2}kd\cos\theta}|^2\sin^2\theta \\ &\propto \cos^2(\frac{1}{2}kd\cos\theta)\sin^2\theta. \end{aligned}$$

The function on the right maximizes at a value of 1 when $\theta = 90°$ — see its polar plot in the margin for $d=\frac{\lambda}{2}$, $d=2\lambda$, and $d=8\lambda$. Therefore, the gain of our two element array (for all possible d) is

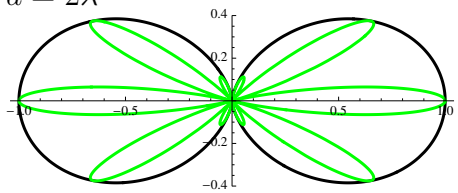
$$G(\theta,\phi) = D\cos^2(\frac{1}{2}kd\cos\theta)\sin^2\theta.$$

Polar plots of $G(\theta,\phi)/D$ for two-element array (compared to the short-dipole, shown in black):

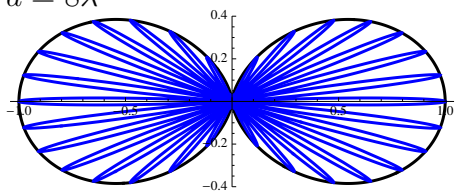
$d=\frac{\lambda}{2}$



$d=2\lambda$



$d=8\lambda$



Question: which of the above arrays has the largest D and smallest Ω_o?
Explain qualitatively.

Example 4: For the two-element antenna array examined in Examples 1-3, with the gain function

$$G(\theta,\phi) = D\cos^2(\frac{1}{2}kd\cos\theta)\sin^2\theta,$$

determine all angles θ for which $G(\theta,\phi) = 0$ if $d = 2\lambda$.

Solution: Clearly, $G(\theta,\phi) = 0$ at $\theta = 0°$ and $180°$ because of $\sin^2\theta$ factor. But also, because of factor $\cos^2(\frac{1}{2}kd\cos\theta)$, we have $G(\theta,\phi) = 0$ for all θ for which

$$\frac{1}{2}kd\cos\theta = \frac{\pi}{2}(2n+1)$$

where $n = 0, \pm1, \pm2\cdots$. This condition can be satisfied when

$$\cos\theta = \frac{\pi}{kd}(2n+1) = \frac{\pi}{\frac{2\pi}{\lambda}d}(2n+1) = \frac{\lambda/2}{d}(2n+1)$$

for all integers n such that the right hand side is bounded by -1 and +1. For $d = 2\lambda$, this condition reduces to

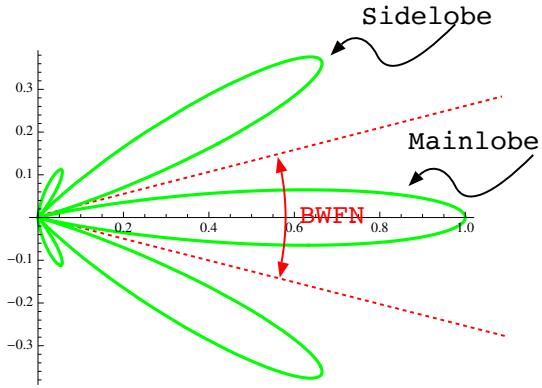
$$\cos\theta = \frac{\lambda/2}{2\lambda}(2n+1) = \frac{2n+1}{4} = \{-\frac{3}{4}, -\frac{1}{4}, \frac{1}{4}, \frac{3}{4}\}.$$

So, we have

$$G(\theta,\phi) = 0 \text{ for } \theta{=}0°,\ \cos^{-1}\frac{3}{4} = 41.41°,\ \cos^{-1}\frac{1}{4} = 75.52°,$$

$$\cos^{-1}\frac{-1}{4} = 104.78°,\ \cos^{-1}\frac{-3}{4} = 138.6°,\ 180°.$$

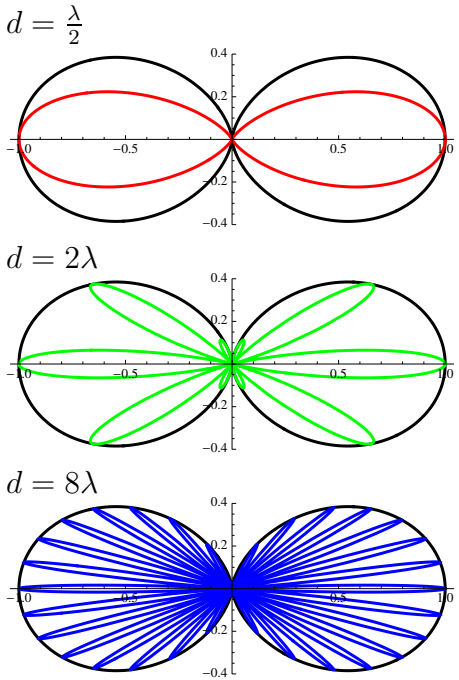
$d = 2\lambda$



BWFN=
Beam-width between first nulls

- The patterns shown for the two-element array in the margin illustrate that larger the element separation d, narrower the angular width of the mainlobe.

  - However, the number of sidelobes also increase with d, so there is no substantial directivity increase *with* distance d because of that (because of power diverted into the relatively large intensity sidelobes).
  - The remedy is to have *multiple-element arrays* analyzed next.



- In the two-element array the distant field, in *paraxial approximation*, was found to be

$$\tilde{\mathbf{E}}(\mathbf{r}) = \tilde{\mathbf{E}}_1(\mathbf{r}) + \tilde{\mathbf{E}}_2(\mathbf{r}) = \tilde{\mathbf{E}}_1(\mathbf{r})[1 + e^{jkd\cos\theta}]$$

  after using

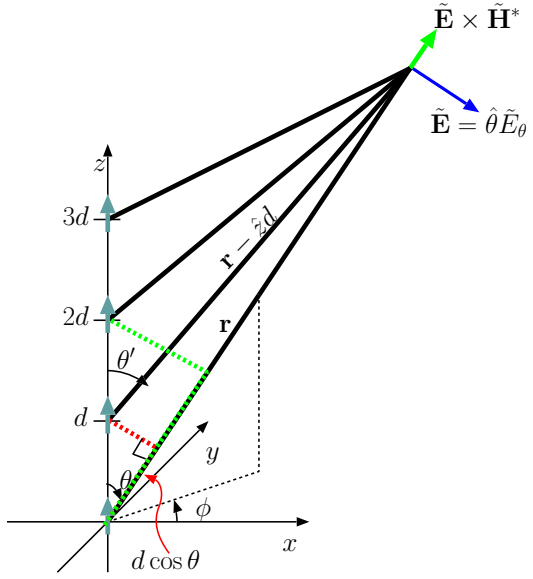
$$\tilde{\mathbf{E}}_2(\mathbf{r}) \approx \tilde{\mathbf{E}}_1(\mathbf{r})e^{jkd\cos\theta}.$$

- For a 3-element array with element locations $(0,0,0)$, $(0,0,d)$, and $(0,0,2d)$ this result can be extended as

$$\tilde{\mathbf{E}}(\mathbf{r}) = \tilde{\mathbf{E}}_1(\mathbf{r}) + \tilde{\mathbf{E}}_2(\mathbf{r}) + \tilde{\mathbf{E}}_3(\mathbf{r}) = \tilde{\mathbf{E}}_1(\mathbf{r})[1 + e^{jkd\cos\theta} + e^{j2kd\cos\theta}],$$

  and, for an N-element array, with elements at $(0,0,nd)$ for n in the interval $0 \cdots N-1$, we can write

$$\tilde{\mathbf{E}}(\mathbf{r}) = \tilde{\mathbf{E}}_1(\mathbf{r})[1 + e^{jkd\cos\theta} + e^{j2kd\cos\theta} + \cdots + e^{j(N-1)kd\cos\theta}].$$



– These superposed field expressions in the antenna far-field imply an effective length of

$$\ell_{eff} = \ell \sum_{n=0}^{N-1} (e^{jkd\cos\theta})^n.$$

- The sum on the right is called **array factor** (A.F.) and we see that the **effective length of the array antenna** is the product of the effective length ℓ of an array element *and* A.F..

- We can write the gain of the N-element array (once again as)

$$G(\theta, \phi) = Df(\theta, \phi),$$

where

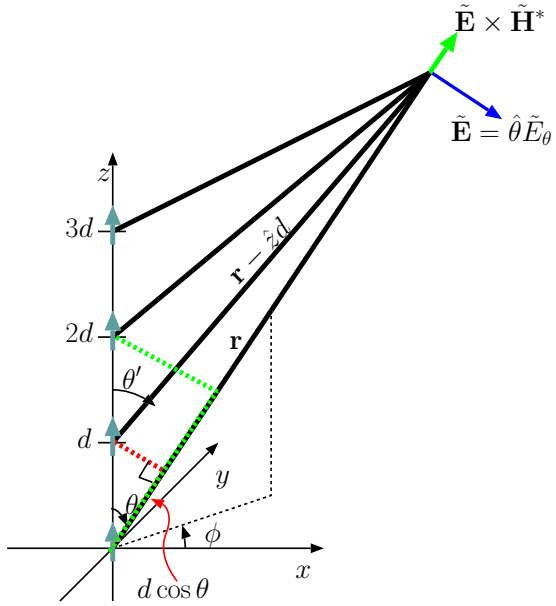
$$f(\theta, \phi) \propto |\ell_{eff} \sin\theta|^2 = |\ell|^2 |\sum_{n=0}^{N-1} (e^{jkd\cos\theta})^n|^2 \sin^2\theta$$

and has a max value of 1. The A.F. maximizes at a value of N at $\theta = 90°$ and thus it works out that

$$G(\theta, \phi) = D\sin^2\theta |\frac{1}{N} \sum_{n=0}^{N-1} (e^{jkd\cos\theta})^n|^2.$$

– To simplify this gain formula we note that

$$s \equiv 1 + w + w^2 + \cdots + w^{N-1} \quad \Rightarrow \quad sw = w + w^2 + \cdots + w^{N-1} + w^N,$$



The same interference principle governs N-element arrays: at locations where field phasors from individual elements have the same angle, constructive interference takes place, and the radiation field of the array is strong. At other locations where field phasors from individual elements cancel one another, the field of the whole array is weak.

and, therefore,

$$s(w-1) = w^N - 1 \quad \Rightarrow \quad s = \frac{w^N - 1}{w - 1}.$$

- Applying this summation formula for s with $w = e^{jkd\cos\theta}$, we obtain

$$\text{A.F.} = \sum_{n=0}^{N-1} (e^{jkd\cos\theta})^n = \frac{e^{jNkd\cos\theta} - 1}{e^{jkd\cos\theta} - 1}.$$

Now,

$$\begin{aligned} |\text{A.F.}| &= |\sum_{n=0}^{N-1} (e^{jkd\cos\theta})^n| = \frac{|e^{jNkd\cos\theta} - 1|}{|e^{jkd\cos\theta} - 1|} \\ &= \frac{|e^{j\frac{N}{2}kd\cos\theta}(e^{j\frac{N}{2}kd\cos\theta} - e^{-j\frac{N}{2}kd\cos\theta})|}{|e^{j\frac{1}{2}kd\cos\theta}(e^{j\frac{1}{2}kd\cos\theta} - e^{-j\frac{1}{2}kd\cos\theta})|} = \frac{|\sin(\frac{N}{2}kd\cos\theta)|}{|\sin(\frac{1}{2}kd\cos\theta)|}. \end{aligned}$$

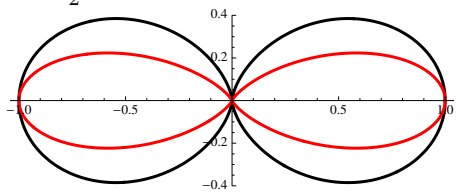
The upshot is,

$$G(\theta,\phi) = D\sin^2\theta \frac{\sin^2(\frac{N}{2}kd\cos\theta)}{N^2\sin^2(\frac{1}{2}kd\cos\theta)}$$

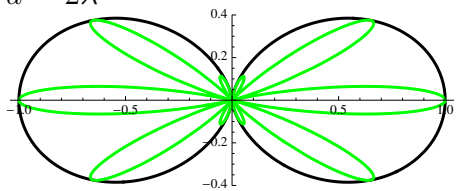
for an N-element array with a physical size Nd.

– Plots of $G(\theta,\phi)/D$ for $d = \frac{\lambda}{2}$ and $N = 2, 4, 16$ are shown in the margin. Note the reduced sidelobe levels (you can barely see them) and how larger N results in larger directivity D.

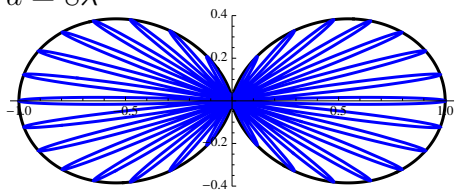
$d = \frac{\lambda}{2}$



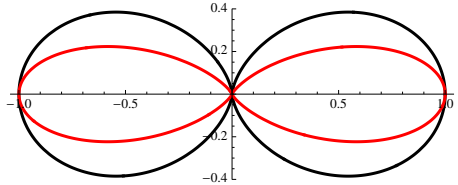
$d = 2\lambda$



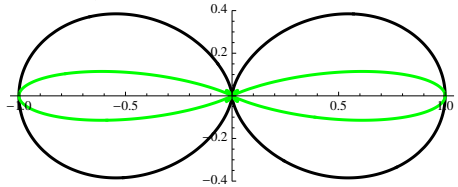
$d = 8\lambda$



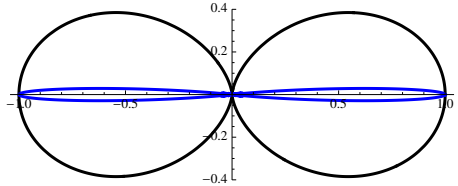
$N = 2,\ Nd = \lambda$



$N = 4,\ Nd = 2\lambda$



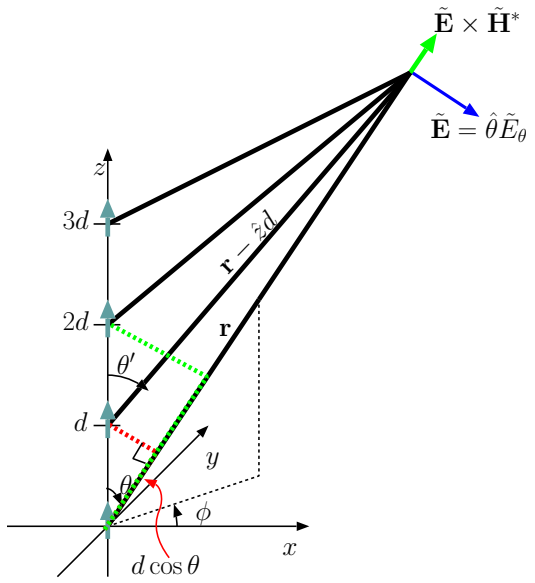
$N = 16,\ Nd = 8\lambda$



12 Interference, antenna arrays — cont'd.

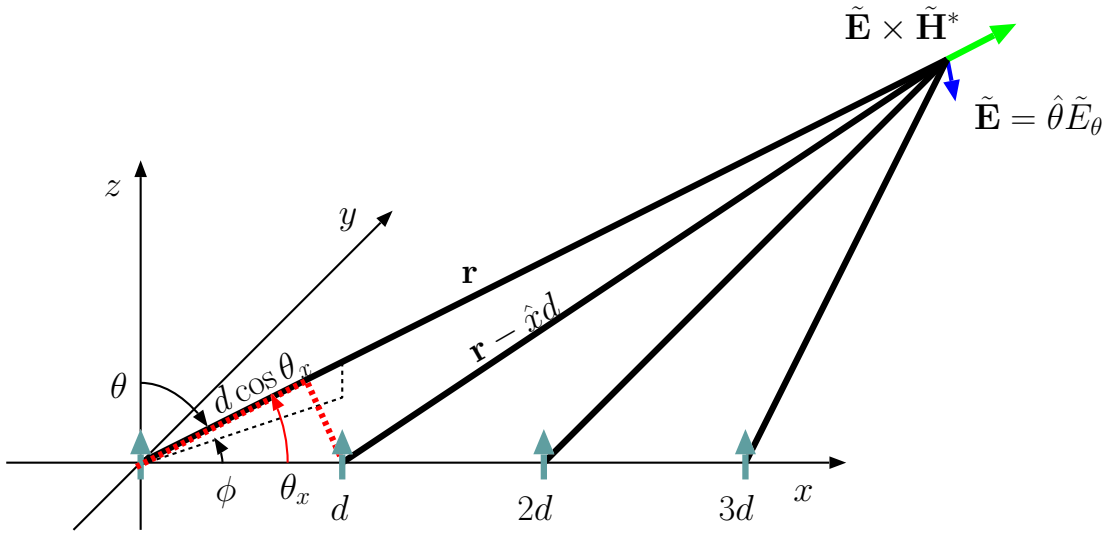
We continue our study of interference effects and antenna arrays.

- Beam patterns of N-element antenna arrays examined last lecture were isotropic in ϕ direction — the main effect of increasing the array size Nd appeared to be *narrowing* the mainlobe of the pattern in θ direction.
- These so-called **broadside arrays** — meaning that they mainly radiate in the "broadside direction" of the "array axis" — are good for *broadcasting purposes* at relatively high frequencies $\frac{\omega}{2\pi}$ in the FM band ($\sim$100 MHz),
  - where array sizes Nd, in excess of many λ's, become practicable (as opposed to in AM band where $\frac{\omega}{2\pi}$ $\sim$1 MHz and $\lambda \sim 300$ m).
- They may also be used as "elements" of arrays built along x- or y-axis directions which we will consider next.
  - In that case it will be possible to produce antenna beam patterns *anisotropic* in the azimuth plane (in ϕ direction).
  - We will also consider *phasing* the element input currents so that the mainlobe of the beam can be steered into desired directions in the azimuth plane.



New vocabulary:

- **Broadside arrays**
- **Array axis**
- **Broadside direction**



- Consider an array of elements polarized in $\hat{z}$-direction positioned along the x-axis as shown above. Our initial analysis of this array will assume equal input currents I_o for all the elements. Let

  $\tilde{\mathbf{E}}_0(\mathbf{r}) \propto \dfrac{e^{-jk|\mathbf{r}|}}{|\mathbf{r}|}$ denote the field at the observation point $\mathbf{r}$ due to the element at the origin.

  - Then, using the paraxial approximation, the field phasor at a distant observation point due to the next element at $(d, 0, 0)$ can be expressed in terms of $\tilde{\mathbf{E}}_0(\mathbf{r})$ as

    $$\tilde{\mathbf{E}}_1(\mathbf{r}) \approx \tilde{\mathbf{E}}_0(\mathbf{r}) e^{jkd\cos\theta_x}$$

    where θ_x is the angle between vectors $\mathbf{r}$ and $\hat{x}$, i.e.,

    $\cos\theta_x = \hat{r} \cdot \hat{x} = (\sin\theta\cos\phi\hat{x} + \sin\theta\sin\phi\hat{y} + \cos\theta\hat{z}) \cdot \hat{x} = \sin\theta\cos\phi,$

    known as a **direction cosine**.

– Likewise,

$$\tilde{\mathbf{E}}_2(\mathbf{r}) \approx \tilde{\mathbf{E}}_0(\mathbf{r}) e^{j2kd\cos\theta_x}, \text{ etc., so that,}$$

– For an N-element array,

$$\tilde{\mathbf{E}}(\mathbf{r}) = \tilde{\mathbf{E}}_0(\mathbf{r})[1 + e^{jkd\cos\theta_x} + e^{j2kd\cos\theta_x} + \cdots + e^{j(N-1)kd\cos\theta_x}].$$

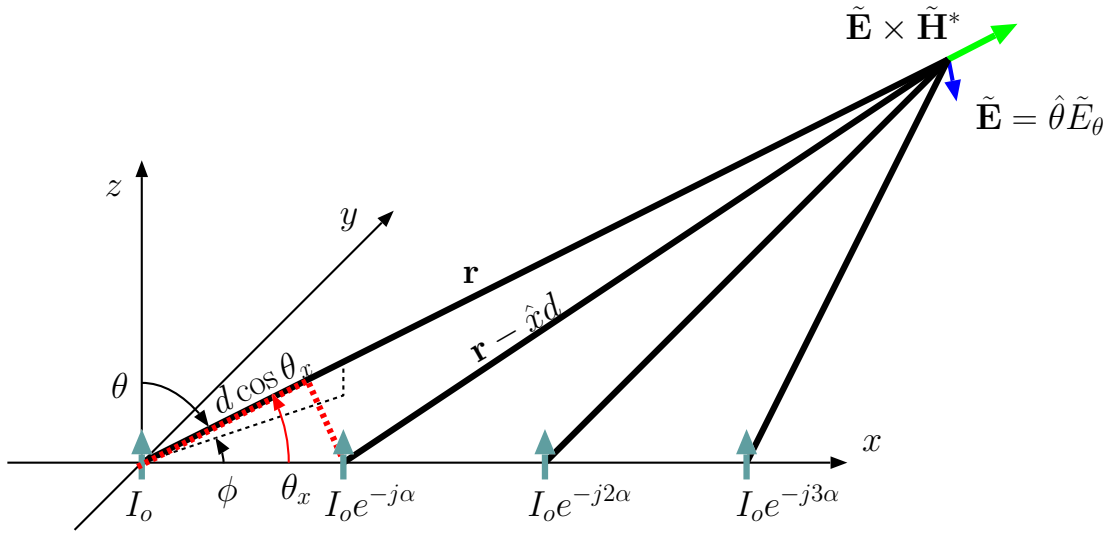
- The field expression above is identical in essence with the field expression for the N-element array examined in the last lecture except for the replacement of $\cos\theta$ by $\cos\theta_x$. Therefore, assuming that $\tilde{\mathbf{E}}_0(\mathbf{r})$ is due a short dipole (so that $\ell = \frac{L}{2}$ is independent of direction), we obtain the A.F. for our new array by exchanging $\cos\theta$ by $\cos\theta_x$ in the A.F. obtained in the last lecture — by that procedure we arrive at

$$\begin{aligned} G(\theta,\phi) &= K\sin^2\theta \frac{\sin^2(\frac{N}{2}kd\cos\theta_x)}{N^2\sin^2(\frac{1}{2}kd\cos\theta_x)} \\ &= K\sin^2\theta \frac{\sin^2(\frac{N}{2}kd\sin\theta\cos\phi)}{N^2\sin^2(\frac{1}{2}kd\sin\theta\cos\phi)}. \end{aligned}$$

We have *at last* obtained a gain function that does actually depend on both θ and ϕ. The scaling constant K above is to be determined by requiring $\int G d\Omega = 4\pi$.

– Even more complicated gain expressions would be obtained if the array elements were themselves arrays (like those examined last lecture) having angle dependent ℓ's! (see HW problems).

Phased array:



- Next let's examine what happens when the element input currents are not identical, but follow a progressive phase pattern with

$$I_n = I_o e^{-jn\alpha} \text{ for } n\text{-th element located at } (nd, 0, 0)$$

where α is a phasing increment specified in radians.

In that case — since $\tilde{\mathbf{E}}_n(\mathbf{r}) \propto I_n$ — we would have a **phased array** with element field phasors

$$\tilde{\mathbf{E}}_1(\mathbf{r}) \approx \tilde{\mathbf{E}}_0(\mathbf{r}) e^{j(kd\cos\theta_x - \alpha)}, \quad \tilde{\mathbf{E}}_2(\mathbf{r}) \approx \tilde{\mathbf{E}}_0(\mathbf{r}) e^{j2(kd\cos\theta_x - \alpha)}, \quad \text{etc.},$$

and an array gain function (again, assuming short-dipole elements)

$$G(\theta, \phi) = K \sin^2\theta \frac{\sin^2(\frac{N}{2}(kd\sin\theta\cos\phi - \alpha))}{N^2 \sin^2(\frac{1}{2}(kd\sin\theta\cos\phi - \alpha))}$$

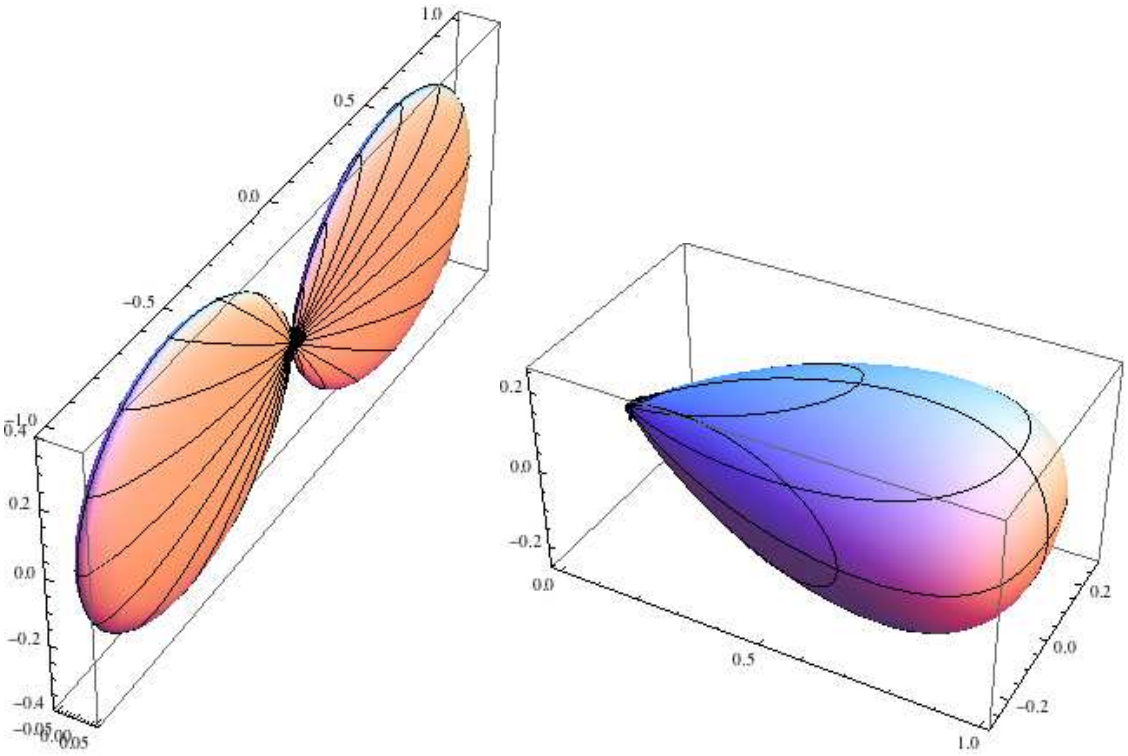
where, once again, constant K is to be determined by requiring $\int G d\Omega = 4\pi$ — demonstarted in Mathematica notebook demo shown in class.

- On $\theta = 90°$ plane we have

$$G(90°, \phi) = K \frac{\sin^2(\frac{N}{2}(kd\cos\phi - \alpha))}{N^2 \sin^2(\frac{1}{2}(kd\cos\phi - \alpha))}$$

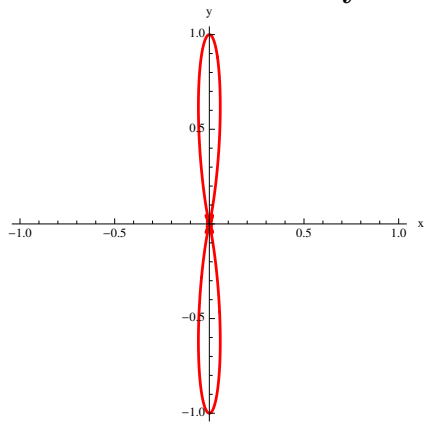
  which leads to azimuth plane patterns (G/K) shown in the margin shown for $d = \frac{\lambda}{4}$, $N = 16$, and $\alpha = 0$, $\frac{\pi}{4}$, and $\frac{\pi}{2}$ radians.

- Also, 3D plots of $G(\theta, \phi)/K$ for $d = \frac{\lambda}{4}$, $N = 16$, and $\alpha = 0$ and $\frac{\pi}{2}$ radians are shown below:

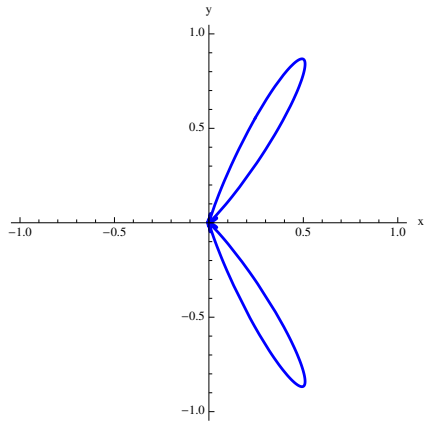


Question: The broadside pattern shown on the left is clearly a "fan beam" — how would you make the "fans" narrower (in order to increase D) in θ direction?

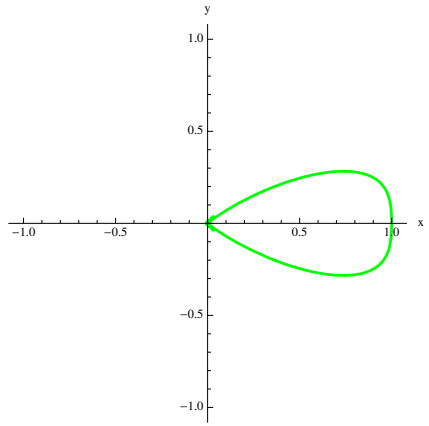
$\alpha = 0$**: Broadside array**



$\alpha = \frac{\pi}{4}$**: Steered beam:**



$\alpha = \frac{\pi}{2}$**: End-fire array**

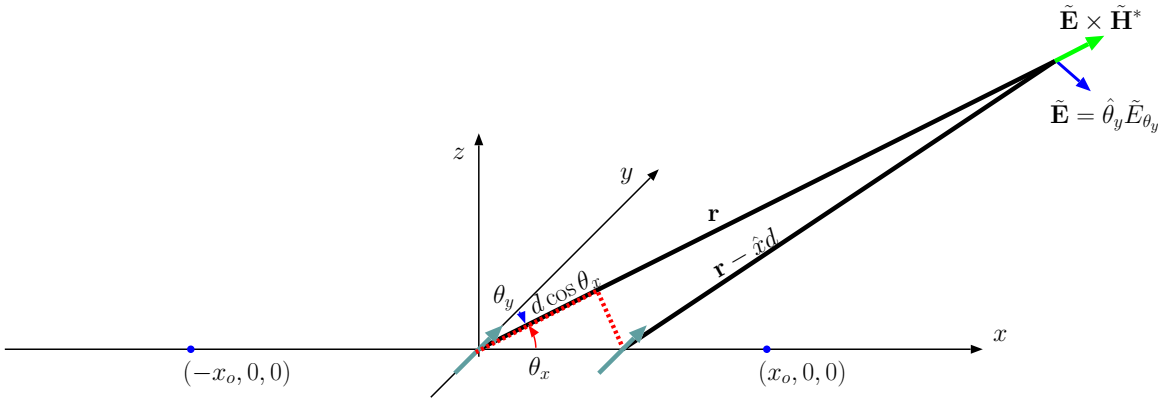


- Our discussion so far have focused on 1D arrays.
  - One class of 2D arrays consist of 1D arrays having other 1D arrays as their elements, in which case their gain functions can be formulated after multiplying the array factors of two 1D arrays and the effective length of the smallest element of the arrays.
  - We have shown examples illustrating how antenna beams with relatively small solid angles Ω_o and directivities $D = \frac{4\pi}{\Omega_o}$ can be generated.
  - We have shown how phasing can be used to steer the antenna beam patterns.

- Interference effects which are fundamental to antenna array design are mainly sensitive to
  1. antenna locations, and
  2. phases of antenna input currents.

  The next example examines these parameters in more detail.

Example 1: We have two identical $\hat{y}$-polarized short dipoles. We want to "place" them and "phase" their input currents in such a way that no power is radiated in $+x$ direction and there is a gain maximum in $-x$ direction. Determine the required positions of the dipoles and the relative phases of their input currents.



Solution: Let's place the two $\hat{y}$-polarized short dipoles at $(0,0,0)$ and $(d,0,0)$ as shown above and drive them with input currents I_1 and I_2, respectively. For $I_2 = I_1 e^{-j\alpha}$, the field phasors of the dipoles at an observation point $(x_o, 0, 0)$, $x_o \gg d$, will vary as

$$\tilde{\mathbf{E}}_1 \propto e^{-jkx_o} \quad \text{and} \quad \tilde{\mathbf{E}}_2 \propto e^{-j\alpha} e^{-jk(x_o - d)} = e^{-jkx_o} e^{j(kd-\alpha)}$$

having identical proportionality constants. Likewise, at an observation point $(-x_o, 0, 0)$, $x_o \gg d$, we will have field phasors

$$\tilde{\mathbf{E}}_1 \propto e^{-jkx_o} \quad \text{and} \quad \tilde{\mathbf{E}}_2 \propto e^{-j\alpha} e^{-jk(x_o + d)} = e^{-jkx_o} e^{-j(kd+\alpha)}.$$

Now, in order to have *destructive* interference between $\tilde{\mathbf{E}}_1$ and $\tilde{\mathbf{E}}_2$ at $(x_o, 0, 0)$ we need to have

$$e^{j(kd-\alpha)} = -1 \quad \Rightarrow \quad kd - \alpha = \pi.$$

Also to have *constructive* interference between $\tilde{\mathbf{E}}_1$ and $\tilde{\mathbf{E}}_2$ at $(-x_o, 0, 0)$ we need to have

$$e^{-j(kd+\alpha)} = 1 \quad \Rightarrow \quad kd + \alpha = 0.$$

Adding and subtracting these equations we find that

$$kd = \frac{\pi}{2} \quad \text{and} \quad \alpha = -\frac{\pi}{2}.$$

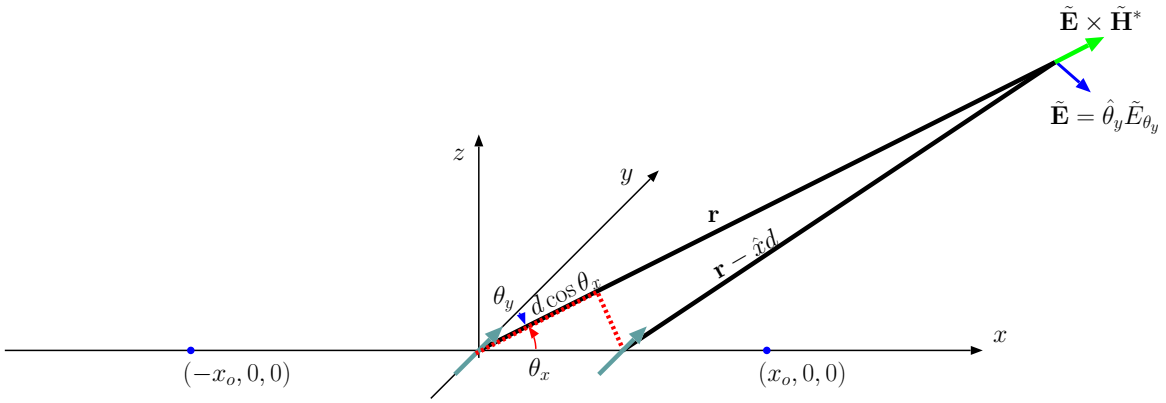
Thus

$$d = \frac{\pi}{2k} = \frac{\pi}{2\frac{2\pi}{\lambda}} = \frac{\lambda}{4}.$$

This result makes sense because, with I_1 lagging I_2 by 90° of phase, and $\tilde{\mathbf{E}}_1$ traveling an extra $\frac{\lambda}{4}$ compared to $\tilde{\mathbf{E}}_2$ to lose an additional phase of 90°, $\tilde{\mathbf{E}}_1$ ends up being 180° out of phase with $\tilde{\mathbf{E}}_1$ at $(x_o, 0, 0)$, which is the condition for destructive interference.

Conversely, with I_2 leading I_1 by 90° of phase, but $\tilde{\mathbf{E}}_2$ traveling an extra $\frac{\lambda}{4}$ compared to $\tilde{\mathbf{E}}_1$ to lose that phase lead of 90°, $\tilde{\mathbf{E}}_2$ ends up being *in phase* with $\tilde{\mathbf{E}}_1$ at $(-x_o, 0, 0)$, which is the condition for constructive interference.

Example 2*: (Difficult example) Obtain the gain function $G(\theta, \phi)$ for the 2-element array examined in Example 2.



Solution: With $I_2 = I_1 e^{-j\alpha}$, $\alpha = -\frac{\pi}{2}$ and $d = \frac{\lambda}{4}$, we have in the antenna far-field (i.e.,

the region where paraxial approximation justified)

$$\tilde{\mathbf{E}}(\mathbf{r}) = \tilde{\mathbf{E}}_1(\mathbf{r})(1 + e^{-j\alpha}e^{jkd\cos\theta_x}) = \tilde{\mathbf{E}}_1(\mathbf{r})(1 + e^{j\frac{\pi}{2}(\cos\theta_x+1)})$$

where

$$\tilde{\mathbf{E}}_1(\mathbf{r}) = j\eta_o I_1 k\ell \sin\theta_y \frac{e^{-jkr}}{4\pi r}\hat{\theta}_y.$$

Consequently,

$$G(\theta,\phi) = Df(\theta,\phi)$$

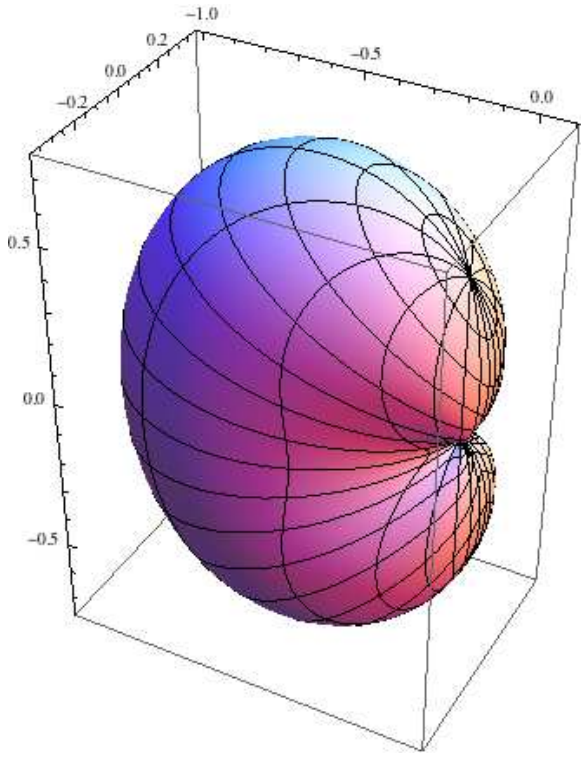
such that

$$f(\theta,\phi) \propto |\ell|^2 |\sin\theta_y|^2 |1 + e^{j\frac{\pi}{2}(\cos\theta_x+1)}|^2$$

and is normalized to a peak value of 1. This is compatible with

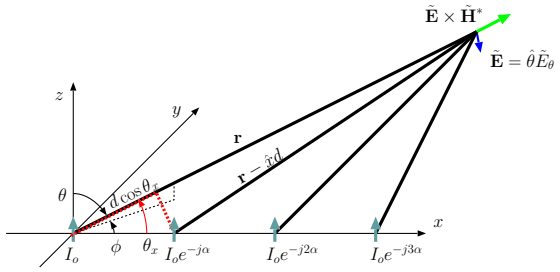
$$\begin{aligned} G(\theta,\phi) &= D\sin^2\theta_y \frac{|1 + e^{j\frac{\pi}{2}(\cos\theta_x+1)}|^2}{4} \\ &= D\sin^2\theta_y \frac{2 + e^{j\frac{\pi}{2}(\cos\theta_x+1)} + e^{-j\frac{\pi}{2}(\cos\theta_x+1)}}{4} \\ &= D(1-\cos^2\theta_y)\frac{1+\cos(\frac{\pi}{2}(\cos\theta_x+1))}{2} \\ &= D(1-\sin^2\theta\sin^2\phi)\frac{1+\cos(\frac{\pi}{2}(\sin\theta\cos\phi+1))}{2}. \end{aligned}$$

A 3D plot of $G(\theta,\phi)/D$ is shown in the margin.



13 Arrays and feed networks

Performance of antenna arrays depends on our ability to feed the array elements with input currents having accurate phase relationships. This can be accomplished by using appropriately designed "feed networks" consisting of transmission line (TL) segments. We continue our study of antenna arrays with examples illustrating feed network design issues.



Example 1: Consider a 4-element phased array with $\hat{z}$-polarized short dipole elements, progressive phasing with increments α, and element-to-element spacings $d = \frac{\lambda}{2}$ along the x-axis. It is desired that the array has a gain maximum on $\theta = 90^\circ$ plane along $\phi = 45^\circ$ directions. Determine α such that $I_1 = I_0 e^{-j\alpha}$ and suggest a TL feed network that can be used to distribute the required input currents of the array elements.

Solution: Let $\tilde{\mathbf{E}}_0(\mathbf{r})$ denote the far-field phasor due to the array element at the origin. The phasor due to the element at $(d, 0, 0)$ can then be expressed (in the far-field) as

$$\tilde{\mathbf{E}}_1(\mathbf{r}) = \tilde{\mathbf{E}}_0(\mathbf{r}) e^{-j\alpha} e^{jkd\cos\theta_x} = \tilde{\mathbf{E}}_0(\mathbf{r}) e^{-j\alpha} e^{jkd\sin\theta\cos\phi}.$$

Since we are interested in $\theta = 90^\circ$ case, we can simplify this as

$$\tilde{\mathbf{E}}_1(\mathbf{r}) = \tilde{\mathbf{E}}_0(\mathbf{r}) e^{j(kd\cos\phi - \alpha)}.$$

Since we want constructive interference in $\phi = 45^\circ$ direction, we will demand that

$$e^{j(kd\cos\phi-\alpha)} = 1 \quad \Rightarrow \quad \alpha = kd\cos\phi \ \text{ for } \ \phi = 45^\circ,$$

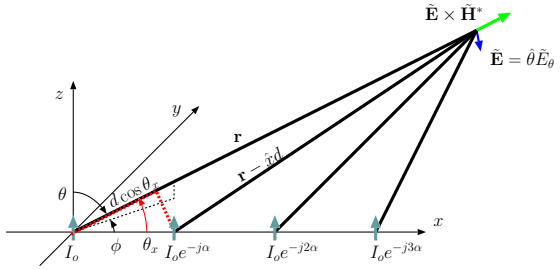
i.e,

$$\alpha = \frac{2\pi}{\lambda}\frac{\lambda}{2}\cos 45^\circ = \frac{\pi}{\sqrt{2}}\,\text{rad}.$$

Thus, the required current inputs of the array elements are

$$I_n = I_0 e^{-jn\frac{\pi}{\sqrt{2}}}, \quad n = 0, 1, 2, 3.$$

With this phasing, radiation coming from all four elements will interfere constructively along the $\phi = 45^\circ$ direction.



Designing the feed network: The required phase delays

$$\alpha = \frac{2\pi}{\lambda}\frac{\lambda}{2}\cos 45^\circ = \frac{\pi}{\sqrt{2}}\,\text{rad}$$

to be applied progressively are *in effect* to compensate for the fact that propagation distance from **element 1** to the observation point (along the $\phi = 45^\circ$ line) is shorter than the distance from **element 0** by an amount

$$\Delta = d\cos\phi = \frac{\lambda}{2}\cos 45^\circ.$$

Let's make the current signal arriving at the input terminals of **element 1** travel on a TL (with propagation speed $v = c$) an *extra distance* Δ compared to the current going to **element 0** (coming from the same source) — that procedure, repeated progressively for all the elements, will produce the required current inputs

$$I_n = I_0 e^{-jn\frac{\pi}{\sqrt{2}}}, \quad n = 0, 1, 2, 3,$$

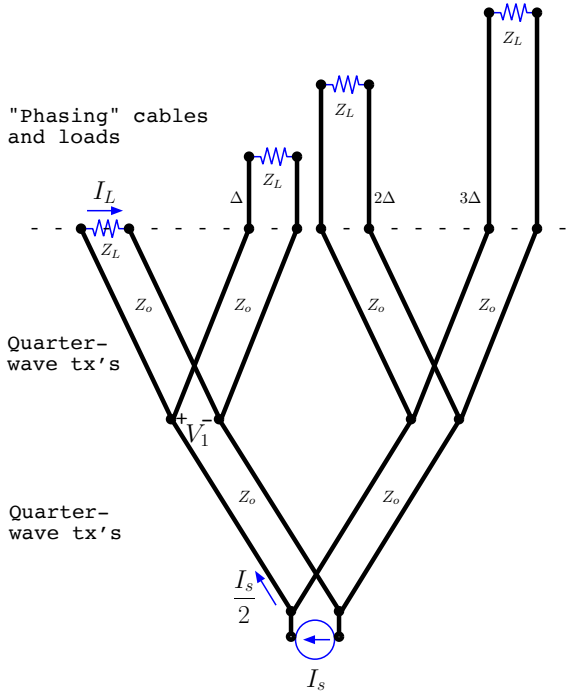
derived from a common current source, provided reflected waves from the elements can be avoided.

To avoid reflections from the antenna elements it is necessary to match the antenna impedance Z_L to the characteristic impedance of the TL (e.g., single-stub tuning).

Assume that all the elements have been identically matched to a TL with a characteristic impedance Z_o. Then we can connect the elements to a common current source I_s via a corporate-ladder network shown in the margin and have

$$I_n = \underbrace{-\frac{I_s}{2}}_{I_0} e^{-jn\frac{\pi}{\sqrt{2}}}, \quad n = 0, 1, 2, 3.$$

The verification of this formula would require the use of *quarter-wave transformation formulae* for impedance and terminal voltage and currents reviewed next.



- As shown in the margin a quarter-wave transformer with a characteristic impedance Z_o transforms a load impedance Z_L into an input impedance

$$Z_{in} = \frac{Z_o^2}{Z_L}.$$

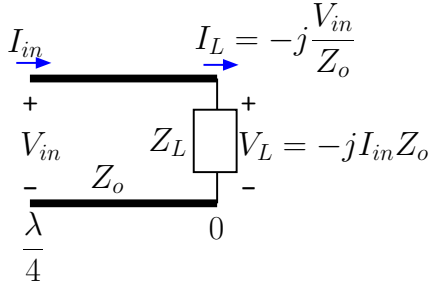
Also, an input voltage V_{in} is transformed to a load current

$$I_L = -j\frac{V_{in}}{Z_o}$$

independent of load impedance Z_L, while an input current I_{in} is transformed into a load voltage

$$V_L = -jZ_oI_{in}.$$

Quarter-wave transformer:



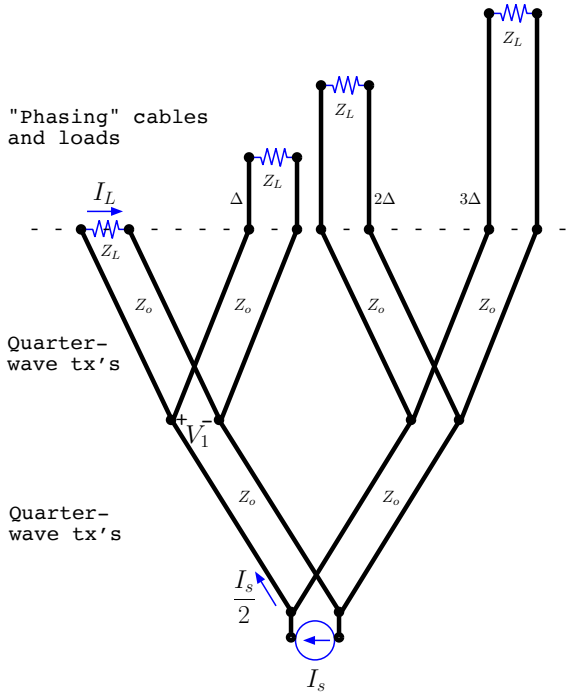
$Z_{in}Z_L = Z_o^2$

- Assuming $Z_L = Z_o$ and applying the quarter-wave transformer formulae repeatedly we note that

1. Junction voltage
$$V_1 = -j\frac{I_s}{2}Z_o;$$

2. Element 1 current
$$I_L = -j\frac{V_1}{Z_o} = -j\frac{(-j\frac{I_s}{2}Z_o)}{Z_o} = -\frac{I_s}{2}.$$

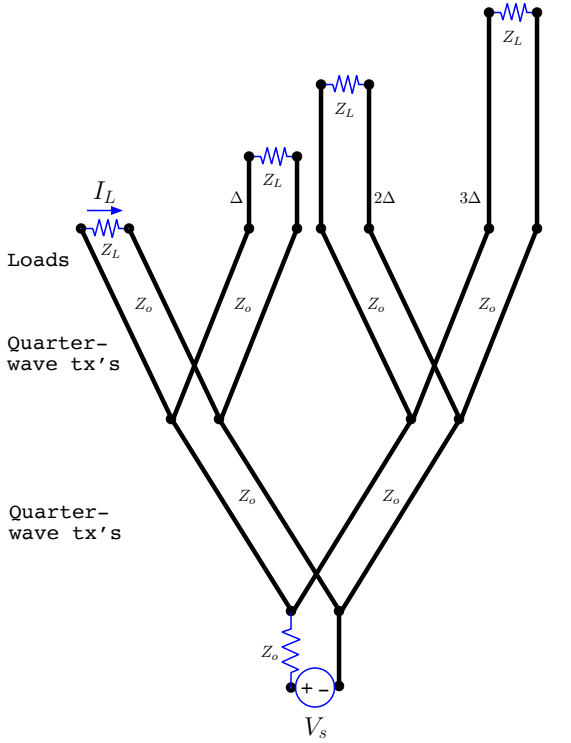
3. Element n current is then
$$I_n = I_L e^{-jnk\Delta} = I_L e^{-jn\frac{\pi}{\sqrt{2}}}$$
as required.



Question: Why did we have to assume $Z_L = Z_o$ above?

Answer: Because otherwise the impedance of the two branches seen by I_s may have been different, and we would not necessarily have equal currents $\frac{I_s}{2}$ flowing into the two branches of the circuit.

Exercise: In the diagram on the margin we have a voltage source V_s in series with a source impedance Z_o feeding an array feed network. Determine I_n using quarter-wave transformation rules assuming $Z_L = Z_o$.



Spatial Fourier transforms of current distributions:

- Consider an array of identical dipoles positioned along the x-axis at locations $x_n = nd$ having input currents I_n, with n in the interval $0, 1, \cdots N-1$.

- The far-field electric field of the array (in paraxial approximation) is then

$$\begin{aligned}\tilde{\mathbf{E}}(\mathbf{r}) &= \tilde{\mathbf{E}}_0(\mathbf{r})[1 + \frac{I_1}{I_0}e^{jkd\cos\theta_x} + \frac{I_2}{I_0}e^{j2kd\cos\theta_x} + \cdots + \frac{I_{N-1}}{I_0}e^{j(N-1)kd\cos\theta_x}] \\ &= \tilde{\mathbf{E}}_0(\mathbf{r})\sum_{n=0}^{N-1}\frac{I_n}{I_0}e^{jnkd\cos\theta_x}\end{aligned}$$

  so that the array factor is

$$\text{A.F.} = \sum_{n=0}^{N-1}\frac{I_n}{I_0}e^{jk\cos\theta_x nd},$$

  a discrete Fourier transform of the sequence $\frac{I_n}{I_0}$ representing the "illumination pattern" of the array into a spatial-frequency domain of $k_x \equiv k\cos\theta_x$.

- Compare the A.F. with the effective length of a single array element (from Lecture 8)

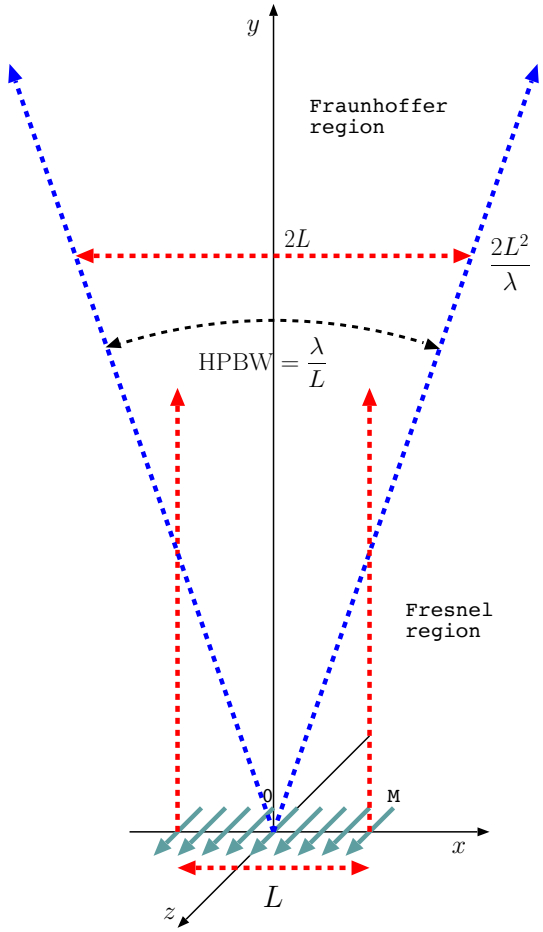
$$\ell = \int \frac{\tilde{I}(z)}{I_o}e^{jk\cos\theta z}dz,$$

  which is also a spatial Fourier transform into the domain $k_z \equiv k\cos\theta$.

- Therefore the system gain that consists of the products of the A.F. and element ℓ is in effect a spatial Fourier transform of the "current distributions" or so-called "illumination pattern" on antenna surfaces of the system.

  - This link will be examined more closely in antenna and imaging courses starting with ECE 454.

14 Interference zones, plane waves



- Let's examine the radiation field of a 1D array of $N = 2M + 1$ identical elements located at $(nd, 0, 0)$, with n in the interval $-M, \cdots -1, 0, 1, \cdots M$ having *spherical* wave field phasors

$$\tilde{\mathbf{E}}_n(\mathbf{r}) = j\eta_o I_n k\ell \sin\theta_n \frac{e^{-jk|\mathbf{r}-\hat{x}nd|}}{4\pi|\mathbf{r}-\hat{x}nd|}\hat{\theta}_n$$

where

$$\cos\theta_n = \hat{z}\cdot\frac{\mathbf{r}-\hat{x}nd}{|\mathbf{r}-\hat{x}nd|}.$$

- The total field phasor

$$\tilde{\mathbf{E}}(\mathbf{r}) = \sum_{n=-M}^{M} \tilde{\mathbf{E}}_n(\mathbf{r})$$

of the array will have different types of spatial variations in different *interference zones* or *regions*:

1. The region

$$|\mathbf{r}| \lesssim \frac{2L^2}{\lambda}, \quad \text{where} \quad L = 2Md$$

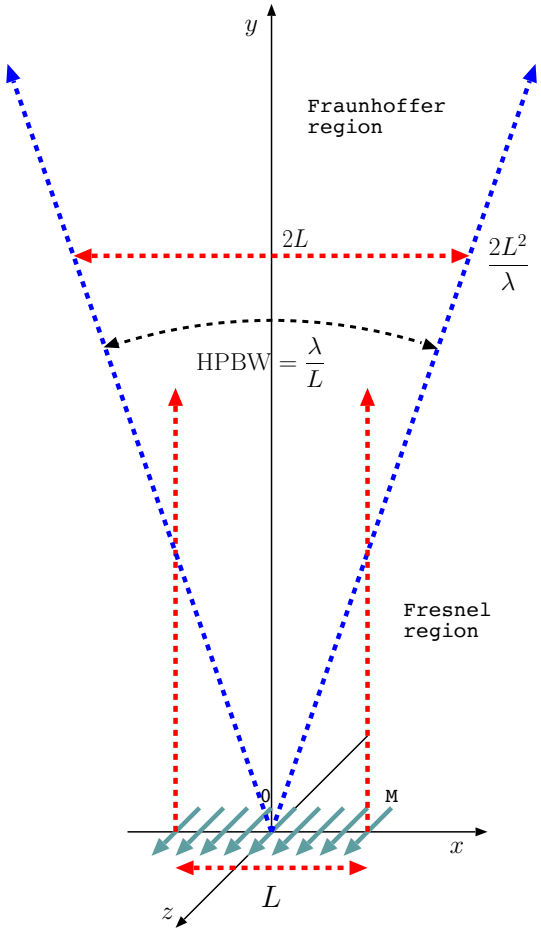
is the physical length of the array, is known as **Fresnel region** or the **near-field** radiation zone — in this zone paraxial approximation cannot be used and the radiation field is highly structured having a prominent magnitude directly above the array (i.e., for $-Md \lesssim x \lesssim Md$).

2. The region

$$|\mathbf{r}| \gtrsim \frac{2L^2}{\lambda}$$

is known as **Fraunhoffer region** or the **far field** — this is the zone in which paraxial approximation works well, and spherical waves arriving from individual array elements merge to become a single spherical wave of a higher directivity.

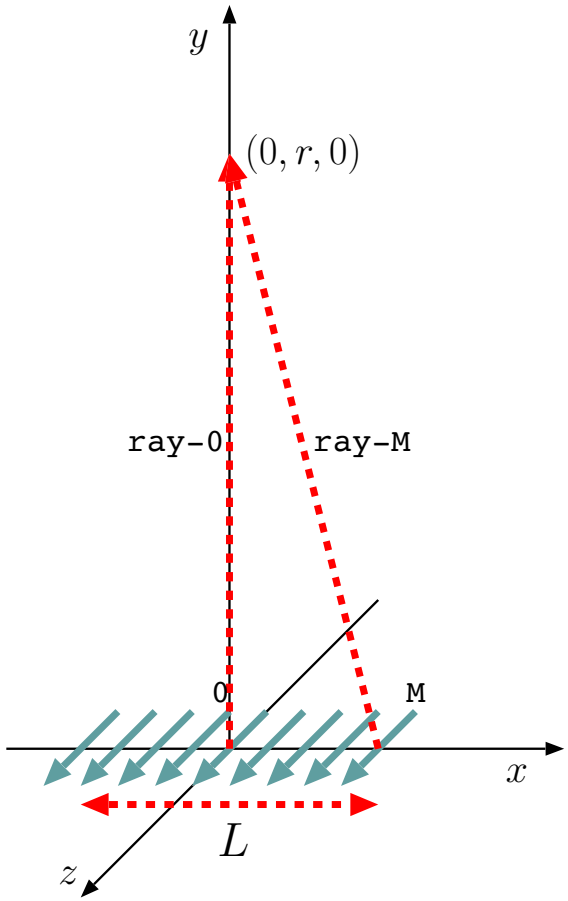
The concept of *antenna beam* applies only in the Fraunhoffer region. A beam with a fixed angular width emerges out of the Fresnel region as Fraunhoffer region is approached, as shown in the cartoon in the margin (in which an "unphased" broadside array has been assumed in sketching the far-field beam).



- In addition, it should be noted that
  - the region $|\mathbf{r}| \lesssim$ few λ will include strong storage fields, whereas
  - for $|\mathbf{r}| \gg \frac{2L^2}{\lambda}$, deep in Fraunhoffer region, spherical waves will "locally" look like plane waves.

We will next examine the transition between Fresnel and Fraunhoffer regions and then examine how spherical waves can be treated as plane waves over limited regions of space in the far-field.

- Consider the "phase-delay" of signals arriving from individual elements of a broadside array on the x-axis to a location $(0, r, 0)$ on the y-axis as shown in the margin.



  - Clearly, the sample "rays" shown in the margin connecting different array elements to $(0, r, 0)$ have different lengths even though in *paraxial approximation* only one length, r, would be assigned to all them since $nd\cos\theta_x = 0$ for $\theta_x = 90°$.

This discrepancy between r and the *actual* ray length $|r\hat{y} - nd\hat{x}|$ would be the cause of the failure of paraxial approximation, except when the "phase error" caused by the discrepancy is *unimportant* (because it is small in radian units).

  - The exact phase delay along ray-0 is

$$\Phi_0 = kr$$

    since the field phasor arriving along this path from element $n = 0$ is $\propto e^{-jkr}$.

  - The exact phase delay along ray-M is

$$\Phi_M = k|r\hat{y} - Md\hat{x}| = k|r\hat{y} - \frac{L}{2}\hat{x}| = k\sqrt{r^2 + (\frac{L}{2})^2}$$

    since the field phasor arriving along this path from element $n = M$ is $\propto e^{-jk|r\hat{z} - Md\hat{x}|}$.

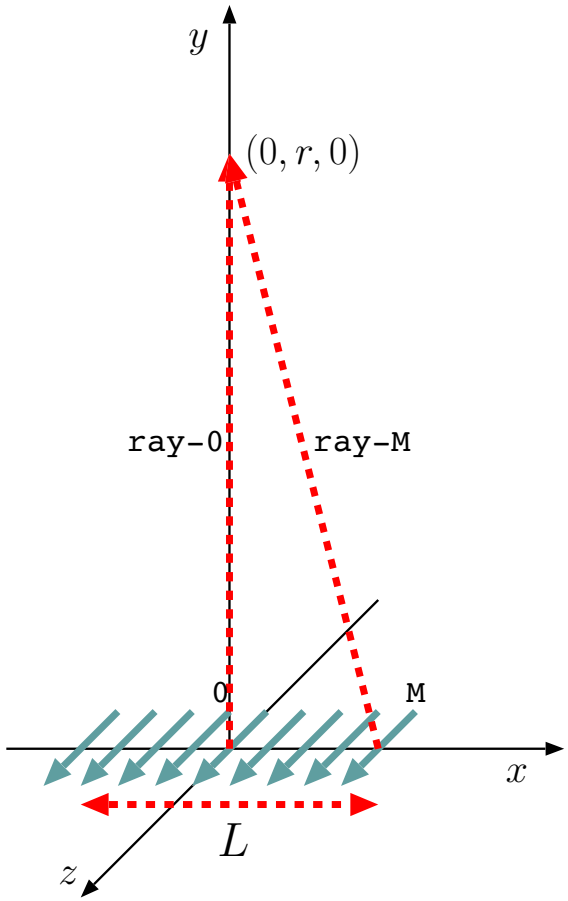
– The maximum phase error made in paraxial approximation is then

$$\begin{aligned}\Delta\Phi &= \Phi_M - \Phi_0 = k\sqrt{r^2 + (\frac{L}{2})^2} - kr = k(\sqrt{r^2 + (\frac{L}{2})^2} - r) \\ &= kr(\sqrt{1 + (\frac{L}{2r})^2} - 1).\end{aligned}$$

Note that this phase error vanishes when $r \to \infty$. But for a finite r, we have, when $r \gg \frac{L}{2}$, a finite error of about

$$\begin{aligned}\Delta\Phi &= kr(\sqrt{1 + (\frac{L}{2r})^2} - 1) \\ &\approx kr(1 + \frac{1}{2}(\frac{L}{2r})^2 - 1) = \frac{2\pi}{\lambda}\frac{L}{8r} = \frac{\pi}{8}\frac{2L}{\lambda r},\end{aligned}$$

using the first two terms of the binomial expansion of $\sqrt{1 + (\frac{L}{2r})^2}$.



– Clearly then, if we were to take

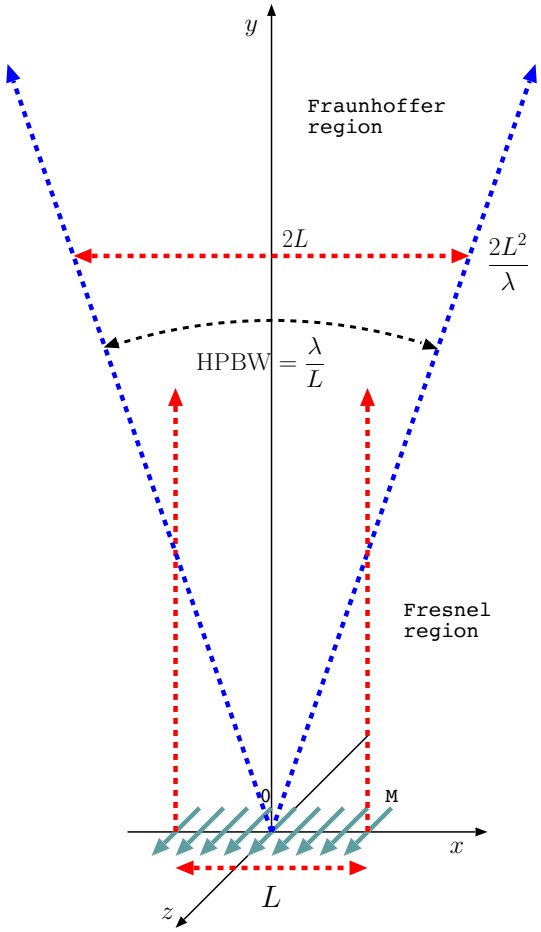
$$\frac{2L^2}{\lambda r} \lesssim 1 \quad \Leftrightarrow \quad r \gtrsim \frac{2L^2}{\lambda} \text{ then we would have } \Delta\Phi \lesssim \frac{\pi}{8}\,\text{rad},$$

which is a small enough of a phase error that can actually be neglected (in particular in multiple-element arrays where the phase errors due to a multitude of other elements will be even smaller than $\frac{\pi}{8}$ rad or 22.5°).

The analysis just concluded indicates that the border between Fresnel and Fraunhoffer zones can be taken as

$$r \sim \frac{2L^2}{\lambda},$$

the so-called *Rayleigh distance*.



- For the N-element broadside array being considered the far-field gain function (adapting from Lecture 12)

$$G(\theta,\phi) = D\sin^2\theta \frac{\sin^2(\frac{N}{2}kd\sin\theta\cos\phi)}{N^2\sin^2(\frac{1}{2}kd\sin\theta\cos\phi)}.$$

The array gain

$$G(\theta,0) = D\sin\theta \frac{\sin^2(\frac{N}{2}kd\sin\theta)}{N^2\sin^2(\frac{1}{2}kd\sin\theta)}$$

on $\phi = 0°$ has its "first nulls" around the mainlobe at angles θ satisfying

$$\frac{N}{2}kd\sin\theta_n = \frac{2\pi/\lambda}{2}\underbrace{Nd}_{L}\sin\theta_n = \pm\pi \quad \Rightarrow \quad L\sin\theta_n = \pm\lambda,$$

so that "beam-width between first nulls" is

$$\text{BWFN} = 2\theta_n \approx \frac{2\lambda}{L}$$

for $L \gg \lambda$.

- Approximately speaking, the "half-power beam width" between the points of $D/2$ in the gain-pattern works out to be

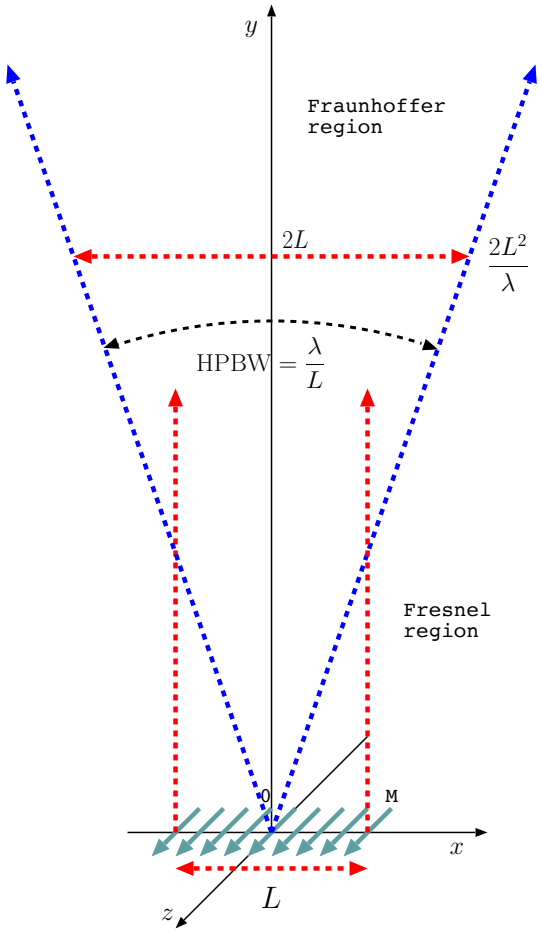
$$\text{HPBW} \approx \frac{1}{2}\text{BWFN} = \frac{\lambda}{L}$$

in radian units.

- Multiplying the HPBW with the Rayleigh distance we find that

$$\text{HPBW} \times \frac{2L^2}{\lambda} = \frac{\lambda}{L} \times \frac{2L^2}{\lambda} = 2L,$$

which indicates that at the border of Fraunhoffer region the "antenna beam" between its half-power points is about twice as wide in the transverse direction as the physical size of the array, as shown in the cartoon in the margin. This is a "physical picture" that should be kept in mind (and can be easily extrapolated into Fresnel and Fraunhoffer regions when needed).

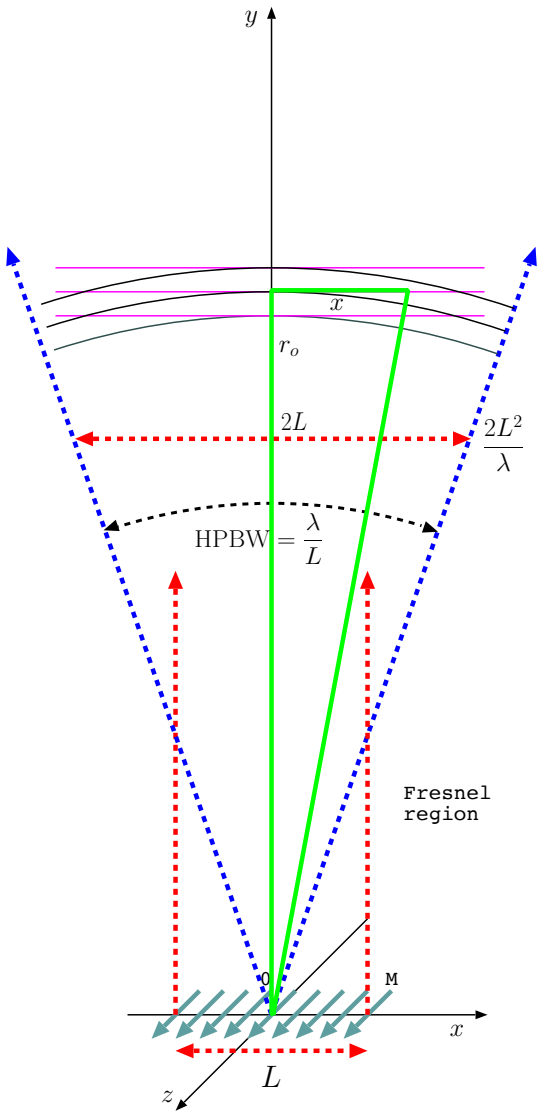


  - Note that increasing the array size L causes:
    1. A larger Rayleigh distance,
    2. A thicker column of radiation field in Fresnel region,
    3. A narrower HPBW in Fraunhoffer region.

The inverse relation between antenna size L and the HPBW, that can be summarized as

$$\text{HPBW} \times L = \lambda,$$

is reminiscent of "uncertainty relation" from quantum mechanics as well as the relation between bandwidth and impulse response length of filter circuits — underlying all such relationships is of course a Fourier transform pair (between frequency response and impulse response in filter circuits; between momentum and position wave functions in quantum mechanics; between effective length function and spatial current distribution in antennas).



- In the Fraunhoffer region, we can express the radiation field of a $\hat{z}$-polarized antenna or antenna array as

$$\tilde{\mathbf{E}}(\mathbf{r}) = j\eta_o I_o k \ell_{eff}(\theta,\phi)\sin\theta\frac{e^{-jkr}}{4\pi r}\hat{\theta}.$$

  - This "globally" spherical-wave field can be considered a plane-wave field "locally" in any neighborhood of

$$\mathbf{r} = \mathbf{r}_o \equiv (x_o, y_o, z_o) = (r_o, \theta_o, \phi_o)$$

    within the Fraunhoffer region.

  - The expression for plane-wave approximation in the neighborhood of $\mathbf{r} = \mathbf{r}_o$ is simply

$$\tilde{\mathbf{E}}_p(\mathbf{r}) \equiv j\eta_o I_o k \ell_{eff}(\theta_o,\phi_o)\sin\theta_o\frac{e^{-jk\hat{r}_o\cdot\mathbf{r}}}{4\pi r_o}\hat{\theta}_o,$$

    where

$$\hat{r}_o \equiv \frac{\mathbf{r}_o}{|\mathbf{r}_o|}.$$

This phasor expression, that approximates the spherical wave phasor in the neighborhood of $\mathbf{r} = \mathbf{r}_o$, and is identical to $\tilde{\mathbf{E}}(\mathbf{r})$ at $\mathbf{r} = \mathbf{r}_o$, is recognized as a *plane wave* because it has the same numerical value (as a complex vector) on *planes* of constant phase defined by

$$k\hat{r}_o \cdot \mathbf{r} = \text{const.}$$

perpendicular to unit vector $\hat{r}_o$. This is a plane wave propagating in direction $\hat{r}_o$ and is *assigned* a **wave vector**

$$\mathbf{k} = k\hat{r}_o.$$

More on the *wave vector* concept later on...

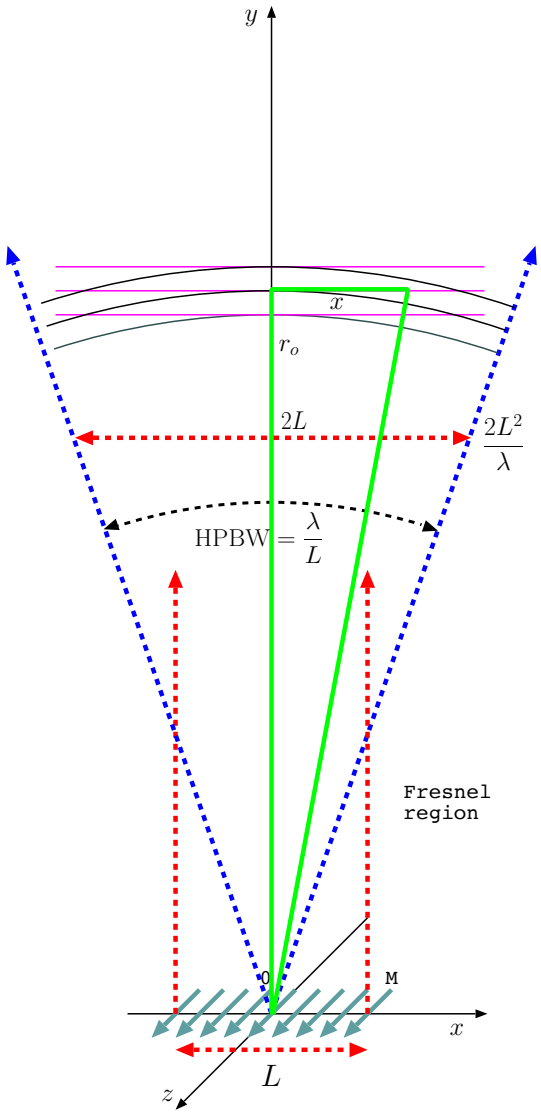
- Notice that the plane-wave field $\tilde{\mathbf{E}}_p(\mathbf{r})$ can also be expressed more compactly as

$$\tilde{\mathbf{E}}_p(\mathbf{r}) = \hat{\theta}_o|\tilde{\mathbf{E}}(\mathbf{r}_o)|e^{-jk\hat{r}_o\cdot\mathbf{r}},$$

disregarding a possible phase offset (position independent) equal to the angle of $jI_o\ell_{eff}(\theta_o, \phi_o)$.

- The deviation of this plane-wave field from the spherical-wave field $\tilde{\mathbf{E}}(\mathbf{r})$, as $\mathbf{r}$ departs from $\mathbf{r}_o$, will be dominated by the *discrepancies* in phase variations of $\tilde{\mathbf{E}}(\mathbf{r})$ and $\tilde{\mathbf{E}}_p(\mathbf{r})$, rather than the much slower variation of $|\tilde{\mathbf{E}}(\mathbf{r})|$ with respect to $|\tilde{\mathbf{E}}_p(\mathbf{r})|$.

  That is, the **wave-front curvature** of spherical $\tilde{\mathbf{E}}(\mathbf{r})$ will be the main cause of the differences that emerge between $\tilde{\mathbf{E}}(\mathbf{r})$ and $\tilde{\mathbf{E}}_p(\mathbf{r})$ as $\mathbf{r}$ departs from $\mathbf{r}_o$.



We next determine the size of a region around $\mathbf{r} = \mathbf{r}_o$ where this wave-front curvature can be neglected. Our criterion will be to keep the phase discrepancy between $\tilde{\mathbf{E}}(\mathbf{r})$ and $\tilde{\mathbf{E}}_p(\mathbf{r})$ due to wave front curvature sufficiently small.

- For simplicity, let

$$\mathbf{r}_o = r_o\hat{y}$$

and compare the phase delay of phasors$\tilde{\mathbf{E}}_p(\mathbf{r})$ and $\tilde{\mathbf{E}}(\mathbf{r})$ at $\mathbf{r} = \mathbf{r}_o + \hat{x}x$.

  - The phase delay of $\tilde{\mathbf{E}}_p(\mathbf{r})$ at $\mathbf{r} = \mathbf{r}_o + \hat{x}x$ is

$$\Phi_p = kr_o$$

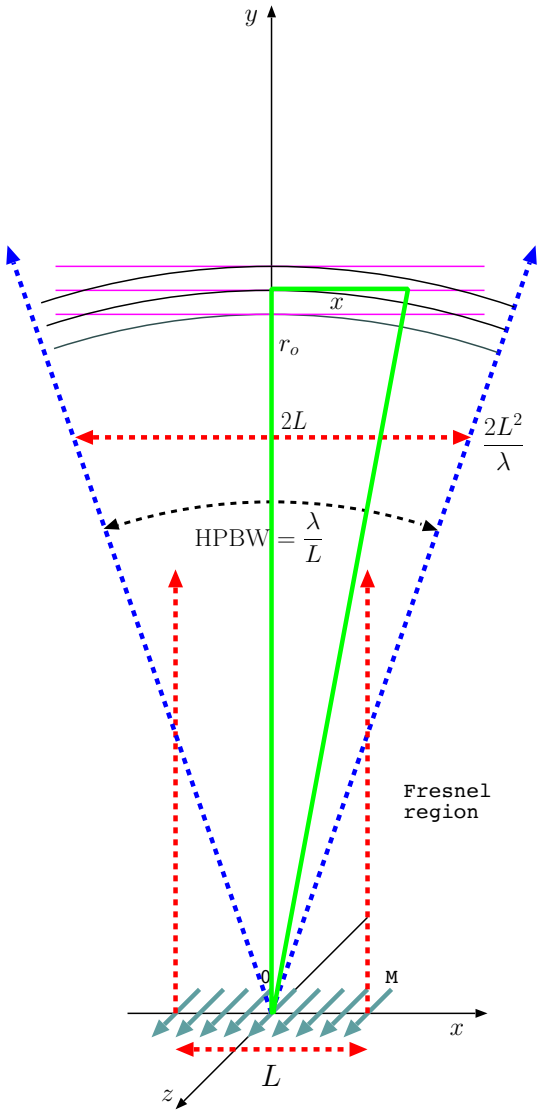
    since $\mathbf{r}_o + \hat{x}x$ and $\mathbf{r}_o$ reside on the same constant phase plane of $\tilde{\mathbf{E}}_p(\mathbf{r})$ (see margin).
  - The phase delay of $\tilde{\mathbf{E}}(\mathbf{r})$ at $\mathbf{r} = \mathbf{r}_o + \hat{x}x$ is

$$\Phi = k|r_o\hat{y} + x\hat{x}| = k\sqrt{r_o^2 + x^2}.$$

  - The phase discrepancy between $\tilde{\mathbf{E}}_p(\mathbf{r})$ and $\tilde{\mathbf{E}}(\mathbf{r})$ at $\mathbf{r} = \mathbf{r}_o + \hat{x}x$ because of wave front curvature is then

$$\begin{aligned}\Delta\Phi &= \Phi - \Phi_p = k\sqrt{r_o^2 + x^2} - kr_o = k(\sqrt{r_o^2 + x^2} - r_o) \\ &\approx kr_o(1 + \frac{1}{2}\frac{x^2}{r_o^2} - 1) = \frac{2\pi}{\lambda}\frac{x^2}{2r_o} = \frac{\pi}{4}\frac{(2x)^2}{\lambda r_o}\end{aligned}$$

    using the first two terms of the binomial expansion of $\sqrt{1 + \frac{x^2}{r_o^2}}$.



– Clearly then, if we were to take

$$\frac{(2x)^2}{\lambda r_o} \ll 1 \quad \Leftrightarrow \quad 2x < \sqrt{\lambda r_o} \text{ then we would have } \Delta\Phi \ll \frac{\pi}{4}\,\text{rad}.$$

Thus, plane-wave approximation of a spherical wave about position $\mathbf{r}_o$ will have negligible errors within a box with dimensions less than $\sqrt{\lambda r_o}$ known as Fresnel size.

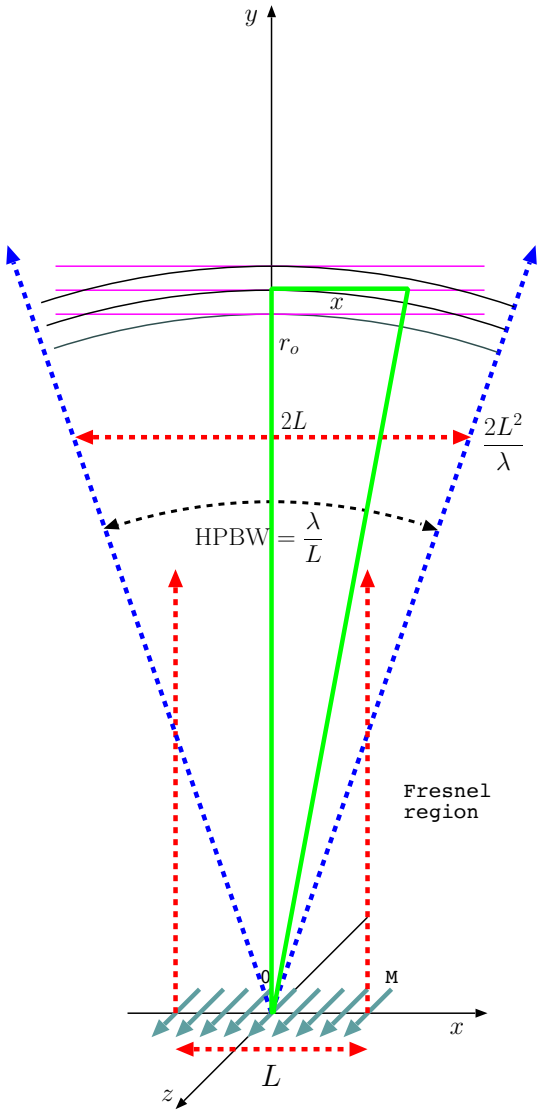
– Note that Fresnel size, the size of the region where plane wave approximation is acceptable, grows as the square root of distance r_o. Smallest meaningful value of Fresnel size is for

$$r_o = \frac{2L^2}{\lambda}, \quad \text{(Rayleigh distance)}$$

in which case

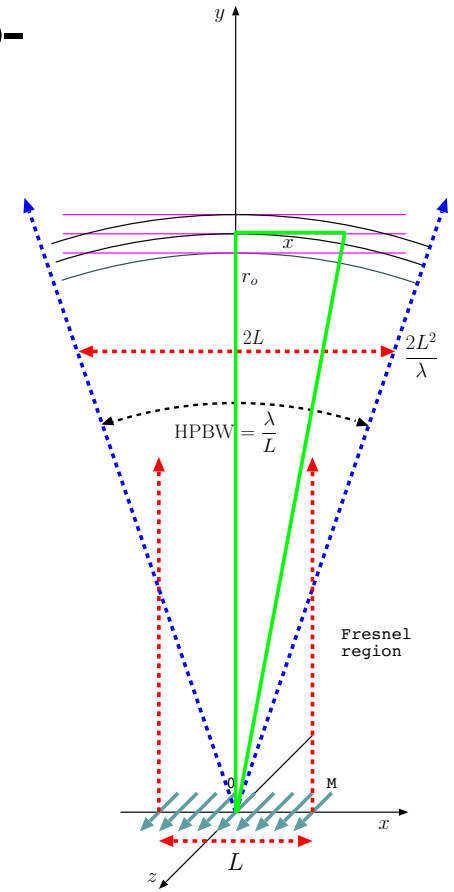
$$\text{Fresnel size} = \sqrt{\lambda r_o} = \sqrt{2}L < 2L = \text{beam size},$$

indicating that the antenna beam is always (at all $|\mathbf{r}_o|$) broader than a Fresnel size and thus only portions of an antenna beam can be well represented by a plane wave. A superposition of many (an infinite number in fact) plane waves would be required for an accurate representation of an entire beam.



Example: For $\lambda = 10$ m (30 MHz) and $r_o = 100$ km, we have $\sqrt{\lambda r_o} = 1$ km. So at a distance of 100 km away from an HF source the wave field looks planar over a neighborhood of about less than a km in extent (or about 100 wavelengths).

15 Plane-wave form of Maxwell's equations, propagation in arbitrary direction



Having seen how EM waves are generated by radiation sources and how spherical TEM waves develop a "planar" character over increasingly large regions as they propagate away from their sources, it is time to shift our attention to *propagation* and *guidance phenomena* using the plane-wave formalism.

Perhaps the most "practical" rationalization of this switch from spherical to plane-wave emphasis is that waves produced by compact sources invariably "look" planar at the scales of practical receiving systems (that will study near the end of this course) situated afar.

- We wish to study wave solutions of Maxwell's equations exhibiting the planar phasor form

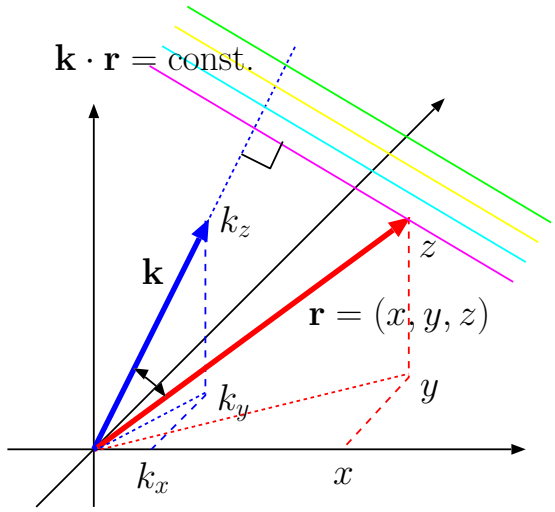
$$\tilde{\mathbf{E}} = \mathbf{E}_o e^{-j\mathbf{k}\cdot\mathbf{r}} = \hat{e} E_o e^{-j\mathbf{k}\cdot\mathbf{r}}$$

  and time-domain variations

$$\begin{aligned} \mathrm{Re}\{\tilde{\mathbf{E}} e^{j\omega t}\} &= \mathrm{Re}\{\mathbf{E}_o e^{j(\omega t - \mathbf{k}\cdot\mathbf{r})}\} \\ &= \hat{e}|E_o| \cos(\omega t - \mathbf{k}\cdot\mathbf{r} + \angle E_o) \end{aligned}$$

  where **wave vector k** is to be found in compliance with ω and Maxwell's equations according to some specific "dispersion relation" including the details of the propagation medium.

– For simplicity, the above phasor has been declared to be linearly polarized. Circular or elliptic polarized wave fields can be constructed later on via superposition methods.



- Linearly polarized wave field phasor above can be expanded as

$$\tilde{\mathbf{E}} = \mathbf{E}_o e^{-j\mathbf{k}\cdot\mathbf{r}} = \mathbf{E}_o e^{-j(k_x x + k_y y + k_z z)}$$

assuming a wave vector

$$\mathbf{k} = (k_x, k_y, k_z) = \hat{x}k_x + \hat{y}k_y + \hat{z}k_z$$

expressed in terms of its projections (k_x, k_y, k_z) along the Cartesian coordinate axes (x, y, z).

- A special case we are familiar with is

$$k_x = k_y = 0,\ k_z > 0,\ \text{ when } \mathbf{k} = k_z\hat{z} = k\hat{z} \text{ and } e^{-j\mathbf{k}\cdot\mathbf{r}} = e^{-jkz}$$

as in plane TEM waves traveling in $+z$ direction having a

$$\text{wavelength } \lambda = \frac{2\pi}{k} \text{ and propagation speed } v_p = \frac{\omega}{k}.$$

– Likewise, the case

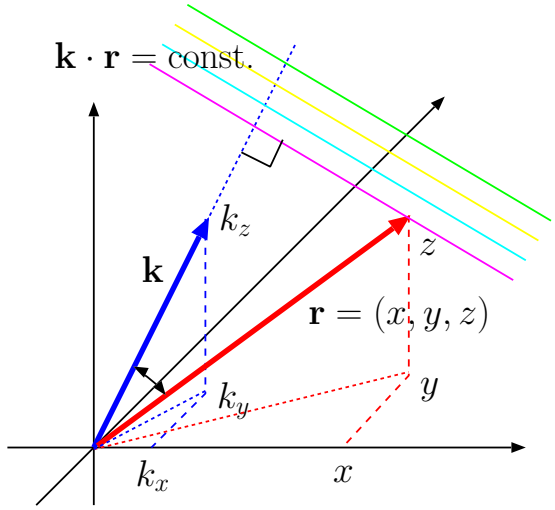
$$k_y = k_z = 0\,, k_x > 0,\ \text{ when } \mathbf{k} = k_x\hat{x} = k\hat{x} \text{ and } e^{-j\mathbf{k}\cdot\mathbf{r}} = e^{-jkx}$$

corresponds to plane TEM waves traveling in $+x$ direction with the same wavelength and propagation speed.

- The general case with non-zero components (k_x, k_y, k_z) corresponds to a plane wave propagating in the direction of unit vector

$$\hat{k} \equiv \frac{\mathbf{k}}{k} = \frac{(k_x, k_y, k_z)}{k} \text{ where } k \equiv |\mathbf{k}| = \sqrt{k_x^2 + k_y^2 + k_z^2} = \frac{2\pi}{\lambda}$$

and also having the same wavelength and propagation speed as above. Wavelength λ now describes the shift invariance of the wave field in spatial $\hat{k}$ direction, i.e., the propagation direction.



Example 1: A plane wave electric field phasor is specified as

$$\tilde{\mathbf{E}} = \hat{z} e^{-j(3\pi x - 4\pi y)} \frac{\text{V}}{\text{m}}.$$

Determine the propagation direction $\hat{k}$, wavenumber $k = |\mathbf{k}|$, wavelength $\lambda = \frac{2\pi}{k}$ and wave frequency $f = \frac{\omega}{2\pi}$ assuming a propagation speed $c = 3 \times 10^8$ m/s.

Solution: Contrasting $\tilde{\mathbf{E}}$ with $e^{-j(k_x x + k_y y + k_z z)}$, we note that

$$k_x = 3\pi \frac{\text{rad}}{\text{m}}, \quad k_y = -4\pi \frac{\text{rad}}{\text{m}}, \quad k_z = 0.$$

Hence, wave vector

$$\mathbf{k} = \hat{x} k_x + \hat{y} k_y + \hat{z} k_z = 3\pi \hat{x} - 4\pi \hat{y} \frac{\text{rad}}{\text{m}},$$

and wave number

$$k = |\mathbf{k}| = \sqrt{k_x^2 + k_y^2 + k_z^2} = \sqrt{(3\pi)^2 + (4\pi)^2 + 0^2} = \sqrt{25\pi^2} = 5\pi \frac{\text{rad}}{\text{m}}.$$

The propagation direction is specified by the unit vector

$$\hat{k} = \frac{\mathbf{k}}{k} = \frac{3\pi\hat{x} - 4\pi\hat{y}}{5\pi} = 0.6\hat{x} - 0.8\hat{y}.$$

The wavelength is

$$\lambda = \frac{2\pi}{k} = \frac{2\pi}{5\pi} = 0.4\,\text{m}.$$

Since

$$c = v_p = \frac{\omega}{k}$$

in general, it follows that

$$\omega = kc = 5\pi \times 3 \times 10^8 = 2\pi \times 7.5 \times 10^8\,\frac{\text{rad}}{\text{s}}$$

and

$$f = \frac{\omega}{2\pi} = 750 \times 10^6\,\text{Hz} = 750\,\text{MHz}.$$

- Based on what we learned in ECE 329, we recognize that the wave analyzed in Example 1 must have been propagating in free space.
- What are the constraints on wave vector $\mathbf{k}$ for plane waves propagating in arbitrary media?

To answer the above question, we will return to macroscopic-form Maxwell's equations written in phasor form (see margin) and examine under which conditions phasor solutions

$$\propto e^{-j\mathbf{k}\cdot\mathbf{r}}$$

can be applicable for all the field quantities in the absence of source currents $\tilde{\mathbf{J}}$ and their accompanying $\tilde{\rho}$.

- First, we note that in view of relation

$$\tilde{\mathbf{D}} = \epsilon\tilde{\mathbf{E}},$$

we can have plane-wave solutions of the form

$$\tilde{\mathbf{D}} = \mathbf{D}_o e^{-j\mathbf{k}\cdot\mathbf{r}} \text{ and } \tilde{\mathbf{E}} = \mathbf{E}_o e^{-j\mathbf{k}\cdot\mathbf{r}}$$

if and only if ϵ does not depend on position $\mathbf{r}$ (why?).

- Likewise, relation

$$\tilde{\mathbf{B}} = \mu\tilde{\mathbf{H}},$$

implies plane-wave solutions

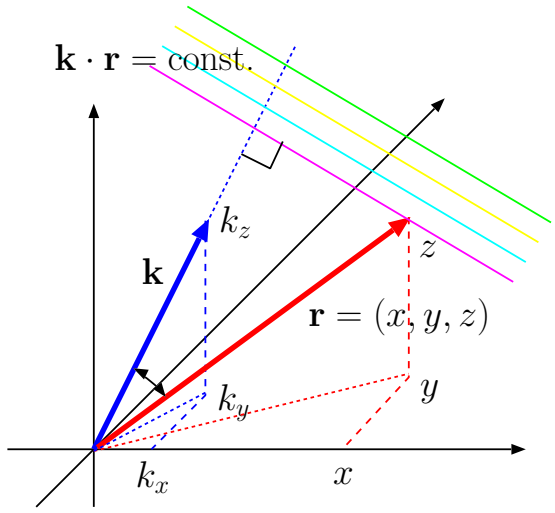
$$\tilde{\mathbf{B}} = \mathbf{B}_o e^{-j\mathbf{k}\cdot\mathbf{r}} \text{ and } \tilde{\mathbf{H}} = \mathbf{H}_o e^{-j\mathbf{k}\cdot\mathbf{r}}$$

if and only if μ does not depend on position $\mathbf{r}$ (why?).

$$\begin{aligned}
\nabla\cdot\tilde{\mathbf{D}} &= \tilde{\rho}\\
\nabla\cdot\tilde{\mathbf{B}} &= 0\\
\nabla\times\tilde{\mathbf{E}} &= -j\omega\tilde{\mathbf{B}}\\
\nabla\times\tilde{\mathbf{H}} &= \tilde{\mathbf{J}} + j\omega\tilde{\mathbf{D}}
\end{aligned}$$

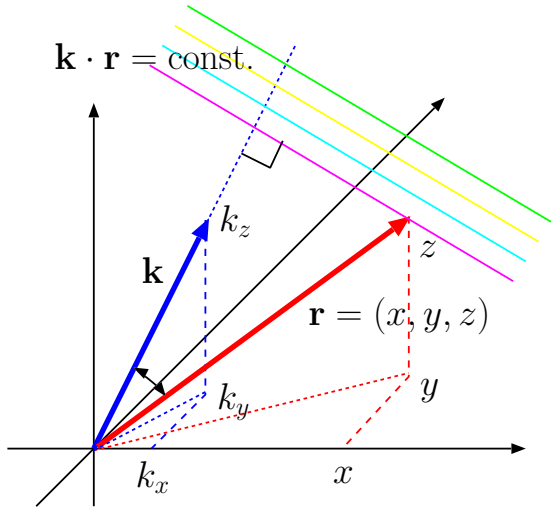
where (constitutive relations)

$$\begin{aligned}
\tilde{\mathbf{D}} &= \epsilon\tilde{\mathbf{E}}\\
\tilde{\mathbf{B}} &= \mu\tilde{\mathbf{H}}\\
\tilde{\mathbf{J}}_c &= \sigma\tilde{\mathbf{E}}.
\end{aligned}$$



- In a homogeneous region where ϵ, μ, and σ are, by definition, independent of $\mathbf{r}$, plane-wave solutions of phasor-form Maxwell's equations given in the margin become possible provided that

$$\begin{aligned} -j\mathbf{k}\cdot\tilde{\mathbf{D}} &= \tilde{\rho} \\ -j\mathbf{k}\cdot\tilde{\mathbf{B}} &= 0 \\ -j\mathbf{k}\times\tilde{\mathbf{E}} &= -j\omega\tilde{\mathbf{B}} \\ -j\mathbf{k}\times\tilde{\mathbf{H}} &= \tilde{\mathbf{J}} + j\omega\tilde{\mathbf{D}}. \end{aligned}$$



We have obtained these vector-algebraic relations from phasor-form Maxwell's equations in the margin after replacing the vector-differential operator ∇ by the vector-algebraic operator $-j\mathbf{k}$.

The justification of this simple procedure is as follows:

If

$$\tilde{\mathbf{D}} = \mathbf{D}_o e^{-j\mathbf{k}\cdot\mathbf{r}} = \mathbf{D}_o e^{-j(k_x x + k_y y + k_z z)} = (D_{xo}, D_{yo}, D_{zo}) e^{-j(k_x x + k_y y + k_z z)}$$

then

$$\begin{aligned} \nabla\cdot\tilde{\mathbf{D}} &= \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial y}\right)\cdot\left(D_{xo}e^{-j(k_x x + k_y y + k_z z)}, D_{yo}e^{-j(k_x x + k_y y + k_z z)}, D_{zo}e^{-j(k_x x + k_y y + k_z z)}\right) \\ &= -jk_x D_{xo}e^{-j(k_x x + k_y y + k_z z)} - jk_y D_{yo}e^{-j(k_x x + k_y y + k_z z)} - jk_z D_{zo}e^{-j(k_x x + k_y y + k_z z)} \\ &= -j(k_x, k_y, k_z)\cdot(D_{xo}, D_{yo}, D_{zo})e^{-j(k_x x + k_y y + k_z z)} = -j\mathbf{k}\cdot\tilde{\mathbf{D}}. \end{aligned}$$

Likewise, if

$$\tilde{\mathbf{E}} = \mathbf{E}_o e^{-j\mathbf{k}\cdot\mathbf{r}} = \mathbf{E}_o e^{-j(k_x x + k_y y + k_z z)} = (E_{xo}, E_{yo}, E_{zo}) e^{-j(k_x x + k_y y + k_z z)}$$

then

$$\begin{aligned}\nabla \times \tilde{\mathbf{E}} &= (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial y}) \times (E_{xo}e^{-j(k_x x+k_y y+k_z z)}, E_{yo}e^{-j(k_x x+k_y y+k_z z)}, E_{zo}e^{-j(k_x x+k_y y+k_z z)}) \\ &= (-jk_x, -jk_y, -jk_z) \times (E_{xo}e^{-j(k_x x+k_y y+k_z z)}, E_{yo}e^{-j(k_x x+k_y y+k_z z)}, E_{zo}e^{-j(k_x x+k_y y+k_z z)}) \\ &= -j\mathbf{k} \times \tilde{\mathbf{E}}.\end{aligned}$$

The vector-algebraic relations above, repeated in the margin (after canceling out some common terms), are known as plane-wave form of Maxwell's equations.

Plane-wave form of Maxwell's equations:

$$\begin{aligned}-j\mathbf{k} \cdot \tilde{\mathbf{D}} &= \tilde{\rho} \\ \mathbf{k} \cdot \tilde{\mathbf{B}} &= 0 \\ \mathbf{k} \times \tilde{\mathbf{E}} &= \omega\tilde{\mathbf{B}} \\ -j\mathbf{k} \times \tilde{\mathbf{H}} &= \tilde{\mathbf{J}} + j\omega\tilde{\mathbf{D}}.\end{aligned}$$

- Plane-wave form ME in the margin provide us with the constraints such plane waves satisfy in various types of propagation media categorized according to ϵ, μ, and σ.
- Focusing first on the case $\tilde{\rho} = \tilde{\mathbf{J}} = 0$ and $\sigma = 0$ (source free and non-conducting), the equations simplify as

$$\begin{aligned}\mathbf{k} \cdot \tilde{\mathbf{D}} &= 0 \\ \mathbf{k} \cdot \tilde{\mathbf{B}} &= 0 \\ \mathbf{k} \times \tilde{\mathbf{E}} &= \omega\tilde{\mathbf{B}} \\ -\mathbf{k} \times \tilde{\mathbf{H}} &= \omega\tilde{\mathbf{D}}.\end{aligned}$$

  The first two constraints tell us that wave vector $\mathbf{k}$ is necessarily orthogonal to both $\tilde{\mathbf{D}} = \epsilon\tilde{\mathbf{E}}$ and $\tilde{\mathbf{B}} = \mu\tilde{\mathbf{H}}$.

  - Hence, the plane waves satisfying the above equations will be TEM.

- Cross-multiplying the third equation with $\mathbf{k}$ and substituting from the fourth equation we get

$$\mathbf{k} \times (\mathbf{k} \times \tilde{\mathbf{E}}) = \omega\mu \mathbf{k} \times \tilde{\mathbf{H}} = \omega\mu(-\omega\tilde{\mathbf{D}}) = -\mu\epsilon\omega^2\tilde{\mathbf{E}}.$$

But we also have

$$\mathbf{k} \times (\mathbf{k} \times \tilde{\mathbf{E}}) = -(\mathbf{k} \cdot \mathbf{k})\tilde{\mathbf{E}}$$

since vectors $\mathbf{k}$ and $\tilde{\mathbf{E}}$ are perpendicular as shown in the margin — cross-multiplying $\tilde{\mathbf{E}}$ twice by $\mathbf{k} = k\hat{k}$ produces $-\tilde{\mathbf{E}}$ times $k^2 \equiv \mathbf{k} \cdot \mathbf{k}$!

  - The above lines are compatible if and only if

$$\mathbf{k} \cdot \mathbf{k} = \omega^2\mu\epsilon \quad \Rightarrow \quad \hat{k} \cdot \hat{k} = 1 \text{ and } k = \omega\sqrt{\mu\epsilon},$$

which is the **dispersion relation** of TEM plane-wave solutions of Maxwell' equations

$$\propto e^{-j\omega\sqrt{\mu\epsilon}\hat{k}\cdot\mathbf{r}}$$

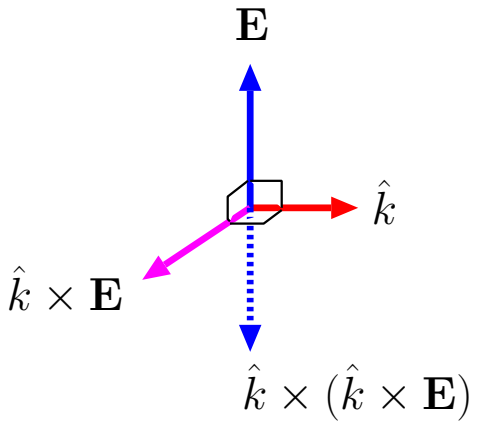
with

$$\hat{k} \cdot \tilde{\mathbf{E}} = 0 \quad \text{and} \quad \hat{k} \cdot \tilde{\mathbf{H}} = 0.$$

as well as (according to the last two equations in the margin)

$$\tilde{\mathbf{H}} = \frac{\hat{k} \times \tilde{\mathbf{E}}}{\eta} \quad \text{and} \quad \tilde{\mathbf{E}} = \eta\tilde{\mathbf{H}} \times \hat{k} \quad \text{with} \quad \eta = \sqrt{\frac{\mu}{\epsilon}}.$$

- TEM plane wave solutions obtained above correspond to *undamped uniform* plane waves when the **wavevector $\mathbf{k}$** obeying the **dispersion relation $\mathbf{k} \cdot \mathbf{k} = \omega^2\mu\epsilon$** is real valued.



Also, the vector identity

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

leads to the same result.

Plane-wave form of Maxwell's equations:

$$\begin{aligned} \mathbf{k} \cdot \tilde{\mathbf{D}} &= 0 \\ \mathbf{k} \cdot \tilde{\mathbf{B}} &= 0 \\ \mathbf{k} \times \tilde{\mathbf{E}} &= \omega\mu\tilde{\mathbf{H}} \\ -\mathbf{k} \times \tilde{\mathbf{H}} &= \omega\epsilon\tilde{\mathbf{E}}. \end{aligned}$$

- Same results also describe *damped* plane waves and/or *non-uniform* plane waves with *complex valued* $\mathbf{k}$:
  - Damped waves: if $\hat{k}$ is real but $k = \omega\sqrt{\mu\epsilon}$ is complex valued with a negative imaginary part - e.g., in Ohmic conductors
  - Non uniform waves: if $\hat{k}$, obeying $\hat{k} \cdot \hat{k} = 1$ is a complex valued unit vector - e.g., with *surface waves*, *evanescent waves* ... to be studied over the next few weeks
- Example: Non-uniform plane waves with real valued $\mathbf{k} \cdot \mathbf{k}$
  - Consider $\mathbf{k} \cdot \mathbf{k} = \omega^2 \mu_o \epsilon_o$ where the right hand side is real valued and equal to the square of ω/c.
  - Let $\mathbf{k} = \mathbf{k}_r + j\mathbf{k}_i$ where $\mathbf{k_r}$ and $\mathbf{k}_i$ are real valued.
  - Then $\mathbf{k} \cdot \mathbf{k} = (\mathbf{k}_r \cdot \mathbf{k}_r - \mathbf{k}_i \cdot \mathbf{k}_i) + j2\mathbf{k}_r \cdot \mathbf{k}_i = \omega^2 \mu_o \epsilon_o$ leading to the constraints

$$\begin{aligned} \mathbf{k}_r \cdot \mathbf{k}_r - \mathbf{k}_i \cdot \mathbf{k}_i &= \omega^2 \mu_o \epsilon_o \\ \mathbf{k}_r \cdot \mathbf{k}_i &= 0. \end{aligned}$$

  - For instance $\mathbf{k} = (k_x, k_y, k_z) = (2\pi, 0, -j\pi)$ will comply with these constraints with $\mathbf{k}_r = (2\pi, 0, 0)$ and $\mathbf{k}_i = (0, 0, -\pi)$ and $\omega^2 \mu_o \epsilon_o = 3\pi^2$, describing a non-uniform plane wave with a phasor

$$e^{-j\mathbf{k}\cdot\mathbf{r}} = e^{-j(2\pi x - j\pi z)} = e^{-j2\pi x} e^{-\pi z}$$

that *propagates* in x direction with a wavelength of $\lambda = 2\pi/k_x = 1$ m and *decays* in z direction ... namely a "surface wave" propagating along, say, $z = 0$ surface.

- Translating the wave phasor back to time domain, we see that it will be described as

$$\mathrm{Re}\{e^{-j\mathbf{k}\cdot\mathbf{r}}e^{j\omega t}\} = \mathrm{Re}\{e^{-j2\pi x}e^{-\pi z}e^{j\omega t}\} = e^{-\pi z}\cos(\omega t - 2\pi x).$$

16 Reflection and transmission, TE mode

- Last lecture we learned how to represent plane-TEM waves propagating in a direction $\hat{k}$ in terms of field phasors

$$\tilde{\mathbf{E}} = \mathbf{E}_o e^{-j\mathbf{k}\cdot\mathbf{r}} \text{ and } \tilde{\mathbf{H}} = \frac{\hat{k}\times\tilde{\mathbf{E}}}{\eta}$$

  such that

$$\eta = \sqrt{\frac{\mu}{\epsilon}}, \quad \mathbf{k} = k\hat{k}, \quad k = \omega\sqrt{\mu\epsilon}, \text{ and } \mathbf{k}\cdot\mathbf{E}_o = 0.$$

  Such waves are only permitted in **homogeneous** propagation media with constant μ and ϵ and zero σ.

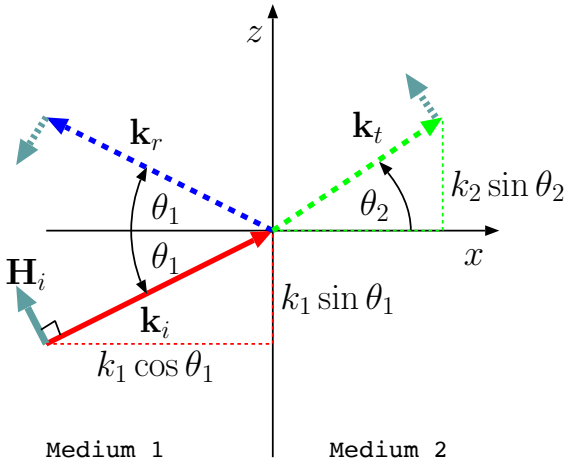
  - The condition of zero σ can be relaxed easily — in that case the above relations would still hold if we were to replace ϵ by $\epsilon + \frac{\sigma}{j\omega}$ as we will see later on.

- In this lecture we will examine the propagation of plane-TEM waves across two distinct homogeneous media having a planar interface between them.

- With no loss of generality we can choose unit vector $\hat{x}$ be the unit-normal of the interface plane separating **medium 1 in the region** $x < 0$ from **medium 2 in the region** $x > 0$.

In 1808 Etienne-Loius *Malus* discovered that light reflected from a surface at an oblique angle will in general be *polarized* differently than the incident wave on the reflecting surface.

This is caused by the difference of the reflection coefficients of TE and TM components of the incident wave as we will learn in this lecture. Practical implementation of the phenomenon include polarizers and polarizing filters used in optical instruments, photography, and LCD displays.



- A plane-TEM wave incident onto the interface from medium 1 is assigned a wavevector

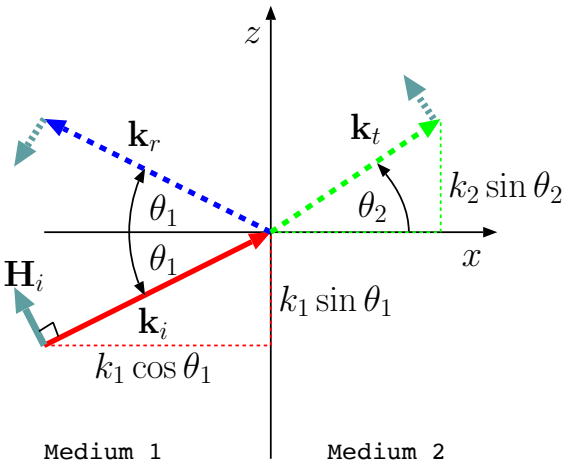
$$\mathbf{k}_i = k_1(\hat{x}\cos\theta_1 + \hat{z}\sin\theta_1)$$

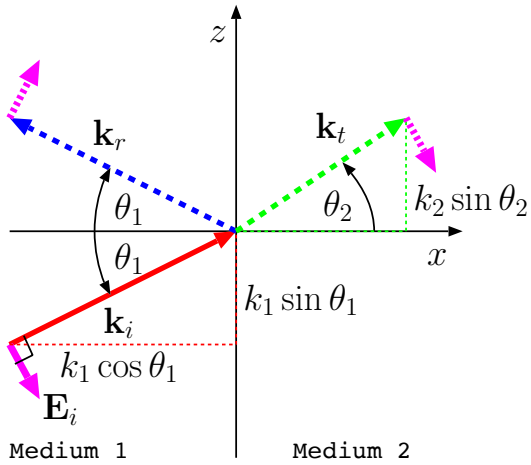
by taking $\hat{y}$ to be orthogonal to $\mathbf{k}_i$ (see margin).

This makes the xz-plane the "plane of incidence" and θ_1 the "angle of incidence", and, furthermore,

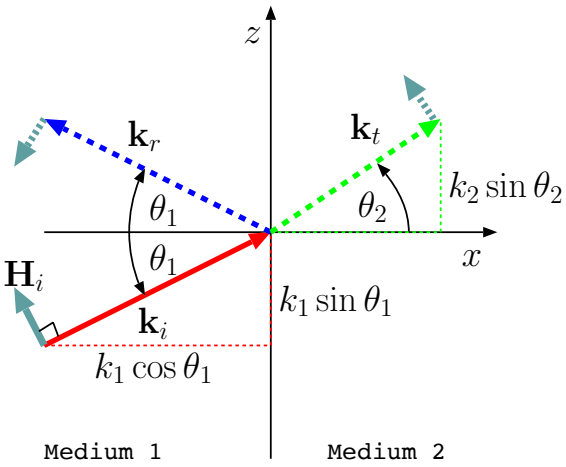
  - if we were to consider the case of $\tilde{\mathbf{E}}_i = \hat{y}E_o e^{-j\mathbf{k}_i\cdot\mathbf{r}}$ we would call the problem a "TE mode" problem, where TE is short for **T**ransverse **E**lectric field, and transverse is with respect to the plane of incidence.
  - if we were to consider the case of $\tilde{\mathbf{H}}_i = \hat{y}H_o e^{-j\mathbf{k}_i\cdot\mathbf{r}}$ we would call the problem a "TM mode" problem, where TM is short for **T**ransverse **M**agnetic field, and transverse is, once again, with respect to the plane of incidence.

This lecture we will examine the TE mode, and next lecture the TM mode. These different modes have different transmission and reflection properties. They are easy to study one at a time, and sufficient in general since all cases can be represented as a superposition of TE and TM cases.





TE mode reflection problem:



- In TE mode reflection problem, the plane-wave field phasors incident on the interface between medium 1 and 2 — $x = 0$ plane — are specified as

$$\tilde{\mathbf{E}}_i = \hat{y} E_o e^{-j\mathbf{k}_i \cdot \mathbf{r}} \text{ and } \tilde{\mathbf{H}}_i = \frac{\mathbf{k}_i \times \tilde{\mathbf{E}}_i}{k_1 \eta_1},$$

where

$$\mathbf{k}_i = k_1(\hat{x}\cos\theta_1 + \hat{z}\sin\theta_1),$$

and

$$k_1 = \frac{\omega}{v_1}, \; v_1 = \frac{1}{\sqrt{\mu_1 \epsilon_1}}, \; \eta_1 = \sqrt{\frac{\mu_1}{\epsilon_1}}.$$

  - The plane-wave field specified above satisfies Maxwell's equations in the homogeneous medium 1 occupying the region $x < 0$, but if there were no other fields in media 1 and 2, *Maxwell's boundary condition equations* requiring the continuity of tangential $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{H}}$ across any boundary not supporting a surface current would be violated.

In order to comply with the boundary condition equations we postulate a set of *reflected* and *transmitted* wave fields in media 1 and 2 as follows:

- In medium 1 we postulate a **reflected plane-wave** with field phasors

$$\tilde{\mathbf{E}}_r = \hat{y} E_o \Gamma_\perp e^{-j\mathbf{k}_r \cdot \mathbf{r}} \text{ and } \tilde{\mathbf{H}}_r = \frac{\mathbf{k}_r \times \tilde{\mathbf{E}}_r}{k_1 \eta_1},$$

where

$$\mathbf{k}_r = k_1(-\hat{x}\cos\theta_1 + \hat{z}\sin\theta_1).$$

- In medium 2 we postulate a **transmitted plane-wave** with field phasors

$$\tilde{\mathbf{E}}_t = \hat{y}E_o\tau_\perp e^{-j\mathbf{k}_t\cdot\mathbf{r}} \text{ and } \tilde{\mathbf{H}}_t = \frac{\mathbf{k}_t \times \tilde{\mathbf{E}}_t}{k_2\eta_2},$$

where

$$\mathbf{k}_t = k_2(\hat{x}\cos\theta_2 + \hat{z}\sin\theta_2).$$

and

$$k_2 = \frac{\omega}{v_2},\ v_2 = \frac{1}{\sqrt{\mu_2\epsilon_2}},\ \eta_2 = \sqrt{\frac{\mu_2}{\epsilon_2}}.$$

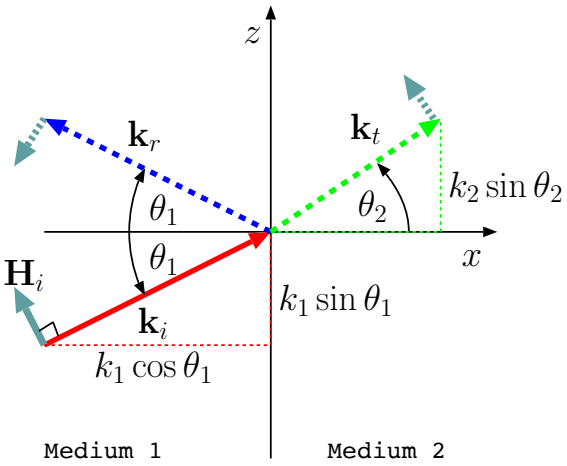
- To justify our postulates and determine a set of *reflection* and *transmission* coefficients $\Gamma_\perp$ and $\tau_\perp$ — defined in terms of electric field components — we will next apply the boundary conditions on $x = 0$ surface, where (using $x = 0$)

$$\tilde{\mathbf{E}}_i = \hat{y}E_o e^{-jk_1\sin\theta_1 z},\ \tilde{\mathbf{E}}_r = \hat{y}E_o\Gamma_\perp e^{-jk_1\sin\theta_1 z},\ \tilde{\mathbf{E}}_t = \hat{y}E_o\tau_\perp e^{-jk_2\sin\theta_2 z}.$$

Clearly, with these field components tangential continuity of the total field phasor $\tilde{\mathbf{E}}$ across $x = 0$ surface will be satisfied for all z if and only if

$$e^{-jk_1\sin\theta_1 z} + \Gamma_\perp e^{-jk_1\sin\theta_1 z} = \tau_\perp e^{-jk_2\sin\theta_2 z},$$

which is only possible (non-trivially) if



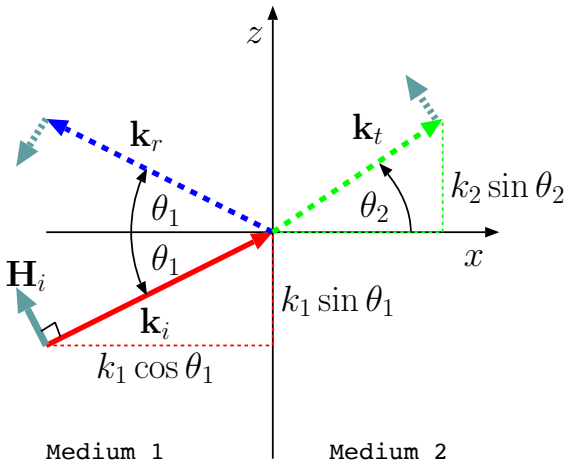
1. A "phase matching" condition[1]

$$k_1 \sin\theta_1 = k_2 \sin\theta_2$$

known as **Snell's law** is satisfied, and

2. $\Gamma_\perp$ and $\tau_\perp$ satisfy

$$1 + \Gamma_\perp = \tau_\perp.$$



- Tangential components of $\tilde{\mathbf{H}}_i$, $\tilde{\mathbf{H}}_r$, and $\tilde{\mathbf{H}}_t$ on $x = 0$ plane are obtained from

$$\tilde{\mathbf{E}}_i = \hat{y} E_o e^{-jk_1 \sin\theta_1 z}, \quad \tilde{\mathbf{E}}_r = \hat{y} E_o \Gamma_\perp e^{-jk_1 \sin\theta_1 z}, \quad \tilde{\mathbf{E}}_t = \hat{y} E_o \tau_\perp e^{-jk_2 \sin\theta_2 z}.$$

as

$$\hat{z}\cdot\tilde{\mathbf{H}}_i = \frac{E_o \cos\theta_1}{\eta_1} e^{-jk_1 \sin\theta_1 z}, \quad \hat{z}\cdot\tilde{\mathbf{H}}_r = -\frac{E_o \Gamma_\perp \cos\theta_1}{\eta_1} e^{-jk_1 \sin\theta_1 z}, \quad \hat{z}\cdot\tilde{\mathbf{H}}_t = \frac{E_o \tau_\perp \cos\theta_2}{\eta_2} e^{-jk_2 \sin\theta_2 z}.$$

Clearly, given Snell's law, tangential continuity of the total field phasor $\tilde{\mathbf{H}}$ across $x = 0$ surface will then be satisfied for all z if and only if

$$\frac{\cos\theta_1}{\eta_1} - \frac{\cos\theta_1}{\eta_1}\Gamma_\perp = \frac{\cos\theta_2}{\eta_2}\tau_\perp.$$

Combining this with

$$1 + \Gamma_\perp = \tau_\perp,$$

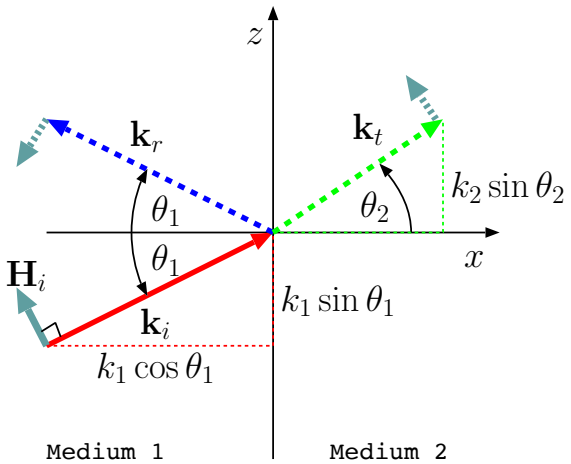
[1]Note that "Snell's law" can also be interpreted as having the components of wavevectors $\mathbf{k}_i$ and $\mathbf{k}_t$ equal along the interface between media 1 and 2.

we find that

$$\frac{\cos\theta_1}{\eta_1} - \frac{\cos\theta_1}{\eta_1}\Gamma_\perp = \frac{\cos\theta_2}{\eta_2}(1+\Gamma_\perp) \quad \Rightarrow \quad \Gamma_\perp = \frac{\eta_2\cos\theta_1 - \eta_1\cos\theta_2}{\eta_2\cos\theta_1 + \eta_1\cos\theta_2}$$

and

$$\tau_\perp = 1 + \Gamma_\perp = \frac{2\eta_2\cos\theta_1}{\eta_2\cos\theta_1 + \eta_1\cos\theta_2}.$$



- **Conclusion:** In TE reflection problem, the Fresnel reflection and transmission coefficients are

$$\Gamma_\perp \equiv \frac{E_{yr}}{E_{yi}} = \frac{\eta_2\cos\theta_1 - \eta_1\cos\theta_2}{\eta_2\cos\theta_1 + \eta_1\cos\theta_2} \text{ and } \tau_\perp = \frac{E_{yt}}{E_{yi}} = \frac{2\eta_2\cos\theta_1}{\eta_2\cos\theta_1 + \eta_1\cos\theta_2},$$

respectively. The coefficients enable us to express the reflected and transmitted wave phasors in terms of the incident-wave electric field phasor at the origin (i.e., E_{yi}).

Example 1: Medium 2 is vacuum while medium 1 has $\mu_1 = \mu_o$ and $\epsilon_1 = 2\epsilon_o$. Given that a TE mode incident field phasor

$$\tilde{\mathbf{E}}_i = \hat{y}5e^{-jk_1(\cos 30^\circ x + \sin 30^\circ z)}\,\frac{\text{V}}{\text{m}},$$

determine $\tilde{\mathbf{E}}_r$, $\tilde{\mathbf{E}}_t$, and $\tilde{\mathbf{H}}_t$.

Solution: We have

$$\eta_1 = \sqrt{\frac{\mu_1}{\epsilon_1}} = \sqrt{\frac{\mu_o}{2\epsilon_o}} = \frac{\eta_o}{\sqrt{2}} \text{ and } \eta_2 = \eta_o.$$

Also, according to Snell's law,

$$k_1 \sin\theta_1 = k_2 \sin\theta_2 \;\Rightarrow\; \sin\theta_2 = \frac{k_1}{k_2}\sin\theta_1 = \frac{\sqrt{\epsilon_1\mu_1}}{\sqrt{\epsilon_2\mu_2}}\sin\theta_1 = \sqrt{2}\sin 30^\circ = \frac{1}{\sqrt{2}},$$

indicating that

$$\theta_2 = 45^\circ.$$

Now, the reflection coefficient is

$$\Gamma_\perp \equiv \frac{E_{yr}}{E_{yi}} = \frac{\eta_2\cos\theta_1 - \eta_1\cos\theta_2}{\eta_2\cos\theta_1 + \eta_1\cos\theta_2} = \frac{\eta_o\frac{\sqrt{3}}{2} - \frac{\eta_o}{\sqrt{2}}\frac{1}{\sqrt{2}}}{\eta_o\frac{\sqrt{3}}{2} + \frac{\eta_o}{\sqrt{2}}\frac{1}{\sqrt{2}}} = \frac{\frac{\sqrt{3}}{2} - \frac{1}{2}}{\frac{\sqrt{3}}{2} + \frac{1}{2}} \approx 0.268.$$

The transmission coefficient is

$$\tau_\perp = \frac{E_{yt}}{E_{yi}} = \frac{2\eta_2\cos\theta_1}{\eta_2\cos\theta_1 + \eta_1\cos\theta_2} = \frac{2\eta_o\frac{\sqrt{3}}{2}}{\eta_o\frac{\sqrt{3}}{2} + \frac{\eta_o}{\sqrt{2}}\frac{1}{\sqrt{2}}} \approx 1.268.$$

Consequently, the reflected and transmitted wave phasors are

$$\tilde{\mathbf{E}}_r = \hat{y}(5 \times 0.268)e^{-jk_1(-\cos 30^\circ x + \sin 30^\circ z)}\,\frac{\text{V}}{\text{m}}$$

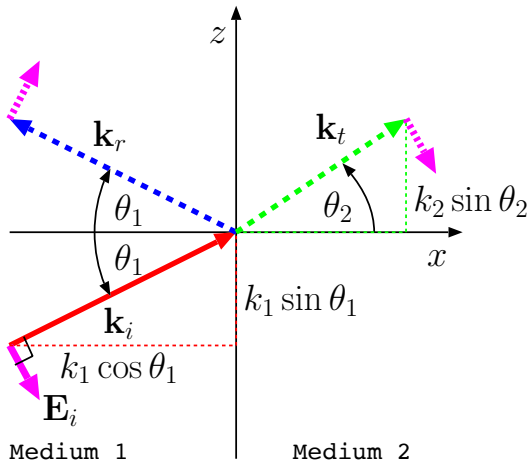
and

$$\tilde{\mathbf{E}}_t = \hat{y}(5 \times 1.268)e^{-jk_2(\cos 45^\circ x + \sin 45^\circ z)} \frac{\text{V}}{\text{m}}.$$

Finally,

$$\begin{aligned} \tilde{\mathbf{H}}_t &= \frac{\mathbf{k}_2 \times \tilde{\mathbf{E}}_t}{k_2\eta_2} = \frac{(\cos 45^\circ \hat{x} + \sin 45^\circ \hat{z}) \times \hat{y}(5 \times 1.268)e^{-jk_2(\cos 45^\circ x + \sin 45^\circ z)}}{\eta_o} \\ &= \frac{(\hat{z} - \hat{x})(5 \times 1.268)e^{-jk_2(\cos 45^\circ x + \sin 45^\circ z)}}{120\pi\sqrt{2}} \frac{\text{A}}{\text{m}}. \end{aligned}$$

17 Reflection and transmission, TM mode



- In TM mode reflection problem, the plane-wave field phasors incident on the interface between medium 1 and 2 — $x = 0$ plane — are specified as

$$\tilde{\mathbf{H}}_i = \hat{y} H_o e^{-j\mathbf{k}_i \cdot \mathbf{r}} \text{ and } \tilde{\mathbf{E}}_i = -\eta_1 \frac{\mathbf{k}_i \times \tilde{\mathbf{H}}_i}{k_1},$$

where

$$\mathbf{k}_i = k_1(\hat{x}\cos\theta_1 + \hat{z}\sin\theta_1),$$

and

$$k_1 = \frac{\omega}{v_1}, \; v_1 = \frac{1}{\sqrt{\mu_1 \epsilon_1}}, \; \eta_1 = \sqrt{\frac{\mu_1}{\epsilon_1}}.$$

- In medium 1 we postulate a **reflected plane-wave** with field phasors

$$\tilde{\mathbf{H}}_r = \hat{y} H_o R e^{-j\mathbf{k}_r \cdot \mathbf{r}} \text{ and } \tilde{\mathbf{E}}_r = -\eta_1 \frac{\mathbf{k}_r \times \tilde{\mathbf{H}}_r}{k_1},$$

where

$$\mathbf{k}_r = k_1(-\hat{x}\cos\theta_1 + \hat{z}\sin\theta_1).$$

- In medium 2 we postulate a **transmitted plane-wave** with field phasors

$$\tilde{H}_t = \hat{y} H_o T e^{-j\mathbf{k}_t \cdot \mathbf{r}} \text{ and } \tilde{\mathbf{E}}_t = -\eta_2 \frac{\mathbf{k}_t \times \tilde{\mathbf{H}}_t}{k_2},$$

where

$$\mathbf{k}_t = k_2(\hat{x}\cos\theta_2 + \hat{z}\sin\theta_2).$$

and

$$k_2 = \frac{\omega}{v_2},\ v_2 = \frac{1}{\sqrt{\mu_2\epsilon_2}},\ \eta_2 = \sqrt{\frac{\mu_2}{\epsilon_2}}.$$

- To justify our postulates and determine a set of reflection and transmission coefficients R and T — defined in terms of magnetic field components — we will next apply the boundary conditions on $x = 0$ surface, where (using $x = 0$)

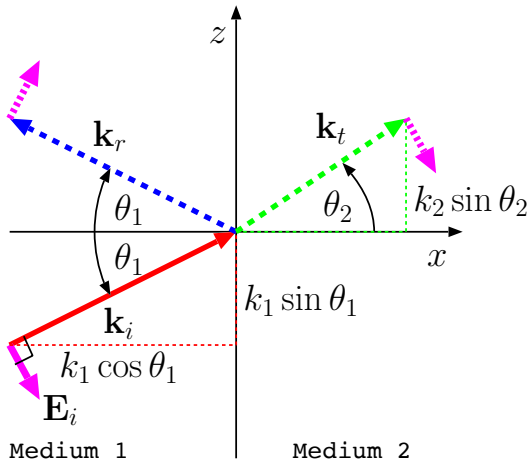
$$\tilde{\mathbf{H}}_i = \hat{y}H_o e^{-jk_1 \sin\theta_1 z},\ \ \tilde{\mathbf{H}}_r = \hat{y}H_o R e^{-jk_1 \sin\theta_1 z},\ \ \tilde{\mathbf{H}}_t = \hat{y}H_o T e^{-jk_2 \sin\theta_2}.$$

Clearly, with these field components tangential continuity of the total field phasor $\tilde{\mathbf{H}}$ across $x = 0$ surface will be satisfied for all z if and only if

$$e^{-jk_1 \sin\theta_1 z} + Re^{-jk_1 \sin\theta_1 z} = Te^{-jk_2 \sin\theta_2}$$

which — given Snell's law — is only possible (non-trivially) if

$$1 + R = T.$$



- Tangential components of $\tilde{\mathbf{E}}_i$, $\tilde{\mathbf{E}}_r$, and $\tilde{\mathbf{E}}_t$ on $x = 0$ plane are then obtained from

$$\tilde{\mathbf{H}}_i = \hat{y}H_o e^{-jk_1 \sin\theta_1 z},\ \ \tilde{\mathbf{H}}_r = \hat{y}H_o R e^{-jk_1 \sin\theta_1 z},\ \ \tilde{\mathbf{H}}_t = \hat{y}H_o T e^{-jk_2 \sin\theta_2 z}.$$

as

$$\hat{z}\cdot\tilde{\mathbf{E}}_i = -\eta_1 \cos\theta_1 H_o e^{-jk_1 \sin\theta_1 z},\ \ \hat{z}\cdot\tilde{\mathbf{E}}_r = \eta_1 \cos\theta_1 H_o R e^{-jk_1 \sin\theta_1 z},\ \ \hat{z}\cdot\tilde{\mathbf{E}}_t = -\eta_2 \cos\theta_2 H_o T e^{-jk_2 \sin\theta_2 z}.$$

Clearly, given Snell's law, tangential continuity of the total field phasor $\tilde{\mathbf{E}}$ across $x = 0$ surface will be satisfied for all z if and only if

$$\eta_1 \cos\theta_1 (1 - R) = \eta_2 \cos\theta_2 T$$

Combining this with

$$1 + R = T,$$

we find that

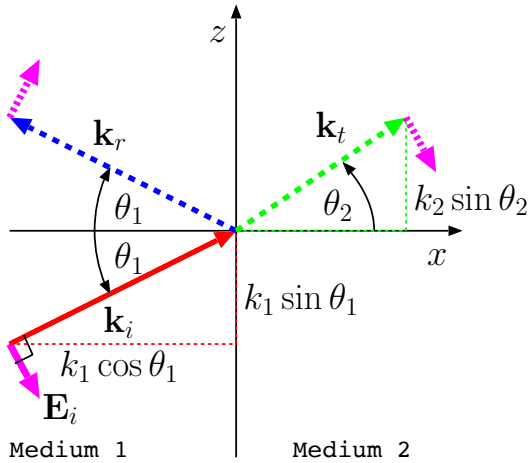
$$\eta_1 \cos\theta_1 (1 - R) = \eta_2 \cos\theta_2 (1 + R) \quad \Rightarrow$$

$$R = \frac{H_{yr}}{H_{yi}} = \frac{\eta_1 \cos\theta_1 - \eta_2 \cos\theta_2}{\eta_1 \cos\theta_1 + \eta_2 \cos\theta_2} = \frac{E_r}{E_i} = -\underbrace{\frac{E_{zr}}{E_{zi}}}_{\equiv \Gamma_\parallel},$$

and

$$1 + \frac{\eta_1 \cos\theta_1 - \eta_2 \cos\theta_2}{\eta_1 \cos\theta_1 + \eta_2 \cos\theta_2} = T \quad \Rightarrow$$

$$T = \frac{H_{yt}}{H_{yi}} = \frac{2\eta_1 \cos\theta_1}{\eta_1 \cos\theta_1 + \eta_2 \cos\theta_2} = \frac{\eta_1}{\eta_2} \underbrace{\frac{E_t}{E_i}}_{\equiv \tau_\parallel}.$$



– E_i, E_r, E_t refer above to the (phasor) amplitudes of the incident, reflected, and transmitted *electric field* vectors at the origin, pointing in the reference unit vector directions along $\tilde{\mathbf{H}} \times \mathbf{k}$ indicated by the arrows shown in magenta in the margin, while

– $\Gamma_\parallel$ and $\tau_\parallel$ introduced above are known as *Fresnel coefficients* for TM polarized reflection and transmission, respectively.

- **Conclusion:** In TM case, the Fresnel reflection and transmission coefficients for plane-wave electric fields are

$$\Gamma_\parallel \equiv \frac{E_{rz}}{E_{iz}} = -\frac{E_r}{E_i} = \frac{\eta_2 \cos\theta_2 - \eta_1 \cos\theta_1}{\eta_2 \cos\theta_2 + \eta_1 \cos\theta_1} \text{ and } \tau_\parallel \equiv \frac{E_t}{E_i} = \frac{2\eta_2 \cos\theta_1}{\eta_2 \cos\theta_2 + \eta_1 \cos\theta_1},$$

respectively. The coefficients enable us to express the reflected and transmitted wave phasors in terms of the incident-wave electric field phasor at the origin (i.e., E_i). Note that $1 + \Gamma_\parallel$is no longer $\tau_\parallel$!!!

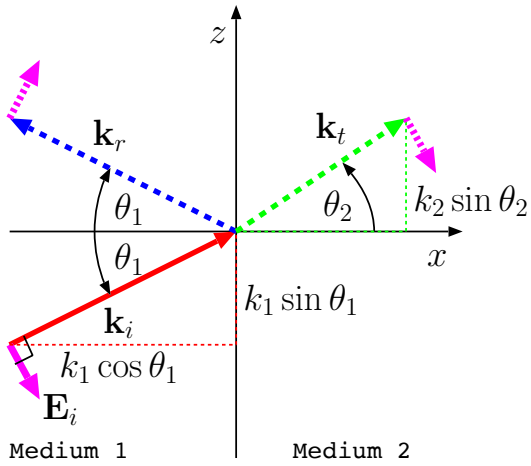
- Note that for $\theta_1 \to 0$, the Snell's law implies $\theta_1 \to 0$ also, in which case

$$\Gamma_\parallel = \frac{E_{zr}}{E_{zi}} = \frac{\eta_2 - \eta_1}{\eta_2 + \eta_1} = \Gamma_\perp = \frac{E_{yr}}{E_{yi}}$$

and

$$\tau_\parallel = \frac{E_t}{E_i} = \frac{2\eta_2}{\eta_2 + \eta_1} = \tau_\perp = \frac{E_{yt}}{E_{yi}}$$

(as one would have hoped for) since the distinction between TE and TM cases vanishes in this limit.



Example 1: Medium 2 is vacuum while medium 1 has $\mu_1 = \mu_o$ and $\epsilon_1 = 2\epsilon_o$. A TM mode plane-wave with an electric field amplitude of 1 V/m is incident on medium 2 with an angle of incidence of $\theta_1 = 30°$. Determine the wave-field phasors $\tilde{\mathbf{E}}_i$, $\tilde{\mathbf{E}}_r$, and $\tilde{\mathbf{E}}_t$.

Solution: The described incident wave field can be represented as

$$\tilde{\mathbf{E}}_i = (\sin 30°\hat{x} - \cos 30°\hat{z})e^{-jk_1(\cos 30° x + \sin 30° z)}\,\frac{\text{V}}{\text{m}}.$$

Also

$$\eta_1 = \sqrt{\frac{\mu_1}{\epsilon_1}} = \sqrt{\frac{\mu_o}{2\epsilon_o}} = \frac{\eta_o}{\sqrt{2}} \text{ and } \eta_2 = \eta_o.$$

Snell's law gives

$$k_1 \sin\theta_1 = k_2 \sin\theta_2 \;\Rightarrow\; \sin\theta_2 = \frac{k_1}{k_2}\sin\theta_1 = \frac{\sqrt{\epsilon_1\mu_1}}{\sqrt{\epsilon_2\mu_2}}\sin\theta_1 = \sqrt{2}\sin 30° = \frac{1}{\sqrt{2}},$$

indicating that

$$\theta_2 = 45°.$$

The TM mode reflection coefficient is

$$\Gamma_\parallel \equiv -\frac{E_r}{E_i} = \frac{\eta_2\cos\theta_2 - \eta_1\cos\theta_1}{\eta_2\cos\theta_2 + \eta_1\cos\theta_1} = \frac{\eta_o\frac{1}{\sqrt{2}} - \frac{\eta_o}{\sqrt{2}}\frac{\sqrt{3}}{2}}{\eta_o\frac{1}{\sqrt{2}} + \frac{\eta_o}{\sqrt{2}}\frac{\sqrt{3}}{2}} = \frac{1-\sqrt{3}/2}{1+\sqrt{3}/2} \approx 0.0718.$$

The transmission coefficient is

$$\tau_\parallel \equiv \frac{E_t}{E_i} = \frac{2\eta_2\cos\theta_1}{\eta_2\cos\theta_2 + \eta_1\cos\theta_1} = \frac{2\eta_o\frac{\sqrt{3}}{2}}{\eta_o\frac{1}{\sqrt{2}} + \frac{\eta_o}{\sqrt{2}}\frac{\sqrt{3}}{2}} \approx 1.3127$$

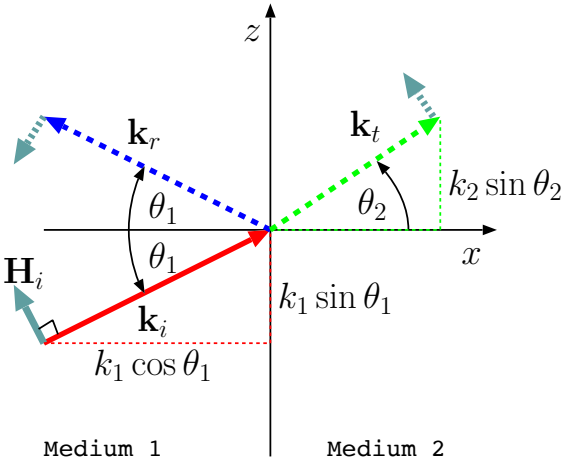
Consequently, the reflected and transmitted wave phasors are

$$\tilde{\mathbf{E}}_r = -0.0718(\sin 30°\hat{x} + \cos 30°\hat{z})e^{-jk_1(-\cos 30° x + \sin 30° z)}\,\frac{\text{V}}{\text{m}}.$$

and

$$\tilde{\mathbf{E}}_t = 1.3127(\sin 45^\circ \hat{x} - \cos 45^\circ \hat{z}) e^{-jk_2(\cos 45^\circ x + \sin 45^\circ z)} \frac{\text{V}}{\text{m}}.$$

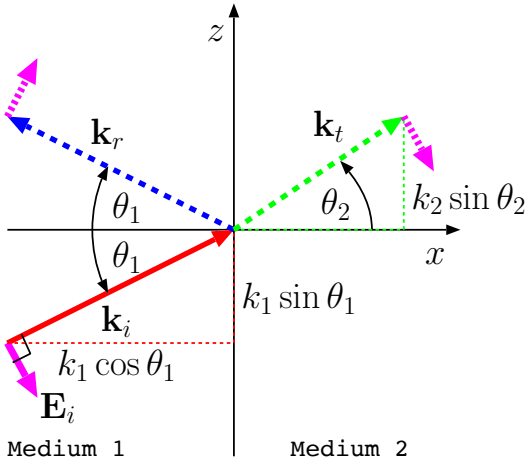
TE reflection:



$$\Gamma_\perp \equiv \frac{E_{yr}}{E_{yi}} = \frac{\eta_2 \cos\theta_1 - \eta_1 \cos\theta_2}{\eta_2 \cos\theta_1 + \eta_1 \cos\theta_2}$$

$$\tau_\perp \equiv \frac{E_{yt}}{E_{yi}} = \frac{2\eta_2 \cos\theta_1}{\eta_2 \cos\theta_1 + \eta_1 \cos\theta_2},$$

TM reflection:



$$\Gamma_\parallel \equiv -\frac{E_r}{E_i} = \frac{\eta_2 \cos\theta_2 - \eta_1 \cos\theta_1}{\eta_2 \cos\theta_2 + \eta_1 \cos\theta_1} = \frac{E_{zr}}{E_{zi}}$$

$$\tau_\parallel \equiv \frac{E_t}{E_i} = \frac{2\eta_2 \cos\theta_1}{\eta_2 \cos\theta_2 + \eta_1 \cos\theta_1}.$$

Brewster's angle:

For *diamagnetic* and *paramagnetic* materials which cover a vast amounts of media of interest in EM and optical applications, we have $\mu \approx \mu_o$. For TE and TM reflection problems between diamagnetic and/or paramagnetic materials it is therefore possible to take $\mu_2 = \mu_1$ and simplify the reflection and transmission coefficient formulae above.

- For the case $\mu_2 = \mu_1$, $\Gamma_\perp = 0$ iff $\eta_2 = \eta_1$, i.e, $\epsilon_2 = \epsilon_1$, but it is possible to have $\Gamma_\parallel = 0$ with $\eta_2 \neq \eta_1$ at a special angle θ_1 known as **Brewster's angle**, θ_p, examined in this section.

TE reflection:

$$\Gamma_\perp \equiv \frac{E_{yr}}{E_{yi}} = \frac{\eta_2 \cos\theta_1 - \eta_1 \cos\theta_2}{\eta_2 \cos\theta_1 + \eta_1 \cos\theta_2}$$

$$\tau_\perp \equiv \frac{E_{yt}}{E_{yi}} = \frac{2\eta_2 \cos\theta_1}{\eta_2 \cos\theta_1 + \eta_1 \cos\theta_2},$$

TM reflection:

$$\Gamma_\parallel \equiv -\frac{E_r}{E_i} = \frac{\eta_2 \cos\theta_2 - \eta_1 \cos\theta_1}{\eta_2 \cos\theta_2 + \eta_1 \cos\theta_1} = \frac{E_{zr}}{E_{zi}}$$

$$\tau_\parallel \equiv \frac{E_t}{E_i} = \frac{2\eta_2 \cos\theta_1}{\eta_2 \cos\theta_2 + \eta_1 \cos\theta_1}.$$

- Note that in view of Snell's law

$$\sqrt{\mu_2\epsilon_2}\sin\theta_2 = \sqrt{\mu_1\epsilon_1}\sin\theta_1$$

we have

$$\Gamma_\parallel = -\frac{E_r}{E_i} = \frac{\eta_2 \cos\theta_2 - \eta_1 \cos\theta_1}{\eta_2 \cos\theta_2 + \eta_1 \cos\theta_1} = 0$$

only when

$$\eta_1 \cos\theta_1 = \eta_2 \cos\theta_2 = \eta_2\sqrt{1-\sin^2\theta_2} = \eta_2\sqrt{1-\frac{\mu_1\epsilon_1}{\mu_2\epsilon_2}\sin^2\theta_1}.$$

Squaring this we get

$$\frac{\mu_1}{\epsilon_1}(1-\sin^2\theta_1) = \frac{\mu_2}{\epsilon_2}(1-\frac{\mu_1\epsilon_1}{\mu_2\epsilon_2}\sin^2\theta_1) \quad\Rightarrow\quad \sin^2\theta_1 = \frac{1-\frac{\mu_2}{\mu_1}\frac{\epsilon_1}{\epsilon_2}}{(1-\frac{\epsilon_1}{\epsilon_2})(1+\frac{\epsilon_1}{\epsilon_2})}.$$

- For $\mu_2 = \mu_1$, this yields

$$\theta_1 = \sin^{-1}\sqrt{\frac{1}{1+\frac{\epsilon_1}{\epsilon_2}}} \equiv \theta_p$$

which simplifies as the **Brewster's angle**

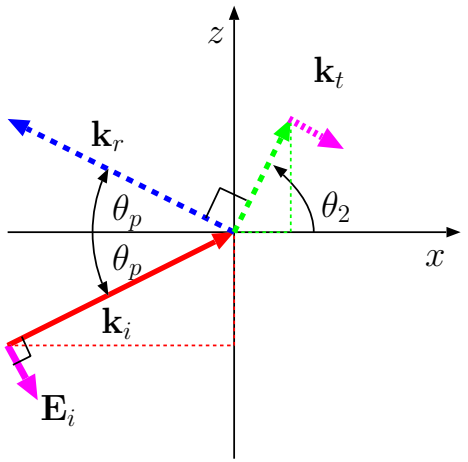
$$\theta_p = \sin^{-1}\sqrt{\frac{\epsilon_2}{\epsilon_2+\epsilon_1}} \quad\Rightarrow\quad \theta_p = \tan^{-1}\sqrt{\frac{\epsilon_2}{\epsilon_1}}.$$

- A physical insight to Brewster's angle θ_p can be gained by noting that (as verified below)

$$\theta_p + \theta_2 = 90^\circ,$$

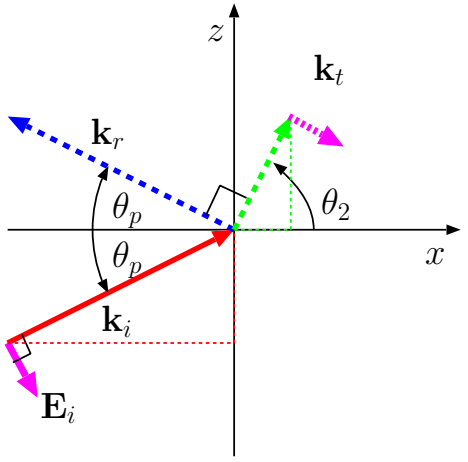
implying that vector $-\tilde{\mathbf{E}}_t$ in medium 2 is co-aligned with wavevector $\mathbf{k}_r$ of the zero-amplitude reflected wave in medium 1 as shown in the margin.

Geometry at Brewster's angle



$\epsilon_1 > \epsilon_2 \quad\Leftrightarrow\quad \theta_2 > \theta_1$

Geometry at Brewster's angle



$\epsilon_1 > \epsilon_2 \quad\Leftrightarrow\quad \theta_2 > \theta_1$

– Now, the physical cause of the plane-wave $\tilde{\mathbf{E}}_r$ is the superposition of dipole radiations of polarized molecules in medium 2 into medium 1, behaving like a giant 3D antenna array.

– However propagation direction $\mathbf{k}_r$ of wave $\tilde{\mathbf{E}}_r$ is the dipole axis of these molecules when $\theta_1 = \theta_p$, in which case radiation amplitude becomes zero because dipoles do not radiate along their axes as we have seen earlier on (they radiate best in the broadside direction)!

Verification of $\theta_p + \theta_2 = 90°$: For $\mu_2 = \mu_1$, Snell's law simplifies as

$$\sqrt{\epsilon_2}\sin\theta_2 = \sqrt{\epsilon_1}\sin\theta_1,$$

yielding

$$\sin\theta_2 = \sqrt{\frac{\epsilon_1}{\epsilon_2}}\sin\theta_1.$$

For $\theta_1 = \theta_p$, we have

$$\sin\theta_2 = \sqrt{\frac{\epsilon_1}{\epsilon_2}}\sin\theta_p = \sqrt{\frac{\epsilon_1}{\epsilon_2}}\sqrt{\frac{\epsilon_2}{\epsilon_2+\epsilon_1}} = \sqrt{\frac{\epsilon_1}{\epsilon_2+\epsilon_1}},$$

implying that

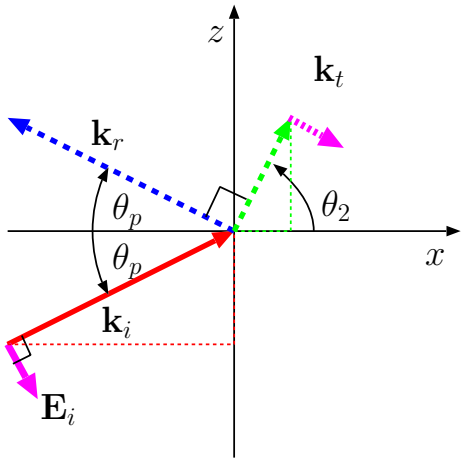
$$\sin^2\theta_2 = \frac{\epsilon_1}{\epsilon_2+\epsilon_1} = \frac{\epsilon_2+\epsilon_1-\epsilon_2}{\epsilon_2+\epsilon_1} = 1 - \frac{\epsilon_2}{\epsilon_2+\epsilon_1} = 1-\sin^2\theta_p = \cos^2\theta_p.$$

Thus

$$\sin\theta_2 = \cos\theta_p,$$

a telltale sign that θ_p and the corresponding θ_2 add up to 90°! (QED)

Geometry at Brewster's angle



$\epsilon_1 > \epsilon_2 \quad \Leftrightarrow \quad \theta_2 > \theta_1$

- Finally, there is no Brewster's angle for TE mode reflections because in the TE case $\mathbf{k}_r$ is unconditionally in the broadside direction of $\tilde{\mathbf{E}}_r$ polarized dipoles (in $\hat{y}$ direction). It is easy to see that for $\mu_2 = \mu_1$, $\Gamma_\perp = 0$ iff $\epsilon_2 = \epsilon_1$:

  Verification: According to Snell's law

  $$\sqrt{\epsilon_2}\sin\theta_2 = \sqrt{\epsilon_1}\sin\theta_1$$

  while

  $$\Gamma_\perp = \frac{E_{yr}}{E_{yi}} = \frac{\eta_2\cos\theta_1 - \eta_1\cos\theta_2}{\eta_2\cos\theta_1 + \eta_1\cos\theta_2} = 0$$

  only when

  $$\eta_1\cos\theta_2 = \eta_2\cos\theta_1.$$

  Dividing Snell's law with this relationship we get

  $$\frac{\sqrt{\epsilon_2}}{\sqrt{\mu_1/\epsilon_1}}\tan\theta_2 = \frac{\sqrt{\epsilon_1}}{\sqrt{\mu_1/\epsilon_2}}\tan\theta_1 \quad \Rightarrow \quad \tan\theta_2 = \tan\theta_1.$$

  But this condition of $\theta_2 = \theta_1$ is only permitted by Snell's law if $\eta_2 = \eta_1$, the trivial case of no practical interest.

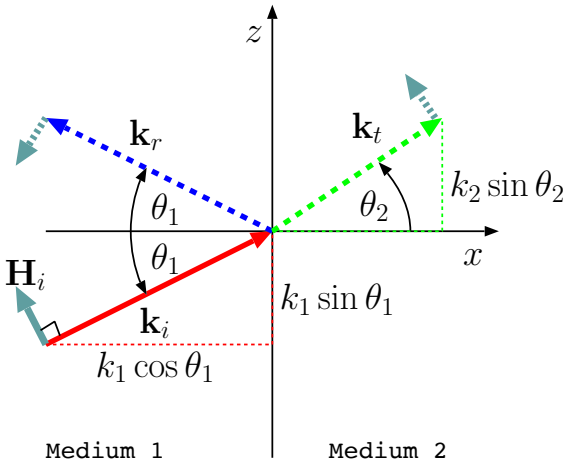
- Read pp 322-323 in Rao for a discussion of the applications of Brewster's angle.

- One simple application: reflected light from ground is typically TE polarized (parallel to the ground) because the TM component of light

reflects poorly because typically θ_1 may be close to θ_p. It is easy to eliminate TE polarized glare from the ground by using polarized eyeglasses which only transmit the TM component of light (polarized vertically). Note that this application also explains why the Brewster's angle θ_p is also known as "polarizing angle".

18 Reflecting plates, monopole antennas, corner reflectors

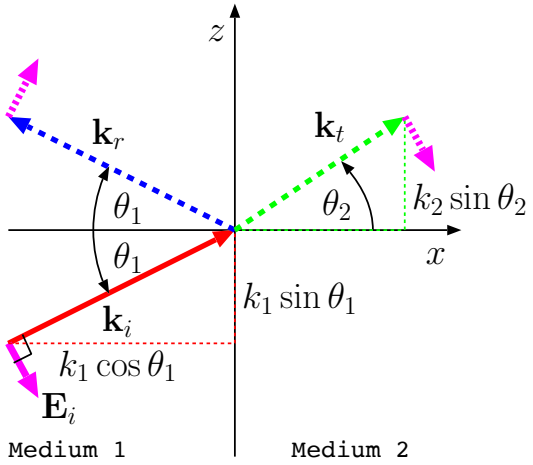
TE reflection:



$$\Gamma_\perp \equiv \frac{E_{yr}}{E_{yi}} = \frac{\eta_2 \cos\theta_1 - \eta_1 \cos\theta_2}{\eta_2 \cos\theta_1 + \eta_1 \cos\theta_2}$$

$$\tau_\perp \equiv \frac{E_{yt}}{E_{yi}} = \frac{2\eta_2 \cos\theta_1}{\eta_2 \cos\theta_1 + \eta_1 \cos\theta_2},$$

TM reflection:



$$\Gamma_\parallel \equiv -\frac{E_r}{E_i} = \frac{\eta_2 \cos\theta_2 - \eta_1 \cos\theta_1}{\eta_2 \cos\theta_2 + \eta_1 \cos\theta_1} = \frac{E_{zr}}{E_{zi}}$$

$$\tau_\parallel \equiv \frac{E_t}{E_i} = \frac{2\eta_2 \cos\theta_1}{\eta_2 \cos\theta_2 + \eta_1 \cos\theta_1}.$$

- In deriving the transmission and reflection rules for TE and TM modes summarized above we assumed lossless propagation media during the last two lectures.
- The equations can be easily modified as described next if either medium

1 or medium 2 or both have non-zero conductivities σ_1 and/or σ_2.

In general, in the case of a non-insulating medium with a finite conductivity σ, we expect a conduction current $\tilde{\mathbf{J}} = \sigma\tilde{\mathbf{E}}$, in which case the plane-wave form of Ampere's law can be cast as

$$\begin{aligned} -j\mathbf{k} \times \tilde{\mathbf{H}} &= \sigma\tilde{\mathbf{E}} + j\omega\epsilon\tilde{\mathbf{E}}, \\ &= j\omega(\epsilon + \frac{\sigma}{j\omega})\tilde{\mathbf{E}}. \end{aligned}$$

Since this equation differs from the non-conducting case only by having $\epsilon + \frac{\sigma}{j\omega}$ in place of ϵ, propagation parameters

$$k = \omega\sqrt{\mu\epsilon} \quad \text{and} \quad \eta = \sqrt{\frac{\mu}{\epsilon}}$$

of non-conducting media are modified as

$$k = \omega\sqrt{\mu(\epsilon + \frac{\sigma}{j\omega})} \quad \text{and} \quad \eta = \sqrt{\frac{\mu}{\epsilon + \frac{\sigma}{j\omega}}},$$

respectively, in homogeneous conducting media. In other words a conducting medium is treated as a dielectric with a permittivity $\epsilon + \frac{\sigma}{j\omega}$.

- Consider the wavenumber

$$k = \omega\sqrt{\mu(\epsilon + \frac{\sigma}{j\omega})}$$

in a medium with $\epsilon \gg \sigma/\omega$. In that case — poor conductor approximation — we can approximate k as

$$\begin{aligned} k &= \omega\sqrt{\mu(\epsilon - j\frac{\sigma}{\omega})} = \omega\sqrt{\mu\epsilon(1 - j\frac{\sigma}{\omega\epsilon})} \approx \omega\sqrt{\mu\epsilon}(1 - j\frac{\sigma}{2\omega\epsilon}) \\ &= \omega\sqrt{\mu\epsilon} - j\frac{1}{2}\sqrt{\frac{\mu}{\epsilon}}\sigma \equiv k' - jk'', \end{aligned}$$

with

$$k' \equiv \mathrm{Re}\{k\} \approx \omega\sqrt{\mu\epsilon} \quad \text{Propagation constant}$$

and

$$k'' \equiv -\mathrm{Im}\{k\} \approx \frac{1}{2}\sqrt{\frac{\mu}{\epsilon}}\sigma \quad \text{Attenuation constant.}$$

These terms are applicable since

$$e^{-j\mathbf{k}\cdot\mathbf{r}} = e^{-jks} = e^{-j(k'-k'')s} = e^{-k''s}e^{-jk's}$$

clearly signify an attenuating plane-wave field with distance s measured in the direction of a unit vector $\hat{k}$ such that $\mathbf{k}$ introduced above relates to $k = k' - jk''$ as in

$$\mathbf{k} = \hat{k}(k' - jk'').$$

- Conversely, in a medium where $\epsilon \ll \sigma/\omega$ — good conductor approximation — we can approximate k as

$$\begin{aligned} k &= \omega\sqrt{\mu(\epsilon - j\frac{\sigma}{\omega})} \approx \omega\sqrt{-j\frac{\mu\sigma}{\omega}} = \sqrt{-j}\sqrt{\omega\mu\sigma} \\ &= \frac{1-j}{\sqrt{2}}\sqrt{\omega\mu\sigma} = (1-j)\sqrt{\pi f\mu\sigma} \equiv k' - jk'' \end{aligned}$$

Clearly, in this case the **penetration depth** δ (recall ECE 329 — distance for the field to decay one e-fold) is

$$\delta = \frac{1}{k''} = \frac{1}{\sqrt{\pi f \mu \sigma}},$$

and this quantity vanishes in the limit $\sigma \to \infty$ — the meaning of this is, TEM waves cannot penetrate regions of perfect electrical conductors.

Example 1: Consider a plane wave of frequency $f = 400$ MHz propagating in a conductive medium with conductivity $\sigma = 4 \times 10^7$ S/m. Given that the wave phasor is

$$\tilde{\mathbf{E}}(x) = \hat{y} E_o e^{-jkx} = \hat{y} E_o e^{-k''x} e^{-jk'x}.$$

determine $\mathbf{E}(x,t)$ and $\mathbf{H}(x,t)$ and the penetration depth (skin depth) δ as well as the propagation velocity $v_p = \omega/k'$. Assume that $\mu = \mu_o$ and $\epsilon = \epsilon_o$ in the medium.

Solution: We first note that in this case

$$\frac{\sigma}{\omega} = \frac{4 \times 10^7}{2\pi \times 400 \times 10^6} = \frac{1}{20\pi} \gg \epsilon = \epsilon_o \approx \frac{1}{36\pi \times 10^9}.$$

Thus

$$k' = k'' \approx \sqrt{\pi f \mu \sigma} = \sqrt{\pi 400 \times 10^6 \times 4\pi \times 10^{-7} \times 4 \times 10^7} = 8\pi \times 10^4 \, \frac{\text{rad}}{\text{m}}.$$

Also,

$$\eta = \sqrt{\frac{\mu}{\epsilon + \frac{\sigma}{j\omega}}} \approx \sqrt{\frac{j\omega\mu}{\sigma}} = \sqrt{\frac{2\pi \times 400 \times 10^6 \times 4\pi \times 10^{-7}}{4 \times 10^7}} \angle 45^\circ = 2\pi\sqrt{2} \angle 45^\circ \, \text{m}\Omega.$$

Therefore, we have

$$\mathbf{E}(x,t) = \hat{y} E_o e^{-8\pi\times 10^4 x} \cos(8\pi \times 10^8 t - 8\pi \times 10^4 x) \, \frac{\mathrm{V}}{\mathrm{m}}$$

and

$$\mathbf{H}(x,t) = \hat{z} \frac{E_o e^{-8\pi\times 10^4 x}}{2\pi\sqrt{2}} \cos(8\pi \times 10^8 t - 8\pi \times 10^4 x - 45^\circ) \, \frac{\mathrm{kA}}{\mathrm{m}}.$$

The penetration depth is

$$\delta = \frac{1}{k''} = \frac{1}{8\pi \times 10^4} = \frac{1}{80\pi} \times 10^{-3} \, \mathrm{m},$$

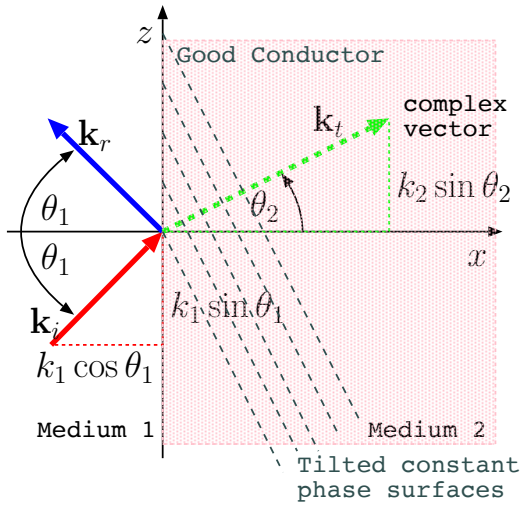
clearly a small fraction of a millimeter. Finally the propagation velocity is

$$v_p = \frac{\omega}{k'} = \frac{8\pi \times 10^8}{8\pi \times 10^4} = 10^4 \, \frac{\mathrm{m}}{\mathrm{s}}.$$

Example 2: A plane TEM wave propagating in air with a phasor

$$\tilde{\mathbf{E}}_i = \hat{y} e^{-j\mathbf{k_i}\cdot\mathbf{r}} = \hat{y} e^{-jk_1(\cos\theta_1 x + \sin\theta_1 z)}$$

is incident $x = 0$ plane at the planar boundary of a good conductor for which $\sigma \gg \omega\epsilon$ — see figure in the margin. Determine the transmitted field phasor $\tilde{\mathbf{E}}_t$.



Solution: Formally, using transmission TE coefficient $\tau_\perp$,

$$\tilde{\mathbf{E}}_t = \hat{y}\tau_\perp e^{-j\mathbf{k}_t\cdot\mathbf{r}} = \hat{y}\tau_\perp e^{-j(k_{2x}x + k_{2z}z)}$$

where, according to *Snell's law*,

$$\underbrace{k_{2z}}_{\equiv\, k_2 \sin\theta_2} = \underbrace{k_{1z}}_{k_1 \sin\theta_1}$$

with

$$k_{2x}^2 + k_{2z}^2 = \mathbf{k_t}\cdot\mathbf{k}_t = \omega^2\mu(\epsilon + \frac{\sigma}{j\omega}) \equiv k_2^2$$

and

$$k_{2x} = \sqrt{k_2^2 - k_{2z}^2} = \sqrt{k_2^2 - k_2^2\sin^2\theta_2} = k_2\sqrt{1-\sin^2\theta_2} \equiv k_2\cos\theta_2.$$

Since

$$k_2 = \sqrt{\omega^2\mu(\epsilon + \frac{\sigma}{j\omega})} \approx \sqrt{-j\mu\sigma\omega} = (1-j)\sqrt{\pi f\mu\sigma}$$

is complex valued

$$\sin\theta_2 \equiv \frac{k_1}{k_2}\sin\theta_1 \quad \text{and} \quad \cos\theta_2 \equiv \sqrt{1-\sin^2\theta_2}$$

are also complex valued and hence θ_2 is not the familiar real-valued angle variable confined to $[0, \pi/2]$ rad range. Postponing a discussion of this complex θ_2 (with its *defined* sine and cosine above), we can express the transmitted field as

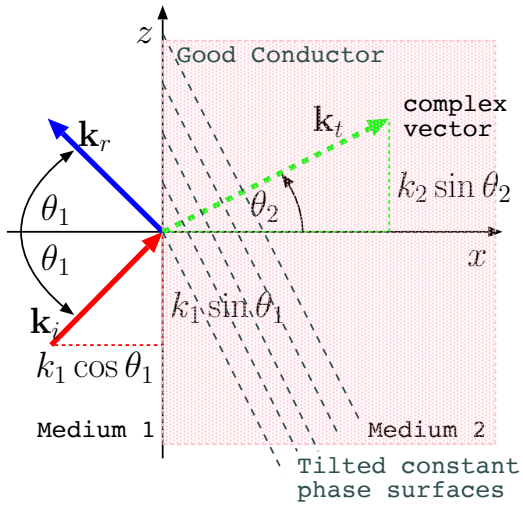
$$\tilde{\mathbf{E}}_t = \hat{y}\tau_\perp e^{-j(k_{2x}x+k_{2z}z)} \approx \hat{y}\tau_\perp e^{-jk_1 \sin\theta_1 z} e^{-j(1-j)\sqrt{\pi f\mu\sigma}x}$$

since $\cos\theta_2 \approx 1$ with $|\sin\theta_2| \ll 1$ when $|k_1/k_2| \ll 1$ with $k_1 = 2\pi f/c$ and $k_2 = (1-j)\sqrt{\pi f\mu\sigma}$ in the good conductor regime. This is a ***non-uniform plane wave*** since $\tilde{\mathbf{E}}_t$ is not a constant on planes of constant phase ($=k_1 \sin\theta_1 z + \sqrt{\pi f\mu\sigma}x$) as a consequence of the $e^{-\sqrt{\pi f\mu\sigma}x}$ factor.

The transmission coefficient $\tau_\perp$ can be computed using

$$\tau_\perp = 1 + \Gamma_\perp = 1 + \frac{\eta_2 \cos\theta_1 - \eta_1 \cos\theta_2}{\eta_2 \cos\theta_1 + \eta_1 \cos\theta_2} \approx \frac{2\eta_2 \cos\theta_1}{\eta_1 \cos\theta_2}$$

with $\cos\theta_2 \approx 1$ obtained for the good conducting case — since $|\eta_2| \ll \eta_1$ we have $|\tau_\perp| \approx |\frac{2\eta_2 \cos\theta_1}{\eta_1 \cos\theta_2}| \ll 1$ (and $\Gamma_\perp \approx -1$).



Note that $|\mathbf{k_t}| \gg |\mathbf{k}_i|$ and as such the diagram above is "not to scale" in a proper way!

The very small tilt angle of the constant phase planes of the transmitted non-uniform plane wave can be shown to be $k_{2z}/\mathrm{Re}\{k_{2x}\} = 2\mathrm{Re}\{\theta_2\} \ll \theta_1$, where $\theta_2 \approx \frac{k_1}{k_2}\sin\theta_1$.

The "complex angle" $\theta_2 = \theta_{2r} + j\theta_{2i}$ encountered in Ex 2 is the *analytic continuation* of real θ_2 from the *real line* into the *complex plane* obeying *Euler's identity*

$$e^{j\theta_2} = \cos\theta_2 + j\sin\theta_2$$

just like *real* θ_2. With $\cos\theta_2 \approx 1$ and $\sin\theta_2 = \frac{k_1}{k_2}\sin\theta_1$, we obtain, from Euler's identity,

$$e^{j\theta_2} \approx 1 + j\frac{k_1}{k_2}\sin\theta_1.$$

From the series expansion $e^{j\theta_2} = 1 + j\theta_2 + \mathcal{O}(\theta_2^2)$ it follows that complex

$$\theta_2 \approx \frac{k_1}{k_2}\sin\theta_1 \approx \frac{2\pi f/c}{(1-j)\sqrt{\pi f\mu\sigma}}\sin\theta_1 = (1+j)\frac{\pi f/c}{\sqrt{\pi f\mu\sigma}}\sin\theta_1,$$

having equal real and imaginary parts (in the good conducting case).

- Assume that medium 2 is a perfect electrical conductor (PEC), i.e., $\sigma_2 \to \infty$. In that case

$$\eta_2 = \sqrt{\frac{\mu}{\epsilon + \frac{\sigma}{j\omega}}} = 0 \text{ (PEC is like a "short"), and}$$

TE reflection:

$$\Gamma_\perp \equiv \frac{E_{yr}}{E_{yi}} = \frac{\eta_2 \cos\theta_1 - \eta_1 \cos\theta_2}{\eta_2 \cos\theta_1 + \eta_1 \cos\theta_2} \to -1$$

$$\tau_\perp \equiv \frac{E_{yt}}{E_{yi}} = \frac{2\eta_2 \cos\theta_1}{\eta_2 \cos\theta_1 + \eta_1 \cos\theta_2} \to 0$$

TM reflection:

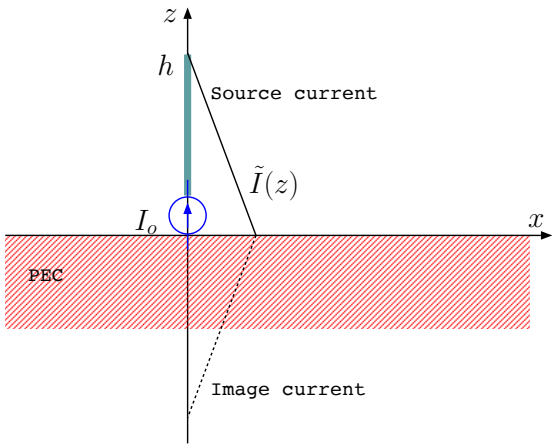
$$\Gamma_\parallel \equiv -\frac{E_r}{E_i} = \frac{\eta_2 \cos\theta_2 - \eta_1 \cos\theta_1}{\eta_2 \cos\theta_2 + \eta_1 \cos\theta_1} = \frac{E_{zr}}{E_{zi}} \to -1$$

$$\tau_\parallel \equiv \frac{E_t}{E_i} = \frac{2\eta_2 \cos\theta_1}{\eta_2 \cos\theta_2 + \eta_1 \cos\theta_1} \to 0.$$

Conclusion: all plane waves incident on a PEC boundary will reflect in such a way that $\tilde{\mathbf{E}}_{tan}$ at the bounding surface (this is the tangential component of the total field summed on the dielectric side of the boundary) is everywhere zero, whereas the accompanying $\tilde{\mathbf{H}}_{tan}$ at the bounding surface will imply a surface current $\tilde{\mathbf{J}}_s$ such that $|\tilde{\mathbf{H}}_{tan}|^2 = |\tilde{\mathbf{J}}_s|^2$.

- In "perturbation method" solutions *good conductors* are *treated as* PEC's to infer $|\tilde{\mathbf{H}}_{tan}|^2 = |\tilde{\mathbf{J}}_s|^2$ and then the *power loss per unit area* of the good conductor is subsequently calculated as the Poynting flux on the conductor side surface — works out to be $\frac{1}{2}|\tilde{\mathbf{H}}_{tan}|^2 \mathrm{Re}\{\eta_2\}$ with $\mathrm{Re}\{\eta_2\} \approx \sqrt{\frac{\pi f \mu}{\sigma}}$ referred to as *surface resistance*!

- We will next examine the consequences of this conclusion on antennas placed near conducting planes.



Monopole above reflecting surface:

- Consider a a straight wire of a length h "end-fed" by an independent current source $\tilde{I}(0) = I_o$ connected as shown in the margin between the wire and an infinite ground plane. For $h \ll \lambda = \frac{2\pi}{k}$ we may assume a current distribution

$$\tilde{I}(z) = I_o \triangle(\frac{z}{2h}) u(z)$$

that drops linearly from I_o to 0 across the length of the wire.

We next construct the radiation field of this so-called *vertical monopole* antenna by postulating that the monopole will radiate, into the half-space $z > 0$, like a short-dipole of a length $L = 2h$ having a triangular current distribution

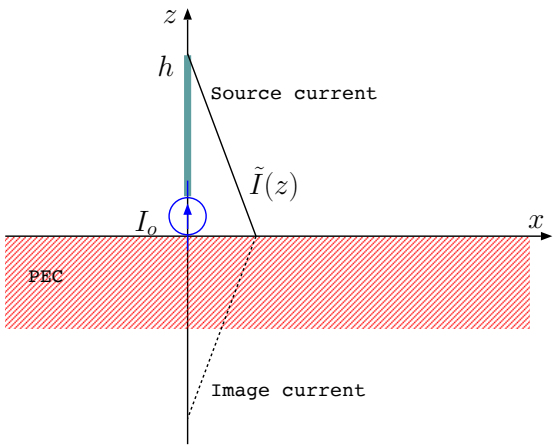
$$\tilde{I}_d(z) = I_o \triangle(\frac{z}{2h})$$

that matches $\tilde{I}(z)$ for $z > 0$ and is considered to be a "image" current of $\tilde{I}(z)$ for $z < 0$.

We justify the postulate by observing that:

1. the field generated by $\tilde{I}_d(z)$ is in $\hat{\theta} = -\hat{z}$ direction on $z = 0$ surface, and therefore it satisfies the boundary condition of having zero tangential $\tilde{\mathbf{E}}$ on the perfectly conducting surface,

2. the field generated by $\tilde{I}_d(z)$ also satisfies the boundary condition of having zero tangential $\tilde{\mathbf{E}}$ on the perfectly conducting surface of the h-long monopole wire, since the wire is just the upper half of a two-wire dipole of length $L = 2h$ (on which the condition is satisfied *ipso facto*),
3. Maxwell's equations (ME) necessarily have a *unique* solution for each possible configuration of boundary conditions (BC) — *a* solution of ME's matching the given BC is *the* solution!



In view of above, the radiation field of the monopole is

$$\tilde{\mathbf{E}} = \begin{cases} j\eta_o I_o k \frac{L}{2} \sin\theta \frac{e^{-jkr}}{4\pi r}\hat{\theta} & \text{for } z > 0 \\ 0 & \text{for } z < 0, \end{cases}$$

where $\frac{L}{2}$ should be replaced by h. As this result indicates, a monopole radiates its entire power to one hemisphere as opposed to a dipole in free space radiating equally into two hemispheres.

- In the above description of the radiation of the monopole, the bottom half of the current distribution $\tilde{I}_d(z)$ is said to be the "image" of the source current $\tilde{I}(z)$ on the monopole. The image is really "imaginary" in the sense that the "real", actual, sources of the radiated field $\tilde{\mathbf{E}}$ in the upper hemisphere are

1. $\tilde{I}(z)$ on the monopole, and

2. a surface current $\tilde{\mathbf{J}}_s(x,y)$ induced on $z=0$ surface in order to satisfy the BC $\hat{z}\times\tilde{\mathbf{H}}(x,y,0)=\tilde{\mathbf{J}}_s(x,y)$.

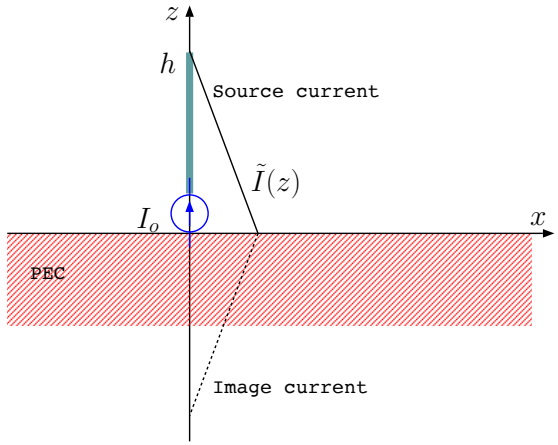
- A *short* monopole of length $h \ll \lambda$ radiates half as much power as a short-dipole of length $L=2h$ having an equal input current I_o. Therefore R_{rad} for monopole is half of

$$R_{rad}=20\pi^2(\frac{L}{\lambda})^2=20\pi^2(\frac{2h}{\lambda})^2$$

of the dipole, i.e.,

$$R_{rad,mono}=10\pi^2(\frac{2h}{\lambda})^2=40\pi^2(\frac{h}{\lambda})^2$$

in ohms.



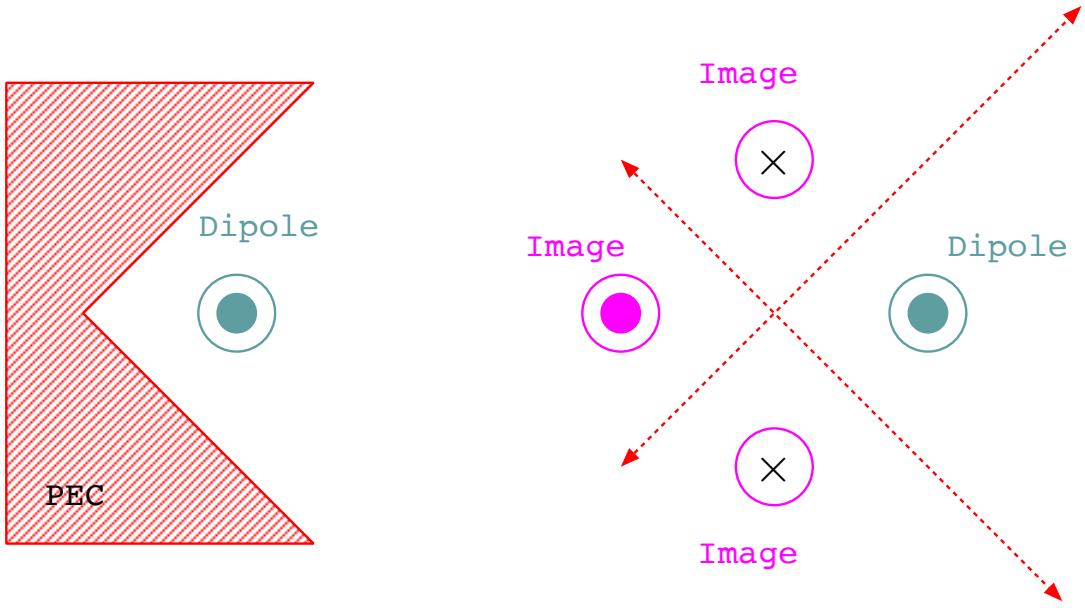
- Directivity of monopole is twice the directivity of short-dipole, i.e., $D=3$, since the beam solid angle Ω_o of the monopole is half the solid angle of the dipole (why?).

- Finally, a monopole of length $h=\frac{\lambda}{4}$ is called a *quarter-wave monopole*.

  - In analogy with a half-wave dipole, the quarter-wave monopole has a radiation resistance of about 36 ohms and a directivity of 3.28.

Corner reflector antenna

- The following diagram depicts a "corner reflector" antenna on the left, and its image based model as a 4-element array.



(a) Corner reflector (b) Image based model

  - Note that the image elements in the model have been so selected that the 4-element array has tangential electric field nulls along the conducting walls of the corner reflector placed next to the $\hat{z}$-polarized dipole antenna (seen from the top) shown on the left.
  - The field of the corner reflector antenna matches the field of the 4-element array in the rightmost quadrant in the diagram bounded by the diagonal lines. The field can be calculated easily by using an array factor that depends on the distance of the dipole from the reflecting corner (see HW).

19 Total internal reflection (TIR) and evanescent waves

- Consider a TE- or TM-polarized wave (or a superposition) incident on an interface at $x = 0$ surface as depicted in the margin at an incidence angle θ_1.

- Independent of the polarization of the incident wave, the angle of transmitted wave θ_2 can be found using Snell's law

$$k_1 \sin\theta_1 = k_2 \sin\theta_2 \quad \Rightarrow \quad \sqrt{\mu_{1r}\epsilon_{1r}}\sin\theta_1 = \sqrt{\mu_{2r}\epsilon_{2r}}\sin\theta_2$$

  assuming lossless media on either side of the interface, where

$$\mu_r \equiv \frac{\mu}{\mu_o} \text{ and } \epsilon_r \equiv \frac{\epsilon}{\epsilon_o}$$

  are the relative permeability and permittivity, respectively, of the propagation media. Moreover,

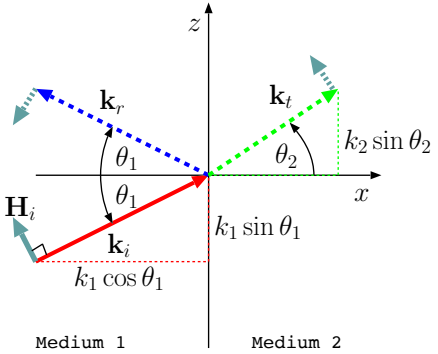
$$\sqrt{\mu_r\epsilon_r} = \frac{\sqrt{\mu\epsilon}}{\sqrt{\mu_o\epsilon_o}} = \frac{c}{v_p} \equiv n$$

  above can be referred to as the **refractive index** of the propagation medium.

- Snell's law, expressed in terms of refractive index,

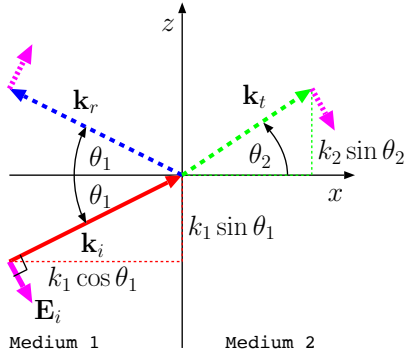
$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \quad \Rightarrow \quad \sin\theta_2 = \frac{n_1}{n_2}\sin\theta_1$$

TE reflection:



$$\frac{E_{yr}}{E_{yi}} = \frac{\eta_2\cos\theta_1 - \eta_1\cos\theta_2}{\eta_2\cos\theta_1 + \eta_1\cos\theta_2}$$
$$\frac{E_{yt}}{E_{yi}} = \frac{2\eta_2\cos\theta_1}{\eta_2\cos\theta_1 + \eta_1\cos\theta_2}$$

TM reflection:



$$-\frac{E_r}{E_i} = \frac{\eta_2\cos\theta_2 - \eta_1\cos\theta_1}{\eta_2\cos\theta_2 + \eta_1\cos\theta_1}$$
$$\frac{E_t}{E_i} = \frac{2\eta_2\cos\theta_1}{\eta_2\cos\theta_2 + \eta_1\cos\theta_1}.$$

Refractive index:

$$n = \frac{c}{v_p} = \sqrt{\mu_r\epsilon_r}$$

shows that for a given θ_1, the corresponding $\sin\theta_2$ can be in excess of 1 when $n_1 > n_2$, that is, for propagation from a high refractive index (optically thick) material such as glass into a lower refractive index (optically thin) material such as air.

– ***For example***: if $\frac{n_1}{n_2} = 1.5$ and $\theta_1 = 45°$, then

$$\sin\theta_2 = \frac{n_1}{n_2}\sin\theta_1 = 1.5\sin 45° = \frac{1.5}{\sqrt{2}} \approx \frac{1.5}{1.41} > 1.$$

But, $\sin\theta_2$ in excess of 1 cannot be solved for θ_2 as if it were a "regular" angle[1] describing the elevation of vector $\mathbf{k}_t$ above the x-axis.

- In general when $n_1 > n_2$ and the incidence angle

$$\theta_1 > \sin^{-1}\frac{n_2}{n_1} = \sin^{-1}\sqrt{\frac{\mu_2\epsilon_2}{\mu_1\epsilon_1}} \equiv \theta_c$$

we will have $\sin\theta_2$ in excess of 1 and $\cos\theta_2 = \sqrt{1-\sin^2\theta_2}$ purely imaginary.

Critical angle:

$$\theta_c = \sin^{-1}\frac{n_2}{n_1}$$

– in such situations use $\sin\theta_2$ and $\cos\theta_2 = \sqrt{1-\sin^2\theta_2}$ directly in the expressions for $\mathbf{k}_t$, Γ, and τ as illustrated below.

[1] Nor if $\sin\theta_2$ is complex valued because medium 2 is lossy and we need to use $\epsilon_{2r} = \frac{\epsilon_2}{\epsilon_o} + \frac{\sigma_2}{j\omega\epsilon_o}$ in Snell's law (as we already did in Lecture 18).

– as we will see the situation corresponds to having a **total internal reflection (TIR)** in medium 1 and establishing an **evanescent wave** (a special form of non-uniform plane wave with an imaginary valued k_{2x}) in medium 2.

– ***For example***: for $\frac{n_1}{n_2} = 1.5$ we have

$$\theta_c = \sin^{-1}\frac{n_2}{n_1} = \sin^{-1}\frac{1}{1.5} \approx 41.81^\circ,$$

which is less than $\theta_1 = 45^\circ$ which is why we find $\sin\theta_2 > 1$ in the above example (see margin for an example plot of this configuration in the context of a glass prism with $n = 1.5$).

Total internal reflection (TIR)

and

evanescent wave

- To understand the field topologies for $\theta_1 > \theta_c = \sin^{-1}\frac{n_2}{n_1}$ let us examine the reflected and transmitted field phasors for, say, the TE-polarization as θ_1 approaches and then exceeds θ_c. We will simplify this exercise by taking $\mu_1 = \mu_2 = \mu_o$ so that the refractive index

$$n_{1,2} = \sqrt{\epsilon_{r1,2}}$$

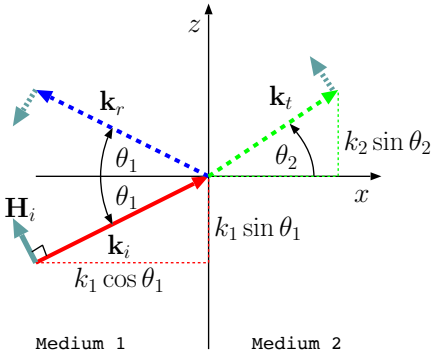
in Snell's law, and so that the reflection and transmission coefficients for the TE-mode shown in the margin can be expressed as

$$\Gamma_\perp = \frac{E_{yr}}{E_{yi}} = \frac{\eta_2\cos\theta_1 - \eta_1\cos\theta_2}{\eta_2\cos\theta_1 + \eta_1\cos\theta_2} \quad \Rightarrow \quad \frac{n_1\cos\theta_1 - n_2\cos\theta_2}{n_1\cos\theta_1 + n_2\cos\theta_2}$$

and

$$\tau_\perp = 1 + \Gamma_\perp = \frac{2n_1\cos\theta_1}{n_1\cos\theta_1 + n_2\cos\theta_2}.$$

TE reflection:



$$\frac{E_{yr}}{E_{yi}} = \frac{\eta_2\cos\theta_1 - \eta_1\cos\theta_2}{\eta_2\cos\theta_1 + \eta_1\cos\theta_2}$$

$$\frac{E_{yt}}{E_{yi}} = \frac{2\eta_2\cos\theta_1}{\eta_2\cos\theta_1 + \eta_1\cos\theta_2}$$

We have, using Snell's law

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \quad \Rightarrow \quad \sin\theta_2 = \frac{n_1}{n_2}\sin\theta_1,$$

and, therefore,

$$\cos\theta_2 = \sqrt{1-\sin^2\theta_2} = \sqrt{1-\frac{n_1^2}{n_2^2}\sin^2\theta_1}$$

in the coefficients above.

- For

$$\theta_1 \geq \theta_c = \sin^{-1}\frac{n_2}{n_1} \quad \Leftrightarrow \quad \sin^2\theta_1 \geq \frac{n_2^2}{n_1^2},$$

we have a ***purely imaginary***

$$\cos\theta_2 = \sqrt{1-\frac{n_1^2}{n_2^2}\sin^2\theta_1} = \pm j\alpha, \text{ with } \alpha = \sqrt{\frac{n_1^2}{n_2^2}\sin^2\theta_1 - 1},$$

in which case

$$\Gamma_\perp = \frac{n_1\cos\theta_1 - n_2\cos\theta_2}{n_1\cos\theta_1 + n_2\cos\theta_2} = \frac{n_1\cos\theta_1 \mp jn_2\alpha}{n_1\cos\theta_1 \pm jn_2\alpha} = 1\angle \mp 2\tan^{-1}(\frac{n_2\alpha}{n_1\cos\theta_1}).$$

and

$$\tau_\perp = 1+\Gamma_\perp = 1+1\angle \mp 2\tan^{-1}(\frac{n_2\alpha}{n_1\cos\theta_1}).$$

Note that

1. $|\Gamma_\perp| = |E_{yr}/E_{yi}| = 1$ **at all** $\theta_1 \geq \theta_c$**, a condition known as** *total internal reflection* **(TIR).**

2. $\tau_\perp = E_{yt}/E_{yi} \neq 0$ **in general and therefore a non-zero transmitted field exists in medium 2 despite TIR — this field in medium 2 has** *evanescent wave* **character described below.**

   - We can express the transmitted TE-mode field phasor for $\theta_1 \geq \theta_c$ as

$$\tilde{\mathbf{E}}_t = \hat{y} E_{yt} e^{-j\mathbf{k}_t \cdot \mathbf{r}} = \hat{y} E_{yt} e^{-jk_2(\cos\theta_2 x + \sin\theta_2 z)}$$

   where

$$E_{yt} = E_{yi}\left(1 + \frac{n_1 \cos\theta_1 \mp j n_2 \alpha}{n_1 \cos\theta_1 \pm j n_2 \alpha}\right),$$

$$k_2 \sin\theta_2 = k_1 \sin\theta_1 \quad \text{(Snell's law)},$$

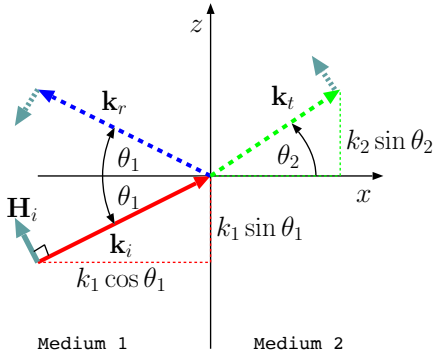
   and

$$k_2 \cos\theta_2 = k_2(\pm j\alpha) = \pm j k_2 \alpha, \quad \text{where} \quad \alpha = \sqrt{\frac{n_1^2}{n_2^2}\sin^2\theta_1 - 1}.$$

   Thus,

$$\begin{aligned} \tilde{\mathbf{E}}_t &= \hat{y} E_{yi}\left(1 + \frac{n_1 \cos\theta_1 \mp j n_2 \alpha}{n_1 \cos\theta_1 \pm j n_2 \alpha}\right) e^{-jk_1 \sin\theta_1 z}\, e^{-j(\pm j k_2 \alpha) x} \\ &= \hat{y} E_{yi}\left(1 + \frac{n_1 \cos\theta_1 \mp j n_2 \alpha}{n_1 \cos\theta_1 \pm j n_2 \alpha}\right) e^{-jk_1 \sin\theta_1 z}\, e^{\pm k_2 \alpha x}. \end{aligned}$$

TE reflection:



$$\frac{E_{yr}}{E_{yi}} = \frac{\eta_2 \cos\theta_1 - \eta_1 \cos\theta_2}{\eta_2 \cos\theta_1 + \eta_1 \cos\theta_2}$$

$$\frac{E_{yt}}{E_{yi}} = \frac{2\eta_2 \cos\theta_1}{\eta_2 \cos\theta_1 + \eta_1 \cos\theta_2}$$

- Depending on which root we select — + or − — we have two candidate solutions for medium 2 satisfying the plane-wave form of Maxwell's equations.

  - In geneal $\tilde{\mathbf{E}}_t$ should be a weighted superposition of these two solutions with weights determined by boundary conditions. In particular the solution that blows up as $x \to \infty$ can be assigned zero weight in problems where the x-extent of the prapagating region is unbounded, leaving us with an *evanescent* wave

    $$\tilde{\mathbf{E}}_t = \hat{y} E_{yi}(1 + \Gamma_\perp) e^{-jk_1 \sin\theta_1 z} e^{-k_2 \alpha x}$$

    in the region $x > 0$, with

    $$\Gamma_\perp = \frac{n_1 \cos\theta_1 + j n_2 \alpha}{n_1 \cos\theta_1 - j n_2 \alpha} \text{ and } \alpha = \sqrt{\frac{n_1^2}{n_2^2} \sin^2\theta_1 - 1},$$

    when the total internal reflection condition

    $$\theta_1 \geq \theta_c = \sin^{-1} \frac{n_2}{n_1}$$

    holds.

  - This solution should fit (as we are about to see) the plane-wave form of Maxwell's equations with

    $$\mathbf{k} = \mathbf{k}_t = k_2(\hat{x} \cos\theta_2 + \hat{z} \sin\theta_2) = -jk_2 \alpha \hat{x} + k_1 \sin\theta_1 \hat{z}$$

    which

1. is perpendicular to $\tilde{\mathbf{E}}_t \propto \hat{y}$ as required — i.e., $\mathbf{k}_t \cdot \tilde{\mathbf{E}}_t = 0$ (Gauss's law),
2. implies a *complex valued* unit vector

$$\hat{k} = \frac{\mathbf{k}_t}{k_2} = -j\alpha\hat{x} + \frac{k_1}{k_2}\sin\theta_1\hat{z}$$

which satisfies

$$\hat{k}\cdot\hat{k} = -\alpha^2 + \frac{k_1^2}{k_2^2}\sin^2\theta_1 = -(\frac{n_1^2}{n_2^2}\sin^2\theta_1 - 1) + \frac{n_1^2}{n_2^2}\sin^2\theta_1 = 1$$

as required, and

3. implies a transmitted magnetic field intensity phasor

$$\tilde{\mathbf{H}}_t = \frac{\hat{k}\times\tilde{\mathbf{E}}_t}{\eta_2} = \frac{E_{yi}}{\eta_2}(1+\Gamma_\perp)e^{-jk_1\sin\theta_1 z}e^{-k_2\alpha x}(-j\alpha\hat{z} - \frac{k_1}{k_2}\sin\theta_1\hat{x})$$

which is of course transverse to $\hat{k}$ *ipso facto*.

– It remains to show that

$$\hat{x}\cdot\langle\mathbf{E}_t\times\mathbf{H}_t\rangle = 0$$

so that

$$\hat{x}\cdot\langle\mathbf{E}_i\times\mathbf{H}_i\rangle + \hat{x}\cdot\langle\mathbf{E}_r\times\mathbf{H}_r\rangle = 0$$

— demanded by TIR — is satisfied.

Verification:

$$\begin{aligned}\langle \mathbf{E}_t \times \mathbf{H}_t \rangle &= \frac{1}{2}\mathrm{Re}\{\tilde{\mathbf{E}}_t \times \tilde{\mathbf{H}}_t^*\} \\ &= \frac{|E_{yi}|^2|\tau_\perp|^2}{2\eta_2} e^{-2k_2\alpha x}\mathrm{Re}\{\hat{y} \times (j\alpha\hat{z} - \frac{k_1}{k_2}\sin\theta_1\hat{x})\} \\ &= \frac{|E_{yi}|^2|\tau_\perp|^2}{2\eta_2} e^{-2k_2\alpha x}\frac{k_1}{k_2}\sin\theta_1\hat{z}.\end{aligned}$$

Evidently, timed averaged power flux is directed along the interface and has no component along $\hat{x}$ normal to the reflecting interface.

Example 1: Consider a uniform plane wave propagating in quartz with $\epsilon_r = 2.25$ and $n = \sqrt{\epsilon_r} = 1.5$ incident on a quartz/air interface at an incidence angle of 45°. Determine the evanescent field phasor $\tilde{\mathbf{E}}_t$ established in air outside the quartz slab. Assume that the incident wave is TE polarized and the wavelength is 1 mm within the quartz.

Solution: With $\theta_1 = 45°$ and $n_1 = 1.5$, $n_2 = 1$, we have $\cos\theta_2 = -j\alpha$ with

$$\alpha = \sqrt{\frac{n_1^2}{n_2^2}\sin^2\theta_1 - 1} = \sqrt{\frac{2.25}{1}\sin^2 45° - 1} = \sqrt{\frac{2.25}{2} - 1} = \sqrt{\frac{9}{8} - 1} = \frac{1}{2\sqrt{2}}.$$

Hence,

$$\Gamma_\perp = \frac{n_1\cos\theta_1 + jn_2\alpha}{n_1\cos\theta_1 - jn_2\alpha} = \frac{\frac{3}{2}\frac{1}{\sqrt{2}} + j1\frac{1}{2\sqrt{2}}}{\frac{3}{2}\frac{1}{\sqrt{2}} + -j1\frac{1}{2\sqrt{2}}} = \frac{3 + j1}{3 - j1} = \frac{8 + j6}{10} = 0.8 + j0.6$$

and

$$\tau_\perp = 1 + \Gamma_\perp = 1.8 + j0.6.$$

Also, since $\lambda_1 = 1$ mm, it follows that

$$k_1 = \frac{2\pi}{\lambda_1} = 2\pi \frac{\text{rad}}{\text{mm}} = k_o n_1 = k_o 1.5, \quad k_o = \frac{k_1}{1.5} = \frac{4\pi}{3} \frac{\text{rad}}{\text{mm}},$$

where $k_o \equiv \omega/c$ is the free space wavenumber also applicable in air (i.e., k_2). Therefore, we have

$$\tilde{\mathbf{E}}_t = \hat{y} E_{yi} (1 + \Gamma_\perp) e^{-jk_1 \sin\theta_1 z} e^{-k_2 \alpha x}$$

$$\hat{y} E_{yi} (1.8 + j0.6) e^{-j\frac{2\pi}{\sqrt{2}} z} e^{-\frac{2\pi}{3\sqrt{2}} x} \frac{\text{V}}{\text{m}}$$

where x and z are used in mm units. This is an example of an evanescent field that "hugs" the the quartz/air interface on the air side.

- **To summarize:** TIR that occurs when $\theta_1 \geq \theta_c$ is accompanied by an evanescent transmitted wave.

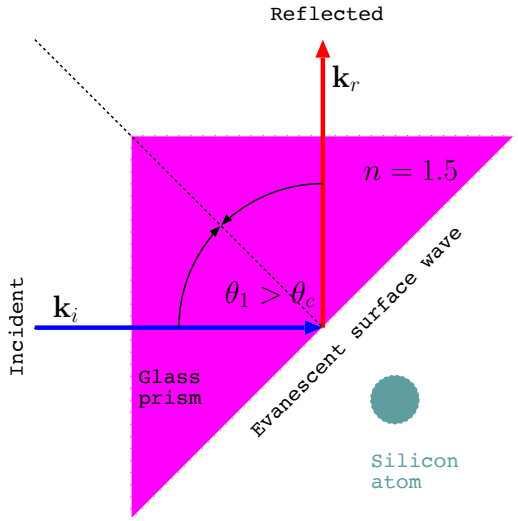
- **The evanescent wave:**

  1. has a decaying amplitude with distance away from the reflecting interface (that is along x) and carries no average power away from the interface,
  2. it exhibits a phase variation along the interface (that is along z) that matches the phase variations of the incident and reflected waves in medium 1,

An evanescent wave is a *non-uniform TEM wave* since the field vector in non-uniform on surfaces of constant phase.

3. it carries average power only along the interface and only close to the interface because of the $e^{-2k_2\alpha x}$ factor — thus it is also known as a **surface wave**,
4. it can be perturbed by introducing some new materials into region 2 to start drawing energy towards medium 2 — this is the topic of frustrated TIR to be examined next.

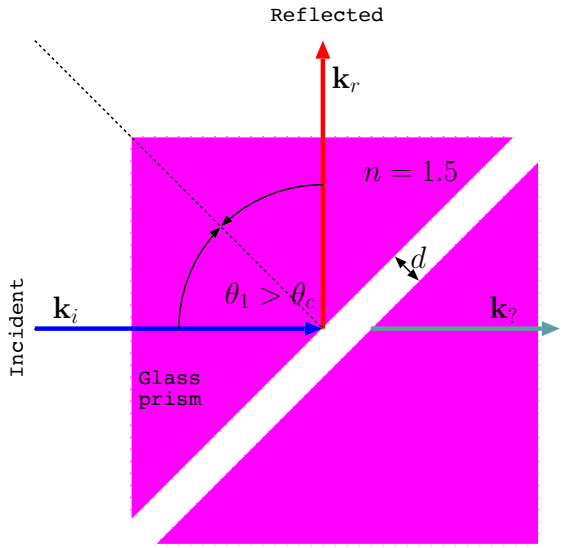
Frustrated TIR and tunneling:



- Suppose a silicon atom is brought to a location right next the prism as shown in the margin where an evanescent wave is present.
  - What happens then to the evanescent and total internal reflected waves to either sides of the diagonal face of the prism?
  - What happens when the atom is replaced by another prism placed, as shown in the second diagram in the margin, at a distance d away from the diagonal face?

- In the first instance, the silicon atom will be stretched into a polarized dipole by the action of the time-varying evanescent electric field outside the prism and therefore it will radiate like an oscillating Hertzian dipole antenna at the frequency of the evanescent wave.

- The radiation field of the atomic dipole will then superpose on the evanescent and internally reflected fields, modifying them both, and enabling the extraction of power from the incident wave to be transported away from the TIR interface.
  - This is an elementary example of what is known as energy "tunneling".



- The tunneling phenomenon becomes more pronounced when the atom is replaced with a second prism (a whole array of silicon atoms mixed

with oxygen atoms) as illustrated in the margin.

- The described phenomenon is known as "frustrated" TIR, because the presence of the second prism will perturb the reflected wave substantially when the gap width d between the prisms is a small fraction of a wavelength λ.
- The double prism arrangement shown in the margin can be used as a "practical" beam splitter at optical frequencies by adjusting d/λ.
- A quantitative treatment of the tunneling problem will be presented in Lecture 24 in a multiple slab geometry involving evanescent regions.

20 Reciprocity and receiving antennas

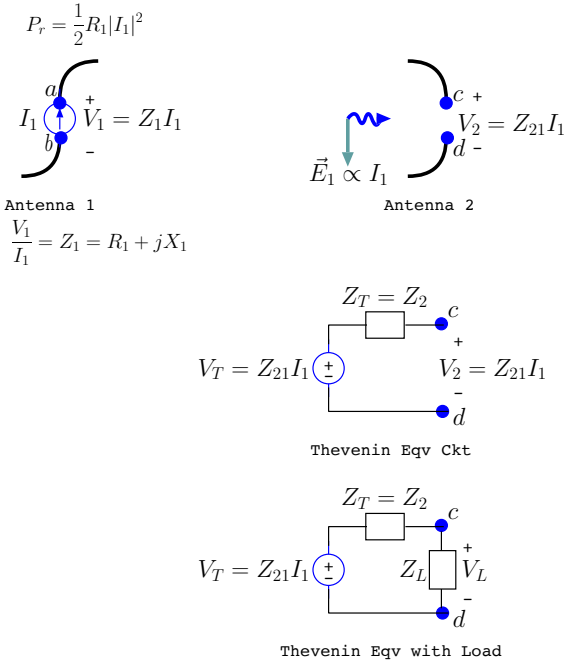
- Antennas exhibit a *reciprocal behavior* in their properties pertinent to the "transmission" and "reception" of electromagnetic waves.

- In the diagram in the margin Antenna 1 on the top left, with an *input current* phasor $\tilde{I}_1$ (applied between its terminals indicated by a and b) and response voltage $\tilde{V}_1 = Z_1\tilde{I}_1$ is "transmitting" — that is, it produces a radiated field indicated as $\tilde{\mathbf{E}}_1 \propto \tilde{I}_1$, which is shown to be incident on Antenna 2 to induce an *open circuit voltage* $\tilde{V}_2 = Z_{21}\tilde{I}_1$ between terminals c and d of Antenna 2.

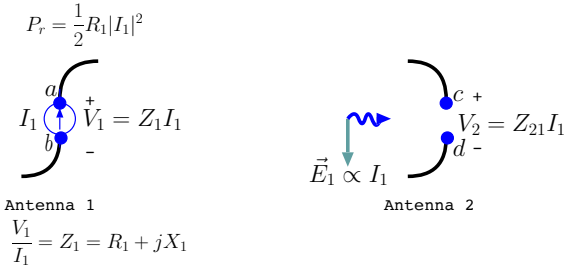
  - *Input voltage* $\tilde{V}_1$ of Antenna 1 and *output voltage* $\tilde{V}_2$ of Antenna 2 are obtained from the *input current* $\tilde{I}_1$ by using the *transmission impedance* $Z_1 = R_1 + jX_1$ of Antenna 1 and the *coupling impedance* Z_{21} of Antenna 2 to Antenna 1, respectively.
  - As an example, for a *half-wave dipole* used as Antenna 1 we would have $Z_1 = 73 + j0\,\Omega$, and Z_{21} would depend on the type, orientation, and the distance of Antenna 2 coupling to the transmitting half-wave dipole considered in this example.

- The open circuit voltage $\tilde{V}_2 = Z_{21}\tilde{I}_1$ of Antenna 2 can be used to construct the *Thevevin equivalent circuit* of the coupled antenna system shown in the second row in the margin where

  - the *Thevenin voltage* $\tilde{V}_T = Z_{21}\tilde{I}_1$,



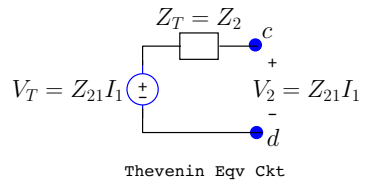
– the *Thevenin impedance* Z_T equals the *transmission impedance* Z_2 of Antenna 2 to produce a response voltage $Z_2\tilde{I}_2$ between terminals c and d when Antenna 2 is used for transmission purposes — this can be confirmed by using the **test current method**[1] from ECE 210 to determine Z_T with $\tilde{I}_{test} = \tilde{I}_2$.

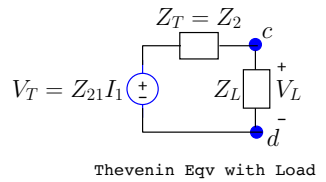


- The Thevenin equivalent ckt of the receiving antenna is useful to calculate the power that can be delivered by the receiving antenna to a load Z_L connected across its terminals c and d, as shown in the bottom row in the margin. Note, in particular, that if $Z_L = Z_2^*$, the average power transferred to Z_L can be calculated as the *available power*

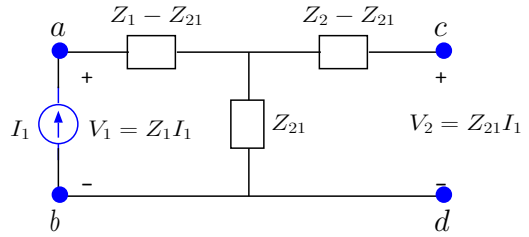
$$P_L = \frac{|\tilde{V}_T|^2}{8\text{Re}\{Z_T\}} = \frac{|Z_{21}\tilde{I}_1|^2}{8\text{Re}\{Z_2^*\}} = \frac{|Z_{21}|^2|\tilde{I}_1|^2}{8R_2}$$

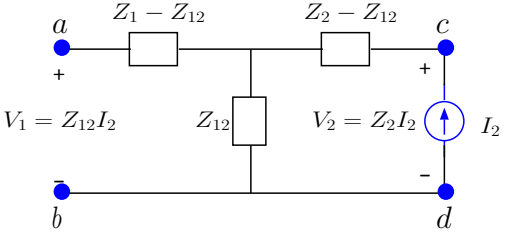
using the *maximum power transfer theorem* from ECE 210.





- The Thevenin equivalent as well as the current-voltage relations discussed above imply a two-port circuit model shown in the margin in the top row including the input source current $\tilde{I}_1$ of Antenna 1:





  – You should examine this T-network topology to confirm all the current-voltage relations from above.
  – Notice that $P_t = \frac{1}{2}\text{Re}\{\tilde{V}_1\tilde{I}_1^*\} = \frac{1}{2}|\tilde{I}_1|^2 R_1$ represents the transmitted power of Antenna 1, while $P_r = \frac{|Z_{21}|^2|\tilde{I}_1|^2}{8R_2}$ is the time-average power

[1] Suppress all the independent sources in the circuit and apply $\tilde{I}_{test}$ at its terminals to measure a $\tilde{V}_{test} \equiv Z_T\tilde{I}_{test}$ response in order to determine Z_T.

delivered by Antenna 2 to its conjugate matched load $Z_L = Z_2^* = R_2 - jX_2$, leading to the received-to-transmitted power ratio

$$\frac{P_r}{P_t} = \frac{|Z_{21}|^2}{4R_1R_2}.$$

- Reversing the transmission and receptions roles of Antennas 1 and 2 we next obtain another two-port model shown in the margin in the second row.

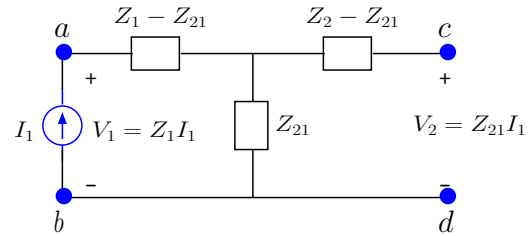
  – In this case we have

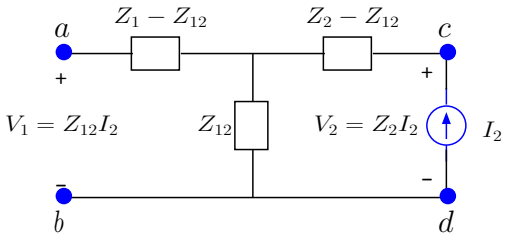
  $$\frac{P_r}{P_t} = \frac{|Z_{12}|^2}{4R_1R_2}$$

  which would be identical with our earlier result if it were the case that $Z_{12} = Z_{21}$.

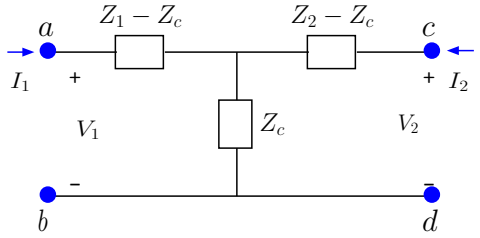
  – Certainly $Z_{12} = Z_{21}$ if/when Antennas 1 and 2 are identical, e.g., a pair of half-wavelength dipoles.

  – **It turns out that $Z_{12} = Z_{21}$ is true even when the antennas are of different types** so long as they are embedded in an *isotropic* propagation medium — this can be demonstrated by actually computing separately and comparing Z_{12} and Z_{21} in specific cases of interest, as we will be doing later in this section. A more general proof can also be furnished as detailed in Lecture 21.





- With $Z_{12} = Z_{21} \equiv Z_c$ we finally obtain our reciprocal two-port model for the coupling of arbitrary antennas as shown in the margin, with the

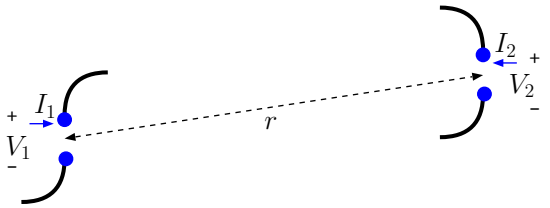


implication that

$$\frac{P_r}{P_t} = \frac{|Z_c|^2}{4R_1R_2},$$

independent of which antenna is used to transmit and which one to deliver maximum available power to a conjugate matched load. This "power reciprocity" property of antennas will be next used to learn how to determine the open circuit voltages and maximum powers of receiving antennas.

- Consider a pair of arbitrary but **co-polarized**[2] antennas separated over some distance r as depicted in the margin.



  - Let's first assume[3] that $I_2 = 0$ so that Antenna 2 can act as a receiving antenna with some Thevenin voltage $V_T = V_2 \propto I_1$, the input current of Antenna 1 in a transmission role.
  - Antenna 1 transmits an average power of $P_t = \frac{1}{2}R_1|I_1|^2$ and produces a Poynting flux of

$$S_{inc} = \frac{P_t}{4\pi r^2}G_1$$

incident on antenna 2 where G_1 denotes the transmission gain of Antenna 1 in the direction of Antenna 2. We know about antenna

[2] *Co-polarized* means that the antennas produce *identically polarized* radiation fields propagating towards one another — e.g., identically weighted mixes of TE and TM polarizations (wrt to any plane containing vector **r** separating the antennas) as well as RC or LC polarized radiation fields.

[3] From here on we will drop "tildes" used in our notation to indicate current and voltage phasors to simplify notation.

gain functions G from our study of transmission antennas.

- For instance $G = \frac{3}{2}\sin^2\theta$ for a z-polarized short dipole antenna.

– Now let P_r denote the average power delivered by Antenna 2 to its **impedance matched load** $Z_L = Z_2^*$ and define the ratio

$$\frac{P_r}{S_{inc}} \equiv A_2$$

as the **effective area** of Antenna 2 f**or reception** so that we can write $P_r = S_{inc}A_2$.

- By combining the expressions for P_r and S_{inc} we notice that

$$\frac{P_r}{P_t} = \frac{G_1 A_2}{4\pi r^2}$$

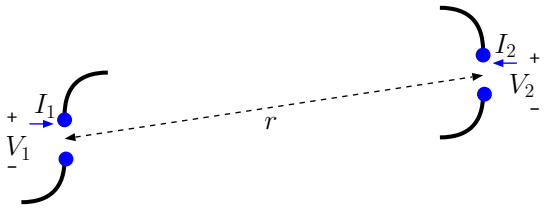
and likewise we can obtain

$$\frac{P_r}{P_t} = \frac{G_2 A_1}{4\pi r^2}$$

for Antenna 2 transmitting towards Antenna 1 with a gain of G_2 and Antenna 1 extracting power P_r from S_{inc} with an effective area of A_1.

- Since $\frac{P_r}{P_t}$ should be identical between Antennas 1 and 2 independent of which one is used for transmission and which one for reception, as we learned above, it follows that

$$G_1 A_2 = G_2 A_1 \quad \Rightarrow \quad \frac{A_2}{G_2} = \frac{A_1}{G_1}.$$



- Since we reached the above conclusion with no restrictions on the types of antennas considered in our discussion, it follows that *all* antennas, independent of their design details, will have to have identical effective area to transmission gain ratios A/G. **If we can compute A/G of any single antenna by any means that result will be valid and usable for all types of antennas.**
  - We will shortly show that

$$\frac{A}{G} = \frac{\lambda^2}{4\pi}$$

    for a **short dipole** antenna and then use this for **any type of antenna to relate its effective area A to its transmission gain G.**

Example 1: Recall that the gain of a z-polarized short dipole antenna is

$$G = \frac{3}{2}\sin^2\theta.$$

Therefore corresponding the effective area must be

$$A = \frac{\lambda^2}{4\pi}G = \frac{3\lambda^2}{8\pi}\sin^2\theta,$$

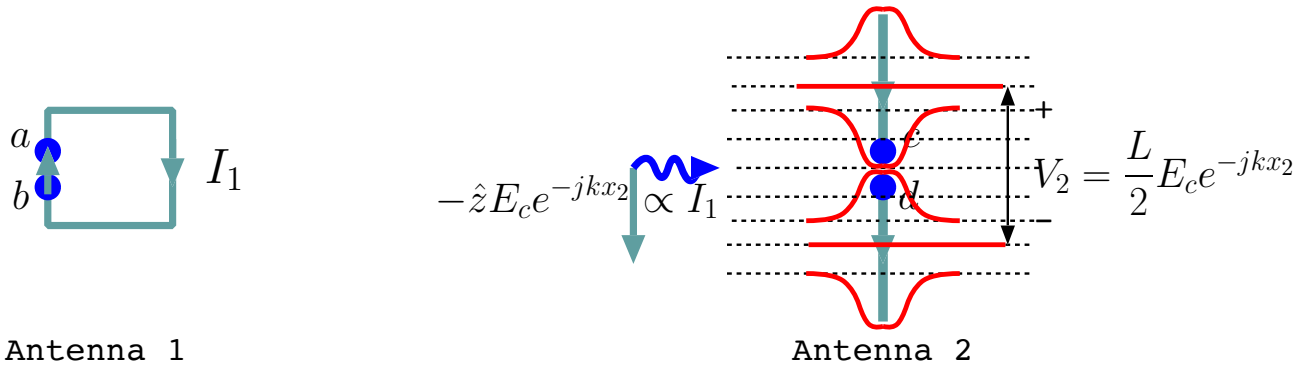
maximizing at a value of $A_{max} = \frac{3\lambda^2}{8\pi}$ for a zenith angle of $\theta = 90°$.

Consequently, a short dipole antenna exposed to a plane wave with wavelength $\lambda = 1$ m and Poynting Flux of $S_{inc} = 8\pi$ W/m^2 arriving from $\theta = 90^\circ$ direction (from the broadside direction of the dipole) will deliver an average power of

$$P_r = S_{inc} A = 8\pi \times \frac{3\lambda^2}{8\pi} \sin^2\theta = 3\,\mathrm{W}$$

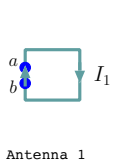
to a matched load connected across its terminals.

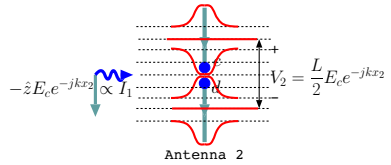
- To show that $A/G = \lambda^2/4\pi$ for a short dipole antenna, consider the response of a short dipole to an incident and *co*-polarized wave field arriving from, say, a loop antenna, as shown in the diagram below:



  - The loop antenna launches the co-polarized wave field $\hat{\theta}E_c e^{-jkx}$ incident on the z-polarized dipole on the right. The length L of the short-dipole is much shorter than the wavelength λ so that over a region of size L surrounding the dipole located at $x = x_2 \gg L$ the field phasor amplitude $E_c e^{-jkx_2}$ can be regarded spatially non-varying and represented in terms of a quasi-static electric potential $V(z)$ linear in z with a slope $\frac{dV}{dz} = E_c e^{-jkx_2}$.

- The equipotentials representing this linear $V(z)$ in the absence of the dipole are shown by dashed lines in the figure, whereas distorted equipotentials caused by the presence of the dipole are sketched by solid red curves. The distortions are such as to render the two wires of the dipole equipotentials sitting at distinct potentials $V_c = V(L/4)$ and $V_d = V(-L/4)$, respectively, as shown in the diagram, so that





$$V_2 = V_c - V_d = \frac{L}{2} E_c e^{-jkx_2}.$$

– This particular derivation of the antenna open circuit voltage in terms of incident co-polarized field $E_c e^{-jkx_2}$ at the antenna location and the antenna length L is only valid for $L \ll \lambda$, i.e., for short dipole antennas, and would not be applicable for, say, half-wave dipole for which $L = \frac{\lambda}{2}$.

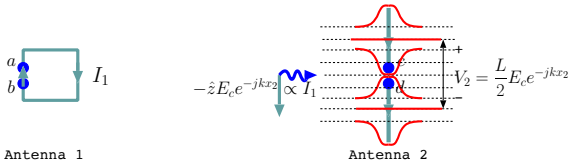
- Still the result allows a quick derivation of $A_2/G_2 = \lambda^2/4\pi$ for a *short dipole in free space* — which is all we need to use $P_r = S_{inc}A$ with all antennas everywhere. Here is the derivation:

$$P_r = \frac{|V_2|^2}{8R_2} = \frac{|\frac{L}{2}E_c e^{-jkx_2}|^2}{8 \times \underbrace{20\pi^2(\frac{L}{\lambda})^2}_{R_2}} = \underbrace{\frac{|E_c|^2}{240\pi}}_{S_{inc}} \overbrace{\underbrace{\frac{\lambda^2}{4\pi}}_{} \underbrace{\frac{3}{2}}_{G_2}}^{A_2}.$$

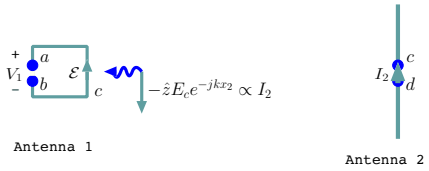
- Since $V_2 = Z_{21} I_1 = \frac{L}{2} E_c e^{-jkx_2}$ and the radiation field of a small loop with area $\Delta x \Delta z$ is $E_c e^{-jkx_2} = \eta_o k^2 \Delta x \Delta z I_1 \frac{e^{-jkx_2}}{4\pi x_2}$ at a distance x_2, it follows that

$$Z_{21} = \frac{V_2}{I_1} = \frac{L}{2} \frac{E_c e^{-jkx_2}}{I_1} = \frac{L}{2} \eta_o k^2 \Delta x \Delta z \frac{e^{-jkx_2}}{4\pi x_2},$$

the coupling impedance from the loop to the dipole.



- Let's next calculate $Z_{12} = V_1/I_2$ by having the short dipole radiate towards the loop antenna centered at the origin — see the margin. The incident field on the receiving loop is once again $-\hat{z}E_c e^{-jkx_2}$ which is accompanied by an incident $\mathbf{H} = -\hat{y}\frac{E_c}{\eta_o} e^{-jkx_2}$ to produce a *magnetic flux* of $\Phi = \mu_o \frac{E_c}{\eta_o} e^{-jkx_2} \Delta x \Delta z$ *linking* the loop and a counterclockwise **emf** of $\mathcal{E} \equiv \oint_c \mathbf{E} \cdot d\mathbf{l} = -j\omega\Phi$ so that an ideal voltmeter connected from a to b would measure (see Example 6 in Lecture 14 of ECE 329 notes) an open circuit voltage[4] drop $V_1 = \mathcal{E} = -j\omega\mu_o \frac{E_c}{\eta_o} e^{-jkx_2} \Delta x \Delta z = V_T$; since $E_c e^{-jkx_2} = j I_2 \eta_o k \frac{L}{2} \frac{e^{-jkx_2}}{4\pi x_2}$ it turns out that



$$Z_{12} = \frac{V_1}{I_2} = -j\omega\mu_o \frac{E_c e^{-jkx_2}}{\eta_o I_2} \Delta x \Delta z = \frac{L}{2} \eta_o k^2 \Delta x \Delta z \frac{e^{-jkx_2}}{4\pi x_2} = Z_{21} = Z_c,$$

showing that the coupling between the loop and the dipole is reciprocal as claimed. A general and rigorous proof of the reciprocity of the coupling impedances of arbitrary antennas is provided in Lecture 21.

[4] Calculating the short circuit current I_N of the loop antenna would require the inclusion of a *back*-emf $-j\omega L_1 I_N$ in addition to the *coupled* emf $V_T = -j\omega\Phi \propto I_2$ to express the *total* emf $R_1 I_N$ of the loop to yield $I_N = V_T/Z_1$, $Z_1 = R_1 + j\omega L_1$, where for a lossless loop R_1 stands for its radiation resistance.

Example 2: Consider a $\hat{z}$-polarized half-wave dipole antenna with an input impedance of $Z_2 = 73\,\Omega$ located at the origin in free space. Determine the time-average power received by the antenna if it has a matched termination and the incident field has an electric field phasor

$$\tilde{\mathbf{E}}_c = 120\pi\hat{z}e^{j2\pi x}\,\mathrm{V/m}.$$

Solution: The incident field has a wave vector

$$\mathbf{k} = -2\pi\hat{x}\,\mathrm{rad/m},$$

indicating the antenna sees the incident wave coming from direction $\theta = 90^\circ$ with a polarization direction $\hat{\theta}$. Since

$$\lambda = \frac{2\pi}{k} = \frac{2\pi}{|\mathbf{k}|} = \frac{2\pi}{2\pi} = 1\,\mathrm{m}$$

and

$$A(\theta,\phi) \;=\; \frac{\lambda^2}{4\pi}G(90^\circ,\phi) = \frac{1}{4\pi}\underbrace{(1.64)}_{\textbf{half-wave-dipole gain at } 90^\circ} = \frac{1.64}{4\pi}\,\mathbf{m}^2,$$

and, furthermore,

$$S_{inc} = \frac{|\tilde{\mathbf{E}}_c|^2}{2\eta_o} = \frac{(120\pi)^2}{2\times 120\pi} = 60\pi\,\mathrm{W/m}^2,$$

it follows that

$$P_r = S_{inc}A(\theta,\phi) = 60\pi\times\frac{1.64}{4\pi} = 24.6\,\mathrm{W}$$

is the time-averaged power received by the antenna with the matched termination.

- Suppose Antenna 2 has been terminated by some $Z_L \neq Z_2^*$. In that case power delivered to Z_L will be less than $P_r = S_{inc}A_2$ and can be computed using the Thevenin ckt as follows:

$$I_L = \frac{V_T}{Z_2 + Z_L} \;\Rightarrow\; V_L = I_L Z_L = \frac{V_T Z_L}{Z_2 + Z_L}$$

and

$$\begin{aligned} P_L &= \frac{1}{2}\mathrm{Re}\{V_L I_L^*\} = \frac{|V_T|^2 \mathrm{Re}\{Z_L\}}{2|Z_2 + Z_L|^2} = \underbrace{\frac{|V_T|^2}{8\mathrm{Re}\{Z_2\}}}_{P_r \text{ with matched load}} \frac{4\mathrm{Re}\{Z_2\}\mathrm{Re}\{Z_L\}}{|Z_2 + Z_L|^2} \\ &= P_r \frac{4\mathrm{Re}\{Z_2\}\mathrm{Re}\{Z_L\}}{|Z_2 + Z_L|^2} < P_r. \end{aligned}$$

Example 3: Repeat Example 2 if the antenna is terminated by $Z_L = 36.5\,\Omega$ by first determining the antenna open circuit voltage magnitude $|V_T|$.

Solution: In this case power received will be less than the available power

$$P_r = S_{inc}A(\theta, \phi) = 60\pi \times \frac{1.64}{4\pi} = 24.6\,\mathrm{W}$$

of the antenna obtained in Example 2. Substituting for P_r,Z_2, and Z_L, we find

$$P_L = P_r \frac{4\mathrm{Re}\{Z_2\}\mathrm{Re}\{Z_L\}}{|Z_2 + Z_L|^2} = 24.6 \frac{4 \times 73 \times 36.5}{|73 + 36.5|^2} = 24.6 \frac{4 \times 2}{9} = 21.87\,\mathrm{W}.$$

Alternatively, we obtain, using $P_r = 24.6$ W, $Z_2 = 73\,\Omega$, and $P_r = \frac{|V_T|^2}{8\mathrm{Re}\{Z_2\}}$,

$$|V_T| = \sqrt{8\mathrm{Re}\{Z_2\}P_r} = 119.86\,\mathrm{V}.$$

Hence

$$I_L = \frac{119.86e^{j\phi}}{73+36.5} \;\Rightarrow\; V_L = I_L Z_L = 36.5\frac{119.86e^{j\phi}}{73+36.5},$$

where $\phi = \angle V_T$ (it's value is not needed to calculate P_L), and

$$P_L = \frac{1}{2}\mathrm{Re}\{V_L I_L^*\} = \frac{1}{2}36.5\frac{119.86e^{j\phi}}{73+36.5}\frac{119.86e^{-j\phi}}{73+36.5} = \frac{1}{2}36.5(\frac{119.86}{73+36.5})^2 = 21.87\,\mathrm{W}.$$

- We have learned how to calculate the power delivered by an antenna to its terminating load, whether matched or not. We have not however outlined how to calculate the open circuit voltage of an arbitrary shaped antenna (except for its magnitude — see Example 3) — we want to address this next:

  - Notice that $V_o = \frac{L}{2}E_c$ denoting the open circuit voltage of any short dipole antenna of length L exposed to a co-polarized incident field E_c arriving from the broadside direction, is a particular realization of

    $$V_o = \ell_{eff}(\theta,\phi)\sin\theta E_c,$$

    the product of the *foreshortened* antenna effective length and co-polarized field amplitude incident from a direction θ. The proof that this is in fact the proper generalization of $V_o = \frac{L}{2}E_c$ for arbi-

trary dipole antennas follows from

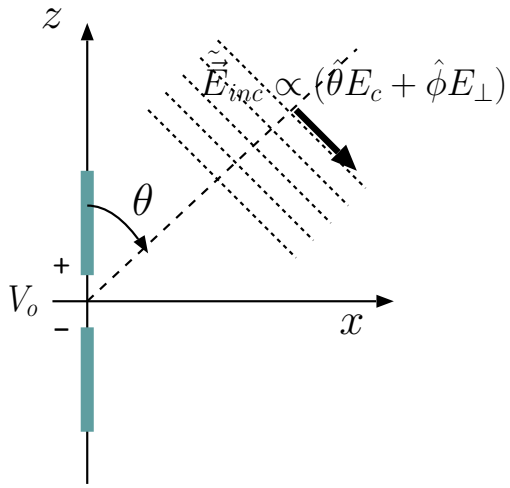
$$P_r = \frac{|V_o|^2}{8R} = \frac{|E_c\,\ell_{eff}(\theta,\phi)\sin\theta|^2}{8\frac{\eta_o}{4\lambda^2}\int d\Omega|\ell_{eff}(\theta,\phi)\sin\theta|^2} = \underbrace{\frac{|E_c|^2}{2\eta_o}}_{S_{inc}}\overbrace{\frac{\lambda^2}{4\pi}\underbrace{\frac{4\pi|\ell_{eff}(\theta,\phi)\sin\theta|^2}{\int d\Omega|\ell_{eff}(\theta,\phi)\sin\theta|^2}}_{G(\theta,\phi)}}^{A_{eff}(\theta,\phi)}$$

where we used the already known expressions (from earlier lectures) for the radiation resistance R and gain G of arbitrary z-polarized dipole antennas with known effective length functions $\ell_{eff}(\theta,\phi)$.

- The formula

$$V_o = \ell_{eff}(\theta,\phi)\sin\theta E_c$$

represents the open circuit voltage response of a receiving z-polarized dipole to a plane wave with co-pol amplitude E_c arriving from θ direction. A further generalization of this formula for an arbitrary shaped and polarized antenna responding to an arbitrary incident wave field (planar, spherical, evanescent, etc.) is given by a path integral formula



$$V_o = -\int_L d\mathbf{l}\cdot\mathbf{E}(\mathbf{r})\frac{I(\mathbf{r})}{I_o}$$

where $\frac{I(\mathbf{r})}{I_o}$ describes the current distribution of the antenna along the path L it occupies. This result, representing a weighted sum of the

projections of the field $\mathbf{E}(\mathbf{r})$ along the antenna body, is derived and further discussed in Lecture 21.

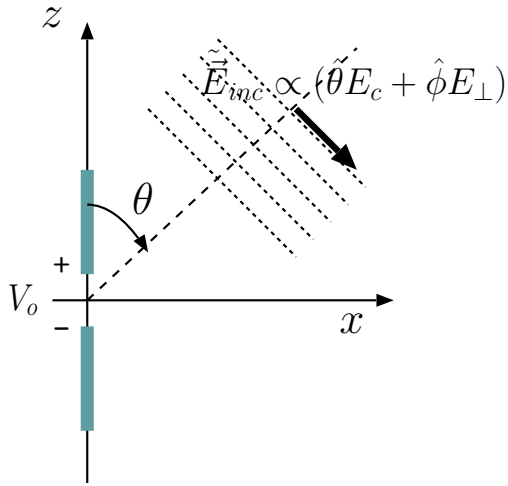
– Using this more general formula for the open circuit voltage V_o together with with an incident plane wave field

$$\tilde{\mathbf{E}}(\mathbf{r}) = (\hat{\theta}E_c + \hat{\phi}E_\perp)e^{j\mathbf{k}\cdot\mathbf{r}}$$

and a z-polarized dipole geometry, such that $d\mathbf{l} = \hat{z}dz$ in the integral formula between $z = -L/2$ and $z = L/2$, we recover, as expected,

$$V_o = -\int_L \frac{I(\mathbf{r})}{I_o} d\mathbf{l}\cdot\mathbf{E}(\mathbf{r}) = -\int_{-L/2}^{L/2} \frac{I(z)}{I_o} dz\, \hat{z}\cdot(\hat{\theta}E_c + \hat{\phi}E_\perp)e^{jk_z z}$$
$$= E_c \sin\theta \int_{-L/2}^{L/2} dz \frac{I(z)}{I_o} e^{jk\cos\theta\, z} = E_c \sin\theta\, \ell_{eff}(\theta,\phi)$$

since $\hat{z}\cdot\hat{\phi} = 0$ and $\hat{z}\cdot\hat{\theta} = -\sin\theta$ and we also recognize the weighted line integral of the antenna current distribution as the *antenna effective length* $\ell_{eff}(\theta,\phi)$ from our earlier lectures.



A z-pol dipole responding to an incoming plane wave $\tilde{\mathbf{E}}_{inc}$ with an open circuit voltage $V_o \propto E_c$.

- The results obtained in this lecture will be sufficient to solve the antenna reception problems assigned in the homework. More general/rigorous/formal derivations of the same results furnished in Lecture 21 will not be needed in the homework ...

21 Antenna reception — formal framework

- Interactions between pairs of antennas 1 and 2 separated by some distance r across an arbitrary but isotropic propagation medium obey a *reciprocity relation*

$$V_1 I_1 = V_2 I_2$$

indicated in the margin — this relation, which holds when the two antennas are excited by input currents I_1 and I_2 one at a time, will be derived starting with Maxwell's equation later in this lecture[1].

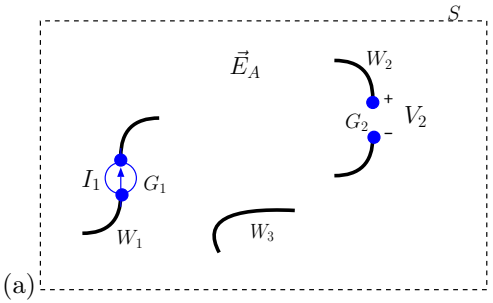
  - Let's take $V_1 I_1 = V_2 I_2$ for now as an experimental finding and notice that it implies

$$\frac{V_1}{I_2} = \frac{V_2}{I_1} \equiv Z_c,$$

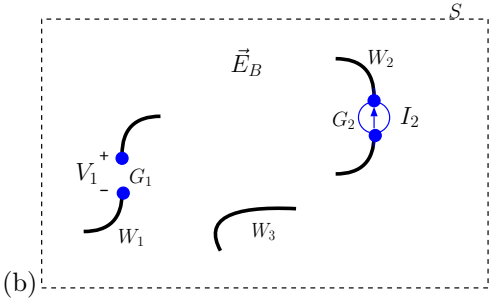
where Z_c stands for the *coupling impedance* of the two interacting antennas.

  - Given Z_c, an equivalent circuit model of the interactions of the antennas can be expressed as shown in the margin using the two-port voltage current relations

$$\begin{aligned} V_1 &= Z_1 I_1 + Z_c I_2 \\ V_2 &= Z_c I_1 + Z_2 I_2 \end{aligned}$$



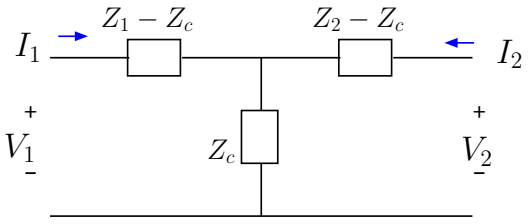
(a) Antenna 1 with input current I_1 causes an open circuit voltage response V_2 at antenna 2 terminals across some distance r (due to field $\mathbf{E}_A$).



(b) Antenna 2 with input current I_2 causes an open circuit voltage response V_1 at antenna 1 terminals across some distance r (due to field $\mathbf{E}_B$).

W_3 represents passive reflectors that may exist in the vicinity of the antennas causing some "scattered" fields. Antenna response voltages $V_{1,2}$ and input currents $I_{1,2}$ satisfy a *reciprocity relation*

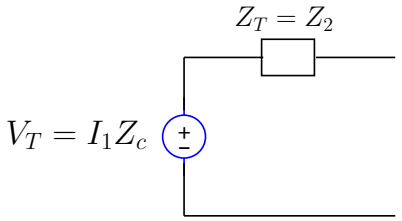
$$V_1 I_1 = V_2 I_2$$



Reciprocal circuit model for antenna coupling

[1] This is a *dense* lecture providing rigorous justifications of the results obtained in Lecture 20 — should be left as *optional reading* for interested students.

- Here V_1 is the terminal voltage of antenna 1, equal to $Z_c I_2$ coupled from antenna 2 if/when antenna 1 terminals are open and hence $I_1 = 0$.
- Likewise V_2 is the terminal voltage of antenna 2, equal to $Z_c I_1$ coupled from antenna 1 if/when antenna 2 terminals are open and hence $I_2 = 0$.
- Z_1 and Z_2 represent the *self-impedances* of the two antennas used to express the antenna input voltages $Z_1 I_1$ and $Z_2 I_2$ when the antennas are used in *transmission* mode. We can call Z_1 and Z_2 as transmission impedances.



Thevenin equivalent circuit of antenna 2 in reception

- The open circuit voltage $Z_c I_1$ of antenna 2 can be regarded as the Thevenin voltage V_T of a **Thevenin equivalent circuit of antenna 2 in reception** as shown in the margin, where Thevenin impedance $Z_T = Z_2$ is obtained from the coupling network using source the suppression method (deactivating the source I_1 by replacing it with an open).

- From the Thevenin equivalent we identify the average power transferred by I_1 to a possible matched termination of antenna 2 as the power received

$$P_r = \frac{|V_T|^2}{8\mathrm{Re}\{Z_T\}} = \frac{|I_1|^2 |Z_c|^2}{8\mathrm{Re}\{Z_2\}}.$$

- To calculate the transmitted power P_t of antenna 1 when antenna 2 is terminated by matched $Z_L = Z_2^*$ (see margin), we first note that

$$V_1 \approx I_1 Z_1$$

if $|Z_c| \ll |Z_{1,2}|$, in which case antenna 1 has a power input

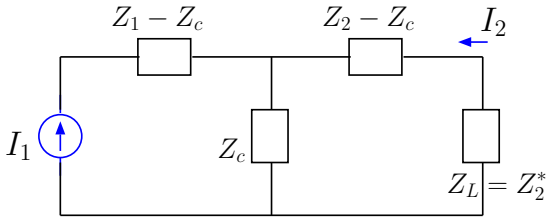
$$P_t = \frac{1}{2}\mathrm{Re}\{V_1 I_1^*\} \approx \frac{1}{2}|I_1|^2 \mathrm{Re}\{Z_1\}.$$

- The above expressions can be combined as

$$\frac{P_r}{P_t} \approx \frac{|Z_c|^2}{4\mathrm{Re}\{Z_1\}\mathrm{Re}\{Z_2\}},$$

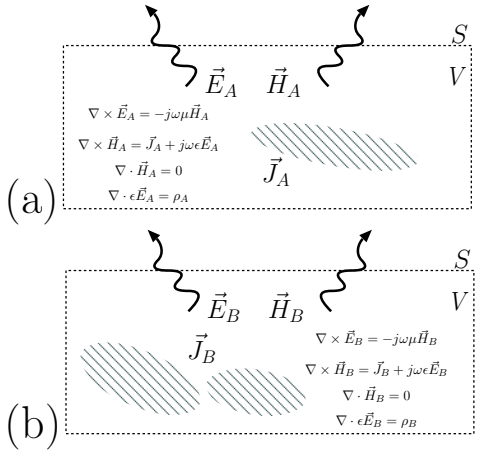
which is a *reciprocal* relation symmetric in Z_1 and Z_2, indicating that received to transmitted ratio between the two antennas would be independent of the transmission and reception roles assigned to the antennas. The same conclusion also holds if the ratio is calculated *exactly*, without making use of $|Z_c| \ll |Z_{1,2}|$ approximation, yielding (to be shown in HW)

$$\frac{P_r}{P_t} = \frac{|Z_c|^2}{4\mathrm{Re}\{Z_1\}\mathrm{Re}\{Z_2\} - 2\mathrm{Re}\{Z_c^2\}}.$$

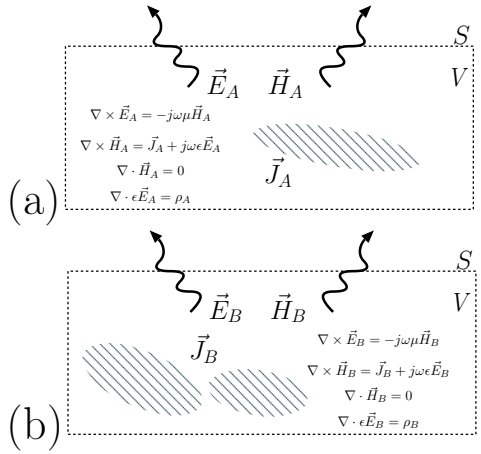


Eqv. ckt. showing antenna 1 in transmission and antenna 2 in reception terminated by a matched load

- The results presented so far regarding antenna reception has focused on the reciprocity between transmission and reception roles of antennas, following from $V_1 I_1 = V_2 I_2$.
  - At this stage we would like to derive this key equality from Maxwell's equations and also find out how the open circuit antenna voltages $V_{1,2} = Z_c I_{2,1}$ can be computed explicitly.
  - Consider the situations (a) and (b) depicted in the margin:



- In situation (a) currents $\mathbf{J}_A$ confined to some volume V surrounded by a surface S generate the fields $\mathbf{E}_A$ and $\mathbf{H}_A$ propagating within and outside the volume V as governed by Maxwell's equations shown within the box.
- Situation (b) depicts the very same volume of space but with a different set of currents $\mathbf{J}_B$ and fields $\mathbf{E}_B$ and $\mathbf{H}_B$ also obeying Maxwell's equations shown in the second box.



– Observe that with the set of equations shown in the boxes

$$\begin{aligned}\mathbf{H}_B \cdot \nabla \times \mathbf{E}_A - \mathbf{H}_A \cdot \nabla \times \mathbf{E}_B &= 0\\ \mathbf{E}_B \cdot \nabla \times \mathbf{H}_A - \mathbf{E}_A \cdot \nabla \times \mathbf{H}_B &= \mathbf{J}_A \cdot \mathbf{E}_B - \mathbf{J}_B \cdot \mathbf{E}_A.\end{aligned}$$

- Subtracting the first equation from the second one, and using the identity $\mathbf{H} \cdot \nabla \times \mathbf{E} - \mathbf{E} \cdot \nabla \times \mathbf{H} = \nabla \cdot (\mathbf{E} \times \mathbf{H})$ used in the derivation of Poynting Theorem we get

$$\nabla \cdot [\mathbf{E}_A \times \mathbf{H}_B - \mathbf{E}_B \times \mathbf{H}_A] = \mathbf{J}_A \cdot \mathbf{E}_B - \mathbf{J}_B \cdot \mathbf{E}_A.$$

- Integrating both sides over an arbitrarily big spherical volume V' centered about surface S and volume V and using the divergence theorem (on the left hand side) we obtain

$$\oint_{S'} d\mathbf{S} \cdot [\mathbf{E}_A \times \mathbf{H}_B - \mathbf{E}_B \times \mathbf{H}_A] = \oint_V dV\, [\mathbf{J}_A \cdot \mathbf{E}_B - \mathbf{J}_B \cdot \mathbf{E}_A],$$

where S' is the surface of the larger volume V', with $d\mathbf{S} = dS\hat{n}$ on S', where $\hat{n}$ is the outward unit vector on surface S'.

– It turns out both sides of this equality are zero and it is easier to see how left hand side vanishes: on S', sufficiently far away from the radiation sources within S, the outward propagating waves appear locally planar with $\mathbf{H}_{A,B} = \hat{n} \times \mathbf{E}_{A,B}/\eta$, implying that $\mathbf{E}_A \times \mathbf{H}_B = \mathbf{E}_B \times \mathbf{H}_A$ (to be shown in HW) leading to the conclusion

$$\oint_V dV\, \mathbf{J}_A \cdot \mathbf{E}_B = \oint_V dV\, \mathbf{J}_B \cdot \mathbf{E}_A$$

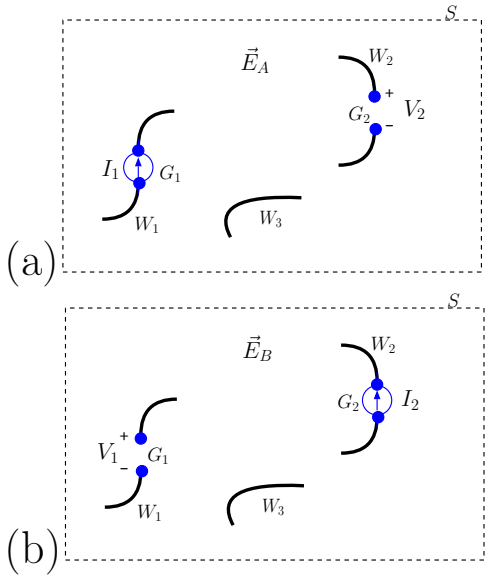
known as the ***reciprocity theorem*** — the left side volume integral is said to be the **reaction**[2] of source $\mathbf{J}_A$ to source $\mathbf{J}_B$ producing $\mathbf{E}_B$, equal to the

$$\oint_V dV\, \mathbf{J}_A \cdot \mathbf{E}_B = \oint_V dV\, \mathbf{J}_B \cdot \mathbf{E}_A$$

[2] see Rumsey, 1954, https://doi.org/10.1103/PhysRev.94.1483.

reaction of source $\mathbf{J}_B$ to source $\mathbf{J}_A$ represented by the right side integral over the same volume.

- We next use the reciprocity theorem just derived to obtain $V_1 I_1 = V_2 I_2$ pertinent to a pair of interacting antennas shown in the situations (a) and (b) depicted in the margin.



  - Current densities $\mathbf{J}_A$ and $\mathbf{J}_B$ have been constrained to flow on surfaces of perfectly conducting and infinitely thin wires contained within S. Under that assumption the volume integrals above reduce to line integrals and the Reciprocity Theorem becomes

$$\int_L I_A d\mathbf{l} \cdot \mathbf{E}_B = \int_L I_B d\mathbf{l} \cdot \mathbf{E}_A,$$

    where it is implied that the line integrals will be carried over all the wires and current paths located within S.

    - We will use the convention that the current reference direction in each segment agrees with the vector direction of $d\mathbf{l}$, and $\mathbf{E}_B I_A$ as well as $\mathbf{E}_A I_B$ vary with position along the integration paths established by the wires contained within S.

  - With the current I_A of situation (a) and $L = G_1 + W_1 + W_2 + W_3$, we have the reaction (of A to B)

$$\int_L I_A d\mathbf{l} \cdot \mathbf{E}_B = I_1 \int_{G_1} d\mathbf{l} \cdot \mathbf{E}_B + \cancelto{0}{\int_{W_1+W_2+W_3} I_A d\mathbf{l} \cdot \mathbf{E}_B},$$

    since $I_A = I_1$ in the gap region G_1 of antenna 1 while the electric field $\mathbf{E}_B$ will have a zero tangential component along the infinitely conducting wire segments W_1, W_2, and W_3.

  - Likewise, with the current I_B of situation (b) and $L = W_1 + W_2 + G_2 + W_3$, we have the reaction (of B to A)

$$\int_L I_B d\mathbf{l} \cdot \mathbf{E}_A = \cancelto{0}{\int_{W_1+W_2+W_3} I_B d\mathbf{l} \cdot \mathbf{E}_A} + I_2 \int_{G_2} d\mathbf{l} \cdot \mathbf{E}_A.$$

$$\oint_V dV\, \mathbf{J}_A \cdot \mathbf{E}_B = \oint_V dV\, \mathbf{J}_B \cdot \mathbf{E}_A$$

becomes

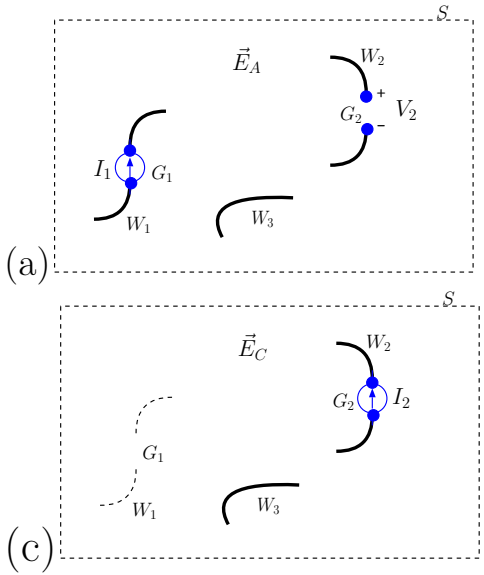
$$\int_L I_A d\mathbf{l} \cdot \mathbf{E}_B = \int_L I_B d\mathbf{l} \cdot \mathbf{E}_A$$

– Hence, the Reciprocity Theorem, the equality of the reactions, implies that

$$I_1 \underbrace{\int_{G_1} d\mathbf{l}\cdot\mathbf{E}_B}_{\equiv -V_1} = I_2 \underbrace{\int_{G_2} d\mathbf{l}\cdot\mathbf{E}_A}_{\equiv -V_2} \Rightarrow I_1V_1 = I_2V_2 \quad \textbf{QED} \;\; ;\text{-)}$$

in terms of *open circuit voltage* V_1 and V_2 across the gaps G_1 and G_2, respectively, with "polarities" shown in the margin - remember, an "E-dot-dl" line integral is always a "voltage drop" in the integration direction!

- Let's next use the reciprocity theorem once more, as applied to situations (a) and (c) shown in the margin, in order to derive an explicit formula for open circuit voltage $V_1 = Z_c I_2$ of antenna 1 in terms of the field radiated by antenna 2 that we will initially call $\mathbf{E}_C$ as shown in the margin — $\mathbf{E}_C$ differs from field $\mathbf{E}_B$ of situation (b), as $\mathbf{E}_B$ contained "scattered" field components caused by currents flowing in wire segments W_1 of antenna 1 which are absent in $\mathbf{E}_C$.



– Starting with current I_A of situation (a) and $L = G_1 + W_1 + W_2 + W_3$, we have the reaction

$$\int_L I_A d\mathbf{l}\cdot\mathbf{E}_C = \int_{G_1+W_1} I_A d\mathbf{l}\cdot\mathbf{E}_C + \cancelto{0}{\int_{W_2+W_3} I_A d\mathbf{l}\cdot\mathbf{E}_C},$$

since $\mathbf{E}_C$ generated in situation (c) is shorted on wires W_2 and W_3, but not at locations $G_1 + W_1$ where antenna 1 used to be before its removal to create situation (c).

– Next with current I_C of situation (c) and $L = W_2 + G_2 + W_3$, we have the reaction

$$\int_L I_C d\mathbf{l}\cdot\mathbf{E}_A = \cancelto{0}{\int_{W_2+W_3} I_C d\mathbf{l}\cdot\mathbf{E}_A} + I_2 \int_{G_2} d\mathbf{l}\cdot\mathbf{E}_A,$$

because I_C can only flow through wires W_2 and W_3 and across gap G_2 (as a source current I_2) and $\mathbf{E}_A$ is naturally shorted out on wires W_2 and W_3.

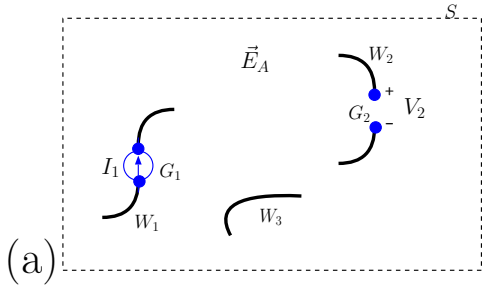
$$\int_L I_A d\mathbf{l}\cdot\mathbf{E}_C = \int_L I_C d\mathbf{l}\cdot\mathbf{E}_A$$

– Hence, the reciprocity theorem implies that

$$\int_{G_1+W_1} I_A d\mathbf{l}\cdot\mathbf{E}_C \;=\; I_2 \underbrace{\int_{G_2} d\mathbf{l}\cdot\mathbf{E}_A}_{-V_2}$$

and since $V_1 I_1 = V_2 I_2$ this result also implies that

$$-V_1 I_1 = \int_{G_1+W_1} I_A d\mathbf{l}\cdot\mathbf{E}_C \;\Rightarrow\; V_1 = -\int_{G_1+W_1} \frac{I_A}{I_1} d\mathbf{l}\cdot\mathbf{E}_C.$$

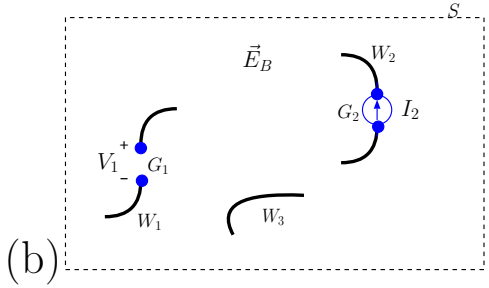


(a)

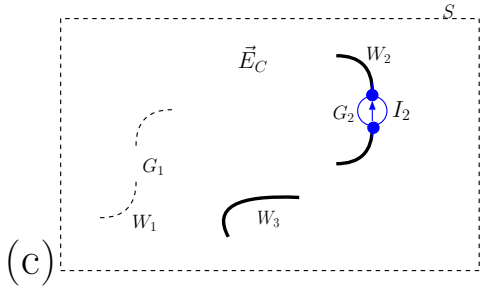
– We now have an explicit formula for the open circuit voltage V_1 of antenna 1 in terms of the undistorted wave field $\mathbf{E}_C \equiv \mathbf{E}_2$ caused solely by I_2-driven antenna 2 and extraneous reflectors/scatterers (like W_3) that may exist in the environment. Since $G_1 + W_1 \equiv L_1$ is the full integration path belonging to antenna 1, we can re-write this result as

$$V_1 = -\int_{L_1} \frac{I_1(\mathbf{r})}{I_1(0)} d\mathbf{l}\cdot\mathbf{E}_2(\mathbf{r})$$

in terms of incident electric field phasor $\mathbf{E}_2(\mathbf{r})$ on antenna 1 and the normalized current distribution $\frac{I_1(\mathbf{r})}{I_1(0)}$ of antenna 1 that it carries when it is used for transmission, as in situation (a), with $I_1(\mathbf{r}) = I_A$ and $I_1(0) = I_1$ representing the input current of antenna 1.



(b)



(c)

- **The upshot:** *any* wire antenna that supports a normalized current distribution $\frac{I(\mathbf{r})}{I_o}$ along its length when it is being used for transmission purposes will exhibit an open circuit voltage of

$$\int_L I_A d\mathbf{l}\cdot\mathbf{E}_C = \int_L I_C d\mathbf{l}\cdot\mathbf{E}_A$$

$$V_o = -\int_L \frac{I(\mathbf{r})}{I_o} d\mathbf{l}\cdot\mathbf{E}(\mathbf{r})$$

across its terminals when it is exposed to a (total) incident field $\mathbf{E}(\mathbf{r})$; the integration path L in the formula is "self-defined" by all the locations with non-zero $\frac{I(\mathbf{r})}{I_o}$ and, *by definition*, $\frac{I(\mathbf{r})}{I_o} = 1$ at the antenna feed point.

  - V_o is the voltage rise across the antenna input terminals in the direction the antenna current $I(\mathbf{r})$ is assumed to flow when the antenna is transmitting.
  - If $d\mathbf{l}$ and $\mathbf{E}(\mathbf{r})$ are orthogonal over the entire integration path L then V_o will be zero and the antenna will not detect such *cross-polarized* fields.

- See Lecture 20 for an example of the use of this general result to compute the open circuit voltage of a z-polarized dipole antenna as

$$V_o = E_c \ell_{eff}(\theta, \phi) \sin\theta$$

in reponse to an incident plane wave field arriving from direction θ. Next we generalize this for any type of antenna located at the origin illuminated by a plane TEM wave with a phasor given by

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_o e^{j\mathbf{k}\cdot\mathbf{r}}$$

arriving from a direction

$$\hat{k} = \frac{\mathbf{k}}{|\mathbf{k}|} = \frac{\mathbf{k}}{k}$$

as shown in the margin.

- In this case

$$V_o = -\int_L \frac{I(\mathbf{r})}{I_o} d\mathbf{l} \cdot \mathbf{E}(\mathbf{r}) = -\mathbf{E}_o \cdot \int_L d\mathbf{l} \frac{I(\mathbf{r}')}{I_o} e^{jk\hat{k}\cdot\mathbf{r}'}.$$

  – Noting that

$$\hat{k} \times (\hat{k} \times d\mathbf{l}) \;=\; (\hat{k} \cdot d\mathbf{l})\hat{k} - \underbrace{(\hat{k} \cdot \hat{k})}_{1} d\mathbf{l} = (\hat{k} \cdot d\mathbf{l})\hat{k} - d\mathbf{l}$$

and, thus,

$$\mathbf{E}_o \cdot \hat{k} \times (\hat{k} \times d\mathbf{l}) \;=\; (\hat{k} \cdot d\mathbf{l}) \underbrace{\hat{k} \cdot \mathbf{E}_o}_{0 \text{ since the field is TEM}} - \mathbf{E}_o \cdot d\mathbf{l} = -\mathbf{E}_o \cdot d\mathbf{l},$$

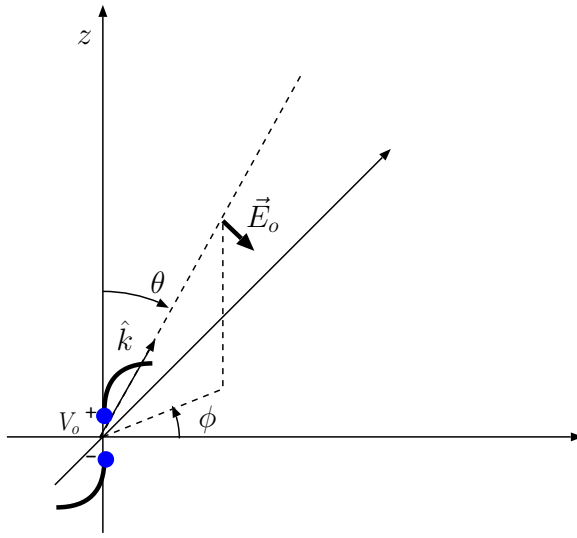
we can write

$$V_o = -\mathbf{E}_o \cdot \int_L d\mathbf{l} \frac{I(\mathbf{r}')}{I_o} e^{jk\hat{k}\cdot\mathbf{r}'} = \mathbf{E}_o \cdot \hat{k} \times \hat{k} \times \int_L d\mathbf{l} \frac{I(\mathbf{r}')}{I_o} e^{jk\hat{k}\cdot\mathbf{r}'} = \mathbf{E}_o \cdot \mathbf{p}(\hat{k}),$$

where

$$\mathbf{p}(\hat{k}) = \hat{k} \times \hat{k} \times \int_L d\mathbf{l} \frac{I(\mathbf{r}')}{I_o} e^{jk\hat{k}\cdot\mathbf{r}'}$$

is a $\hat{k}$-dependent **voltage projection vector** of the receiving antenna to extract from the incoming plane wave field the open circuit voltage response of the antenna — V_o is the projection of $\mathbf{E}_o$ onto $\mathbf{p}(\hat{k})$.



A TEM plane wave $\mathbf{E}_o e^{j\mathbf{k}\cdot\mathbf{r}}$ is incident on an antenna at the origin from a direction $\hat{k}$. Note that $\mathbf{k} \cdot \mathbf{E}_o = 0$.

– For the z-polarized dipole $d\mathbf{l} = dz\,\hat{z}$ and, therefore, the projection vector reduces to

$$\mathbf{p}(\hat{k}) = \hat{k} \times \hat{k} \times \hat{z} \int_{L/2}^{L/2} dz \frac{I(z)}{I_o} e^{jk\cos\theta z}$$
$$= \hat{\theta} \sin\theta \int_{L/2}^{L/2} dz \frac{I(z)}{I_o} e^{jk\cos\theta z} = \hat{\theta} \ell_{eff}(\theta, \phi) \sin\theta,$$

with a *magnitude* given by the foreshortened effective length of the antenna used in radiation electric field calculations and a *direction* $\hat{\theta}$ matching the direction of the radiated electric field vector of the antenna — this is consistent with our earlier finding of $V_o = E_c \ell_{eff}(\theta, \phi)\,\sin\theta$ specific for z-pol dipoles.

- With this knowledge we can always construct the projection vector $\mathbf{p}(\hat{k})$ of any antenna needed to extract V_o out of incident electric field vector $\mathbf{E}_o$ arriving from a direction $\hat{r} = \hat{k}$ provided that we know the foreshortened effective length of the antenna and the polarization direction of its radiated electric field into the same radial direction.
- Note that the direction of projection vector is distinct from the direction of antenna orientation and therefore V_o is in general distinct from the projection of incoming $\mathbf{E}_o$ vector onto the the direction of antenna orientation.

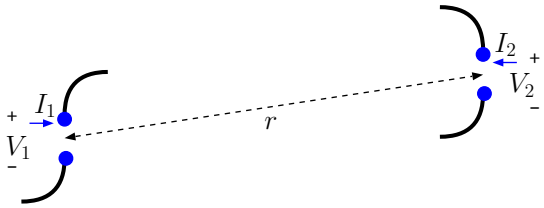
Antenna open circuit voltage as

$$V_o = \mathbf{E}_o \cdot \mathbf{p}(\hat{k})$$

on terms of incident plane wave field phasor $\mathbf{E}_o$ at the antenna location and antenna projection vector

$$\mathbf{p}(\hat{k}) = \hat{k} \times \hat{k} \times \int_L d\mathbf{l} \frac{I(\mathbf{r}')}{I_o} e^{jk\hat{k}\cdot\mathbf{r}'}$$

- The power budget equation of a radio link between a pair of antennas as shown in the margin is known as **Friis transmission formula:**



  - P_t is the average transmitted power of antenna 1,
  - $S_{inc} = \frac{P_t}{4\pi r^2} G_1$ is the Poynting flux incident on antenna 2 at a distance r in terms of antenna 1 gain G_1 in the direction of antenna 2,
  - $P_r = S_{inc} A_2$ is the available power of antenna 2 in terms of its effective area $A_2 = \frac{\lambda^2}{4\pi} G_2$, leading to

$$P_r = P_t \frac{\lambda^2 G_1 G_2}{(4\pi r)^2} \quad \textbf{Friis transmission formula}$$

Example 1: A typical GSM cell tower will radiate 10 W of average power at a carrier frequency of 750 MHz. Assuming a 1 km direct path between the cell tower and your GSM hand held receiver compute the received power P_r assuming that both the tower and the receiver make use of small antennas of directivities $D = 3/2$.

Solution: At the transmission frequency of $f = 750$ MHz the wavelength

$$\lambda = \frac{c}{f} = \frac{3 \times 10^8}{750 \times 10^6} = \frac{300}{750} = 0.4\,\text{m}.$$

Therefore, using $G_1 = G_2 = 1.5$ in Friis Transmission Formula we get an upper bound (assuming best/co-polarized orientations) of

$$P_r = P_t \frac{\lambda^2 G_1 G_2}{(4\pi r)^2} = 10\,(\frac{0.4 \times 1.5}{4\pi 10^3})^2 \approx 2.28 \times 10^{-8}\,\text{W} = 22.8\,\text{nW}$$

or

$$10 \log_{10} \frac{2.28 \times 10^{-8}}{10^{-6}} = -16.42\,\text{dBm},$$

as typically used in the RF world.

dBm stands for dB above a mW power level, e.g., 3 dBm is 2 mW, 10 dBm is 10 mW, 20 dBm is 100 mW and so on.

- The power budget of a radar system is referred to as **radar equation** and can be obtained as follows:
  - A radar transmits an average power of P_t with an antenna G_t in the direction of an object at some distance r that will reflect a portion of average power it intercepts from the incident Poynting flux
    $$S_{inc} = \frac{P_t}{4\pi r^2} G_t$$
    back to the antenna.
  - If the target intercepts an average power of $P_{int} = S_{inc}\sigma_{RCS}$, where σ_{RCS} is an effective area known as **radar cross-section** (RCS) of the target and re-radiates it isotropically, then the radar antenna will receive and deliver to its matched load of an average power of
    $$P_r = \frac{P_t}{4\pi r^2} G_t \sigma_{RCS} \frac{1}{4\pi r^2} A_t$$
    where $A_t = \frac{\lambda^2}{4\pi} G_t$ is the effective area of the radar antenna for reception.

– It follows that

$$P_r = P_t \frac{\lambda^2 G_t^2}{(4\pi)^3 r^4} \sigma_{RCS}$$

which is the **radar equation**.

- As an example, the RCS of a square shaped PEC plate of with a physical area $A_{plate} \gg \lambda^2$ would be

$$\sigma_{plate} = \frac{4\pi}{\lambda^2} A_{plate}^2.$$

This can be derived by noting that the PEC plate will redirect the power $\frac{P_t}{4\pi r^2} G_t A_{plate}$ that it intercepts from S_{inc} back to the radar antenna with an effective gain of $G_{plate} = \frac{4\pi}{\lambda^2} A_{plate}$, behaving as if it were a uniformly excited filled antenna array of area A_{plate}.

  - Accordingly

$$\frac{P_r}{P_t} = \frac{\lambda^2 G_t^2}{(4\pi)^3 r^4} \frac{4\pi}{\lambda^2} A_{plate}^2 = (\frac{G_t A_{plate}}{4\pi r^2})^2 = (\frac{G_t \Omega_{plate}}{4\pi})^2$$

for a PEC plate with a solid angle Ω_{plate} as seen from the radar site!

  - Estimating the solid angle of an airplane at 10 km distance to a airport radar as 10^{-6} steradian and assuming $G_t = 4\pi \times 10^2$ for a typical airport radar, the use of above equation would yield

$$\frac{P_r}{P_t} = (\frac{G_t \Omega_{airplane}}{4\pi})^2 = 10^{-8} = -80\,\mathrm{dB}.$$

22 Doppler shift and Doppler radars

- **Doppler radars** make a use of the **Doppler shift** phenomenon to detect the motion of EM wave reflectors of interest — e.g., a police Doppler radar aims to identify the speed of a vehicle in relative motion.

  – In this lecture we will describe the general principle of how a Doppler radar works and also learn about the Doppler shift phenomenon in non-relativistic and relativistic limits.

Doppler radar:

- Consider a stationary dipole located at the origin excited by a cosinusoidal input current

$$i(t) = I_o \cos(\omega t) \propto e^{j\omega t} + e^{-j\omega t} \equiv e^{j\omega t} + \text{cc}$$

  where "cc" refers to the complex conjugate of the term preceding it.

- The dipole will radiate a spherical wave field

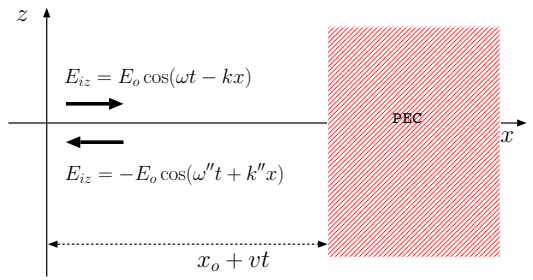
$$\mathbf{E}(\mathbf{r}, t) \propto \cos(\omega t - kr) \propto e^{j(\omega t - kr)} + \text{cc}$$

  where

$$\frac{\omega}{k} = c$$

  assuming propagation in vacuum or air.

- Consider now a car speeding away with velocity v from the dipole along the x-axis having an instantaneous location



$$x(t) = x_o + vt$$

at time t. The field at the location of the car at time t will then be

$$\propto \cos(\omega t - k(x_o + vt)) = \cos((\omega - kv)t - kx_o) \propto e^{j((\omega - kv)t - kx_o)} + \text{cc.}$$

- An induced surface current $\propto \cos(\omega' t - kx_o)$ on the car's body oscillating at a frequency

$$\omega' = \omega - kv$$

will then radiate like a collection of dipoles, producing a "reflected field"

$$\propto \cos(\omega' t - kx_o - k(x_o + vt)) = \cos((\omega - 2kv)t - 2kx_o) \propto e^{j((\omega - 2kv)t - 2kx_o)} + \text{cc}$$

detected back at the location of the original dipole — in this waveform we have included an additional phase delay of $k(x_o + vt)$ to account for the return trip of the reflected wave. Clearly, the reflected field oscillates with the frequency

$$\omega'' = \omega' - kv = \omega - 2kv$$

in the reference frame of the stationary dipole.

  - If the dipole is arranged to detect the reflected wave field (using a T/R switch — a radar jargon implying that the antenna is

switched to connect to the input port of a receiving device shortly after the transmission of a burst of EM wave), then the velocity of the car, v, can be obtained from "two-way" Doppler shifted frequency ω''. That's how police radars work.

Note that

$$\omega' = \omega - kv$$

$$\omega'' = \omega - 2kv$$

- positive v (motion away from the radar antenna) causes $\omega'' < \omega$ and is referred to as **redshift**, whereas
- negative v (motion toward the radar antenna) causes $\omega'' > \omega$ and is referred to as **blueshift**.

Doppler shift in relativistic and non-relativistic limits:

- The "one-way" and "two-way" Doppler shift formulae

$$\omega' = \omega - kv \quad \text{and} \quad \omega'' = \omega - 2kv$$

obtained above, where v is the relative[1] radial recession velocity of the radiator and the observer, are valid only when $|v| \ll c$.

The reason for this is, our analysis above, leading to these formulae, neglected an important detail that according to Maxwell's equations we need to have

$$\text{not only } \frac{\omega}{k} = c, \quad \text{but also } \frac{\omega'}{k'} = c,$$

[1]It does not matter whether the radiator or the observer is "moving" since motion is always *relative*.

whereas we have, in effect, used an inconsistent relation $\frac{\omega'}{k} = c$ at an intermediate stage.

  - This inconsistency produces a negligible error if $|v| \ll c$ (the usual case pertinent for police radar applications) but the errors are unacceptably large if $|v|$ approaches c (like in Fermilab).

- We will refer to the approximate Doppler shift formulae given above as ***non-relativistic*** **Doppler formulae** — they are to be used if and only if $|v/c| \ll 1$, i.e., in the non-relativistic limit.

- ***Relativistic*** **Doppler formulae** that can be used unconditionally (and most importantly for $|v/c|$ approaching unity) are

$$\omega' = \omega\sqrt{\frac{1-\frac{v}{c}}{1+\frac{v}{c}}} \quad \text{and} \quad \omega'' = \omega'\sqrt{\frac{1-\frac{v}{c}}{1+\frac{v}{c}}} = \omega\frac{1-\frac{v}{c}}{1+\frac{v}{c}}.$$

  - Before deriving these relativistic formulae (correct for all v), let us note that they reduce to the non-relativistic formula if $|v/c| \ll 1$. In that case we have, for instance,

$$\begin{aligned}\omega' &= \omega\sqrt{\frac{1-\frac{v}{c}}{1+\frac{v}{c}}} = \omega\frac{(1-\frac{v}{c})^{1/2}}{(1+\frac{v}{c})^{1/2}} = \omega(1-\frac{v}{c})^{1/2}(1+\frac{v}{c})^{-1/2} \\ &\approx \omega(1-\frac{v}{2c})(1-\frac{v}{2c}) \approx \omega(1-\frac{v}{c}) = \omega - kv.\end{aligned}$$

Derivation of the relativistic formula:

To derive the relativistic Doppler shift formulae we will not need complicated relativistic transformation formulae discussed in PHYS 325 (also summarized in ECE 329 notes). It is sufficient that we make a careful use of Maxwell's equations in developing an accurate model of a field reflected from a reflector in motion as shown next:

- Consider a plane TEM wave in free-space,

$$\mathbf{E}_i(x,t) = \hat{z}E_o\cos(\omega t - kx),$$

incident on a conducting surface at $x = 0$ plane from the left such that

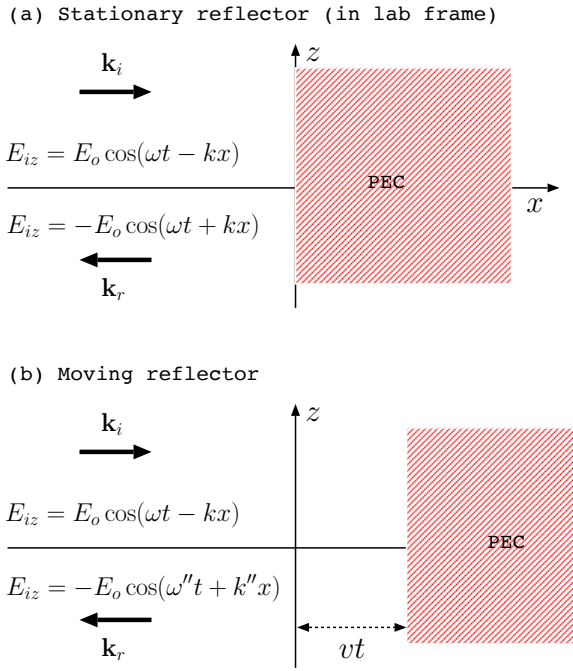
$$k = \frac{\omega}{c}.$$

The wave will be reflected to produce

$$\mathbf{E}_r(x,t) = -\hat{z}E_o\cos(\omega t + kx)$$

so that the total tangential field at $x = 0$ plane

$$\hat{z}\cdot(\mathbf{E}_i(0,t) + \mathbf{E_r}(0,t)) = E_o\cos(\omega t - 0) - E_o\cos(\omega t + 0) = 0.$$



Now, what would $\mathbf{E}_r(x,t)$ be if the conducting reflector were not stationary on the $x = 0$ plane, but rather moving with a steady velocity v to the right, having a trajectory $x = vt$ as depicted in the margin?

- The answer of the question raised above is quite simple: We would have

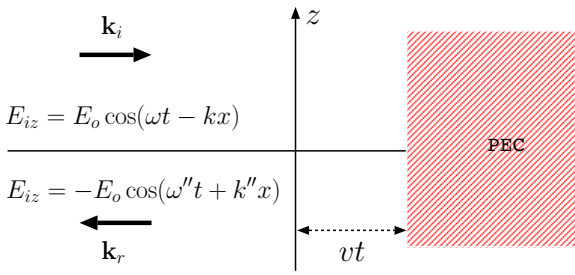
$$\mathbf{E}_r(x,t) = \hat{z} f(t + \frac{x}{c}),$$

where $f(t)$ is to be determined, so that

$$\hat{z} \cdot (\mathbf{E}_i(vt,t) + \mathbf{E_r}(vt,t)) = E_o \cos(\omega t - kvt) + f(t + \frac{vt}{c}) = 0,$$

because

1. $\mathbf{E}_r(x,t) = \hat{z} f(t + \frac{x}{c})$ is a viable (and the *only* viable) $\hat{z}$-polarized wave solution of Maxwell's equations propagating in the $-x$ direction in free space, and
2. the second equation above is the relevant boundary condition to be fulfilled on the surface of the moving reflector at every instant in time.



The boundary condition equation above implies that

$$f(t(1+\frac{v}{c})) = -E_o \cos(\omega t - kvt) = -E_o \cos(\omega t(1-\frac{v}{c})) = -E_o \cos(\omega \frac{1-\frac{v}{c}}{1+\frac{v}{c}} t(1+\frac{v}{c})).$$

Thus,

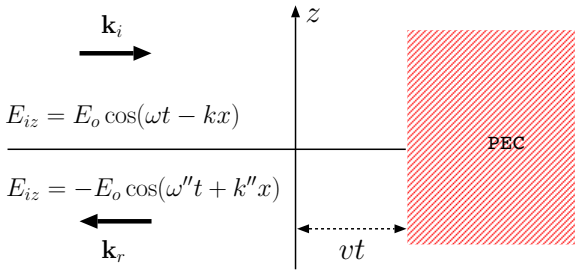
$$f(t) = -E_o \cos(\omega \frac{1-\frac{v}{c}}{1+\frac{v}{c}} t),$$

and

$$\mathbf{E}_r(x,t) = \hat{z} f(t+\frac{x}{c}) = -\hat{z} E_o \cos(\omega \frac{1-\frac{v}{c}}{1+\frac{v}{c}} (t+\frac{x}{c})) = -\hat{z} E_o \cos(\omega'' t + k'' x),$$

with

$$\omega'' = \omega\frac{1-\frac{v}{c}}{1+\frac{v}{c}} \quad \text{and} \quad k'' = \frac{\omega''}{c} = \frac{\omega}{c}\frac{1-\frac{v}{c}}{1+\frac{v}{c}} = k\frac{1-\frac{v}{c}}{1+\frac{v}{c}}.$$



With

$$\omega'' = \omega\frac{1-\frac{v}{c}}{1+\frac{v}{c}} \quad \text{and} \quad k'' = k\frac{1-\frac{v}{c}}{1+\frac{v}{c}}$$

the reflected wave

$$\mathbf{E}_r(x,t) = \hat{z}f(t+\frac{x}{c}) = -\hat{z}E_o\cos(\omega''t + k''x)$$

wave is clearly a co-sinusoid — just like the incident wave — but with *Doppler shifted* **frequency and wavenumbers ω'' and k'', respectively, caused by the motion of the reflector surface (as discussed below).** The result can also be used with negative v corresponding to a reflector moving to the left.

- The Doppler shift formulae given above are relativistically correct — that is, they are valid for all possible values of $\frac{v}{c}$ — even though we did not invoke any "relativistic argument" above.

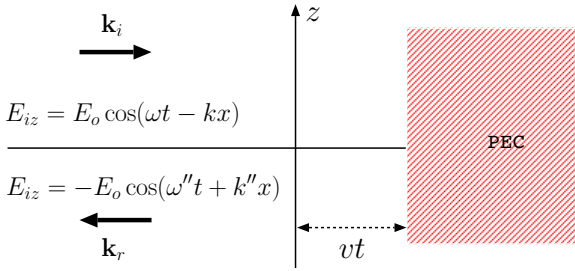
This is true because *relativity* derives from the Maxwell's equations and the accompanying boundary conditions, and so *any* rigorous deduction derived from Maxwell's equations will be *by default* relativistically valid.

- Focusing next on the Doppler shifted frequency formula

$$\omega'' = \omega \frac{1 - \frac{v}{c}}{1 + \frac{v}{c}} = \omega \sqrt{\frac{1 - \frac{v}{c}}{1 + \frac{v}{c}}} \sqrt{\frac{1 - \frac{v}{c}}{1 + \frac{v}{c}}},$$

we can re-express ω'' as

$$\omega'' = \omega' \sqrt{\frac{1 - \frac{v}{c}}{1 + \frac{v}{c}}} \quad \text{with} \quad \omega' \equiv \omega \sqrt{\frac{1 - \frac{v}{c}}{1 + \frac{v}{c}}}.$$



  - We now recognize the Doppler shifted frequency

$$\omega'' = \omega' \sqrt{\frac{1 - \frac{v}{c}}{1 + \frac{v}{c}}}$$

of the reflected wave as a Doppler shifted version of the *wave frequency*

$$\omega' = \omega \sqrt{\frac{1 - \frac{v}{c}}{1 + \frac{v}{c}}}$$

seen in the reflector frame, which is in turn a Doppler shifted version of the frequency ω of the in the incident wave field $\mathbf{E}_i(x,t)$ defined in the so-called[2] "lab frame".

This concludes our derivation of the relativistic Doppler shift formulae stated earlier on.

[2] By definition the frame where the "unprimed" frequency ω is observed is the *lab frame*; it can also be called the *unprimed frame*.

One-way Doppler shift:

When a TEM wave is observed to have a frequency ω in the lab frame (and wavenumber $k = \omega/c$ since we are concerned with free-space propagation at this point), the same TEM wave will appear to have a frequency ω' in a second reference frame which is in motion within the lab frame.

- The *one-way* Doppler shifted frequency ω' will be related to the lab-frame frequency ω as

$$\omega' = \omega\sqrt{\frac{1-\frac{v}{c}}{1+\frac{v}{c}}}$$

  if the moving observer has a velocity v in the lab frame defined to be positive in the direction of wave propagation (away from the wave source).

- For non-relativistic speeds such that $\frac{|v|}{c} \ll 1$ we have

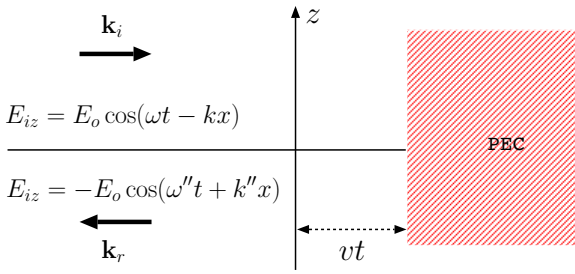
$$\omega' \approx \omega(1-\frac{v}{c}) = \omega - kv$$

  as already seen. This simplified Doppler formula is easy to understand since

$$\mathbf{E}_i(x,t) = \hat{z}E_o\cos(\omega t - kx)$$

  (see margin) implies that the incident field at the location $x = vt$ of the reflector must vary with time t as

$$\mathbf{E}_i(vt,t) = \hat{z}E_o\cos(\omega t - kvt) = \hat{z}E_o\cos((\omega - kv)t) = \hat{z}E_o\cos(\omega' t)$$



where

$$\omega' = \omega - kv$$

as obtained above.

Time dilation and other relativistic effects:

- An astute student may ask at this point: "how come $\mathbf{E}_i(vt, t) \propto \cos((\omega - kv)t)$ and not $\cos(\omega' t) = \cos(\omega\sqrt{\frac{1-\frac{v}{c}}{1+\frac{v}{c}}}t)$ if a rigorous application of electromagnetic solutions should produce relativistically accurate results (as claimed earlier on)?"

  This is the sort of question Albert Einstein asked to himself in his free time at work in a Swiss patent office and figured out that the rigorous conclusion ought to be

  $$(\omega - kv)t = \omega\sqrt{\frac{1-\frac{v}{c}}{1+\frac{v}{c}}}t' \;\Rightarrow\; t' = \sqrt{\frac{1+\frac{v}{c}}{1-\frac{v}{c}}}(1-\frac{v}{c})t = \sqrt{1-\frac{v^2}{c^2}}t,$$

  where t' is the time kept by a clock attached to the reflecting surface.

  Time dilation

  $$t' = \sqrt{1-\tfrac{v^2}{c^2}}t$$

  - The fact that moving clocks tick slower than stationary clocks ($t' < t$) — known as *time dilation* — was one of the surprising results of the work Einstein published in 1905 under the title "*On the Electrodynamics of Moving Bodies*", popularly known as the relativity paper. Furthermore time dilation is a reciprocal effect since moving clocks are stationary in their own reference frames!

  moving clocks tick slower than stationary clocks

- The relativistic Doppler shift equations given above are applicable only if/when v represents the relative speed of a signal source and an observer approaching or receding along a common trajectory. As such these Doppler shift formulae we have learned are known as *longitudinal* Doppler shift formulae.

  Longitudinal vs transverse Doppler

  - If the signal source moves along a trajectory not passing through the location of the observer, then there is an additional *transverse* Doppler effect caused by *time dilation* of the signal

period from the source frame to the observer frame[3] — if the source speed appears as v in the observer frame is v, the source frequencyω_s will be detected by the observer, at the instant of the closest approach of the source to the observer, at a red-shifted frequency given by

$$\omega_o = \omega_s \sqrt{1 - \frac{v^2}{c^2}}$$

before ω_o subsequently approaching the longitudinal red-shifted frequency

$$\omega_s \sqrt{\frac{1 - v/c}{1 + v/c}}$$

as the source recedes to large distances.

— Notice that $|\omega_o - \omega_s|$ is much smaller for transverse Doppler than longitudinal Doppler with $|v| \ll c$, which is the reason why longitudinal Doppler is the better known and utilized phenomenon in practice than transverse Doppler.

Accelerating observers

- All moving observers within a lab frame permeated with a propagating wave of a frequency ω will detect longitudinal Doppler shifted frequencies

$$\omega' = \omega \sqrt{\frac{1 - v/c}{1 + v/c}},$$

where v denotes the instantaneous longitudinal velocity of the observer in the lab frame — *this applies to accelerating observers as well* since co-located observers side-by-side having the same instantaneous velocities will necessarily see the same instantaneous frequency ω' even if their accelerations differ (to cause different velocities and different ω' in future instances).[4]

Accelerating sources

- Consider an accelerating Antenna A, with position $z = 0$ and velocity $v_z = 0$ at time $t = 0$ within some *inertial frame*, emitting a wave pulse of some carrier frequency ω towards the acceleration

[3] Here is how time dilation causes a transverse redshift:
Slower-ticking clocks (because of time dilation) which are *stationary* within the moving source frame will measure a *shorter wave period* $2\pi/\omega_s$ than the period $2\pi/\omega_o = (2\pi/\omega_s)/\sqrt{1 - v^2/c^2}$ measured by the *faster-ticking clocks of the stationary observer,* to lead to $\omega_o = \omega_s \sqrt{1 - v^2/c^2}$ — see The Transverse and Oblique Doppler Effects.

[4] Another way of seeing this is, **special relativity** (SR) is a *local* approximation for **general relativity** (GR):
GR incorporates accelerations and gravity into a relativistic framework of a *globally* "curved spacetime" which is *locally* "flat" and governed by SR. Thus SR-based Doppler formulae also apply to accelerating observers on a instance-to-instance basis.

direction z, to be detected, at time $t = h/c$, by Antenna B also accelerating in z direction to maintain a fixed distance h from Antenna A. Because of its acceleration g, Antenna B will be moving away from $z = 0$ with a velocity gh/c (in the same inertial frame) when the pulse is passing it by — this will cause it to detect the pulse with a *redshifted* frequency[5] $\omega' = \omega\sqrt{\frac{1-gh/c^2}{1+gh/c^2}}$ obtained by using the usual longitudinal redshift formula $\omega' = \omega\sqrt{\frac{1-v/c}{1+v/c}}$ with $v = gh/c$.

Gravitational redshift

$$\omega' = \omega\sqrt{\frac{1-gh/c^2}{1+gh/c^2}}$$

— Since distance h between the antennas is not changing, this frequency change is not a "kinematic" redshift, but it can be interpreted as a "gravitational" redshift caused by the pulse climbing against what an accelerating observer carrying Antenna B perceives as a "gravitational pull" — this follows from the *equivalence principle* of general relativity stating that "being in an accelerating frame[6] is indistinguishable (locally) from being in a gravitational field".

The existence of a gravitational Doppler effect — which comes out from the description offered above — in the vicinity of gravitating masses is of utmost importance in cosmology, and also, somewhat surprisingly, has a substantial impact on the operation of our own GPS system[7]

[5] the same as the Doppler shifted frequency that would be detected by an unaccelerated (constant velocity) antenna, say Antenna C, having the same position and velocity as the accelerating Antenna B at the measurement instant — see page 282 in Einstein Gravity in a Nutshell, A. Zee (2013, Princeton).

Additional facts: A second constant-velocity antenna, antenna D, moving with a velocity Δv in the frame of Antenna C will measure (at the same location and instant) the same ω' multiplied by $\sqrt{1-\Delta v/c}/\sqrt{1+\Delta v/c}$, which can be shown to equal $\omega\sqrt{1-v_2/c}/\sqrt{1+v_2/c}$ in terms of v_2 representing the velocity of this second antenna within the inertial frame — just add the speed gh/c of antenna C and Δv relativistically to get v_2. Hence, Doppler shifts observed by non-accelerated detectors are not dependent on source accelerations — it is sufficient to know the detector velocity within a specific inertial frame where the emitter (i.e., antenna A) of frequency ω appears to have an instantaneous velocity of $v_z = 0$ at the emission time. Finally, for a detector in motion within the accelerated (non-inertial) frame of antenna A with some constant velocity v_d and distance h to antenna A at the time of pulse emission, use $v_d + gh/(c - v_d)$ in place of v in the usual redshift formula $\omega' = \omega\sqrt{\frac{1-v/c}{1+v/c}}$ — this produces an ω' due to a combination of kinematic and gravitational redshifts which can be well approximated by $\omega' \approx \omega(1 - v_d/c - gh/c^2)$ in case of small v_d and weak gravity.

[6] such as the frame in which the accelerating signal source (Antenna A) and its fixed-distance observer (carrying antenna B) described in this example are stationary — a *non-inertial* frame!

Additional facts: Acceleration of a free falling box under the pull of a gravitating mass makes the box an *inertial* frame since the equivalent gravitational force field within the box caused by its acceleration points in opposite direction to and cancels out the gravitational force field produced by the gravitating mass pulling down the box — the interior of the box will have zero-gravity, things will float about and there will be no gravitational Doppler shift of signals emitted by floating sources criss-crossing the interior of the box (although an accelerated source injected into the box will of course produce gravitational Doppler shift in its own non-inertial frame) — see https://www.eftaylor.com/spacetimephysics/02_chapter2.pdf

[7] There is also a *gravitational time dilation* effect described by $t' = t(1 + gh/c^2)$ accompanying the special *relativistic time dilation* effect

Relativistic antenna reception

- How about relativistic antenna reception?
  - In geometries pertinent to longitudinal Doppler shifts use $P_{rec} = S_{inc}A_{eff}$ as usual but take into account that S_{inc} is transformed from source frame to the receiver frame with a factor that is equal to the square of the Doppler shift factor — as an example if the source frequency appears redshifted from 2 GHz to 1 GHz, then the incident Poynting flux is reduced from 4 W/m^2 to 1 W/m^2 to be used in $P_{rec} = S_{inc}A_{eff}$ formula[8].

- Non-relativistic Doppler equations are also valid (in non-relativistic situations) within material media (such as in plasmas) where ω/k may be ω dependent.

Doppler in material media

  - However relativistic Doppler equations are only applicable in vacuum, the reason being, a frequency independent propagation velocity c was assumed in the derivation of ω''!
  - In material media stationary within the lab frame with the refractive index $n(\omega)$ the relativistic one-way Doppler formula takes the form

$$\omega' = \omega \frac{1 - \frac{v}{c} n(\omega)}{\sqrt{1 - \frac{v^2}{c^2}}}$$

    which reduces, with $n(\omega) = 1$, to ω' expression applicable for free space[9].

described by $t' = t\sqrt{1 - v^2/c^2}$ — this is a direct consequence of *gravitational redshift* described as $\omega' = \omega(1 - gh/c^2)$:
We can attribute the *lowered frequency* ω' of a photon climbing against a gravitational field (of a g-strength) to *faster ticking clocks residing at height* h measuring a *longer wave period* of $t' = 2\pi/\omega' = (2\pi/\omega)(1 + gh/c^2)$ compared to period $t = 2\pi/\omega$ of the photon at zero height (assuming $gh/c^2 \ll 1$, i.e., a weak gravity). As such $t' = t(1 + gh/c^2)$, telling us that time passes more rapidly at a height h above a gravitating body as compared to time measured on its surface ($h = 0$), and, equivalently, time slows down as you descend onto the surface of a gravitating body. See https://www.youtube.com/watch?v=JqKa6qyVYgg and https://www.youtube.com/watch?v=2JhQl_d4X7g

[8] of course A_{eff} will need to be computed with the Doppler shifted wavelength.
See Holmes and Ishimaru, IEEE Transactions on Antennas and Propagation (Volume: 17, Issue: 4, July 1969) for a broad discussion of relativistic radio links including arbitrary propagation angles measured from source velocity vectors within the receiver frames.

[9] e.g., in a plasma with $n(\omega) = \sqrt{1 - \omega_p^2/\omega^2}$;
Also note that $v_p > c$ in a plasma, which will be different in different "primed" frames given by $v_p' = \frac{c}{n'}$, where $n' = \sqrt{1 - \omega_{pl}^2/\omega'^2}$ in terms of a Lorentz contracted $\omega_{pl}^2 = \omega_p^2/\sqrt{1 - v^2/c^2}$ — see Achterberg & Wiersma (2007), DOI: 10.1051/0004-6361:20065365.

23 Doppler radars

- Doppler shift phenomenon is essential to the operation of Doppler radars such as weather radars or police radars for purposes of target motion measurements.

- In particular, the two-way Doppler shift equation

$$\omega'' = \omega \frac{1 - \frac{v}{c}}{1 + \frac{v}{c}} \approx \underbrace{\omega(1 - \frac{v}{c})(1 - \frac{v}{c}) \approx \omega(1 - 2\frac{v}{c}) = \omega - 2kv}_{\text{non-relativistic}}$$

  plays a major role because the reflected frequency ω'' from the moving target is compared to the incident frequency ω in order to estimate the target velocity v.

- In many applications the non-relativistic limit applies, i.e., $|v| \ll c$, in which case the target velocity v away from the radar antenna is obtained as

$$v = \frac{\omega - \omega''}{2k} = \frac{c}{2}\frac{\omega - \omega''}{\omega}.$$

  Note that when

  - the target is moving away from the radar, ω'' is red-shifted, $\omega - \omega'' > 0$, and v is positive.
  - the target is moving toward the radar, ω'' is blue-shifted, $\omega - \omega'' < 0$, and v is negative.

Also, the above formula for v gives the component of the target velocity in the direction of propagation of the incident wave from the radar.

Transverse motion of the radar target with respect to the incident wave from the radar does not cause any Doppler shift.

Example 1: Police radars catch you when the magnitude of

$$v = \frac{\omega - \omega''}{2k} = \frac{c}{2}\frac{\omega - \omega''}{\omega} \quad \text{exceeds 70 mph — here}$$

ω is the radar transmission frequency while ω'' is the frequency of the wave that bounces off your car back to the antenna of the police radar.

Also v is the component of your car's vector velocity away from the radar.

When your car is approaching the radar $\omega'' > \omega$ and v is negative. Conversely, when your car is moving away from the radar $\omega'' < \omega$ and v is positive.

- Total reflection is not necessary for Doppler effect and the operation of Doppler radars. Partially reflected waves from a moving dielectric surface, or even scattered fields from atoms in a gas in motion re-radiating like tiny dipole antennas excited by the incident wave, will also produce Doppler shifted returns of the incident (transmitted) radar signal governed by the Doppler shift formulae above.

  – Meteorologists and atmospheric scientists routinely make wind

measurements by bouncing EM waves from atmospheric atoms and ionospheric free electrons — also the research area of S. Franke, E. Kudeki, and J. Makela in the Remote Sensing Lab in our Dept.

- Engineers designing police radars and meteorologists building weather radars find themselves in strictly the non-relativistic domain $\frac{|v|}{c} \ll 1$. They will routinely think of the one-way Doppler shift formula

$$\omega' = \omega - kv$$

as the *time rate of change* of a plane wave phase

$$\omega t - kx$$

evaluated at the location

$$x = vt$$

of a moving observer.

In a broad sense, frequency of a wave in any reference frame is the time rate of change of the wave phase, and observers in relative motion naturally detect different rates in a wave field.

Example 2: A police radar with an operation frequency of $f = 300$ MHz is located at the origin $(x, y, z) = (0, 0, 0)$. A car with the trajectory

$$(x, y, z) = (50t, 50, 0)$$

is passing by, where the coordinates are given in meter units ant time t is measured in seconds.

(a) Determine the vector velocity of the car.

(b) Determine the frequency ω' of the radar signal in the reference frame of the car, by determining the rate of change of the signal phase detected by an antenna connected to the car.

(c) Determine the two-way shifted radar frequency ω'' by considering the rate of change of the phase delay of the reflected signal.

Solution: (a) The vector velocity of the car is

$$\mathbf{v} = \frac{\partial \mathbf{r}}{\partial t} = \frac{\partial}{\partial t}(50t, 50, 0) = (50, 0, 0) = 50\hat{x}\,\frac{\text{m}}{\text{s}}.$$

(b) The radial distance from the radar to the car is given by

$$r = \sqrt{(50t)^2 + 50^2}.$$

Therefore, the spherical wave phasor of the radar signal at the location of the car is proportional to

$$e^{-jkr} = e^{-jk\sqrt{(50t)^2+50^2}},$$

where $k = \omega/c = 2\pi$ rad/m. Therefore, the field at the location of the car varies with time in proportion to the real part of

$$e^{-jkr}e^{j\omega t} = e^{j(\omega t - k\sqrt{(50t)^2+50^2})}.$$

Thus, the phase of the signal detected by the antenna connected to the car is

$$\Phi(t) = \omega t - k\sqrt{(50t)^2 + 50^2}.$$

Finally, the frequency of the incident radar signal detected in the car frame is

$$\omega' = \frac{\partial \Phi}{\partial t} = \omega - k\frac{\frac{1}{2}(2(50t)50 + 0)}{\sqrt{(50t)^2 + 50^2}} = \omega - k50\frac{50t}{\sqrt{(50t)^2 + 50^2}} = \omega - k50\frac{t}{\sqrt{t^2 + 1}}.$$

(c) The field reflected from the moving car corresponds to the real part of

$$e^{-j2kr}e^{j\omega t} = e^{j(\omega t - 2k\sqrt{(50t)^2 + 50^2})}$$

since the phase delay occurs twice over the distance r. This leads to the two-way Doppler shifted radar frequency formula

$$\omega'' = \omega - k100\frac{t}{\sqrt{t^2 + 1}}.$$

Note that $\omega' = \omega'' = \omega$ at $t = 0$ when the car motion is transverse to the propagation direction of the incident radar wave.

Also note that the two-way Doppler shift

$$\omega'' - \omega = -k100\frac{t}{\sqrt{t^2 + 1}}$$

maximizes in magnitude at

$$|\omega'' - \omega| = 100k = 200\pi\,\tfrac{rad}{\mathrm{s}} \quad \Leftrightarrow \quad 100\,\mathrm{Hz}$$

for $t \ll -1$ s and $t \gg 1$ s when the car's motion is nearly collinear with the propagation direction of the incident wave from the radar.

24 Dispersion and propagation in collisionless plasmas

- TEM plane wave propagation in homogeneous conducting media can be described in terms of wavenumbers and intrinsic impedances

$$k = \omega\sqrt{\mu(\epsilon + \frac{\sigma}{j\omega})} \equiv k' - jk'' \quad \text{and} \quad \eta = \sqrt{\frac{\mu}{\epsilon + \frac{\sigma}{j\omega}}} \equiv |\eta|e^{j\tau}$$

  as we have seen in Lecture 18.

  For **real valued** σ these relations imply complex valued k and η as well as an ω *dependent* propagation velocity

$$v_p = \frac{\omega}{k'} = \frac{1}{\text{Re}\{\sqrt{\mu(\epsilon + \frac{\sigma}{j\omega})}\}}.$$

  Having an ω dependent v_p is a telltale sign that propagation of TEM waves in the medium will be *dispersive*, meaning that the shapes of TEM signals waveforms other than co-sinusoids will be distorted as a consequence of propagation — the distortion happens because different co-sinusoid components of the signal having different frequencies ω travel with different velocities v_p and thus fall out of synchronism!

- Dispersion in wave motions can be caused by a variety of reasons including the frequency dependence of the medium parameters as well as

geometrical effects related to the dimensions of the propagation region in relation to a wavelength.

  - For an *ohmic* medium where σ is real — such as seawater or copper — wave propagation is both *lossy* and dispersive.

- An important propagation medium known to be dispersive but *lossless* is the "**collisionless plasma**", an ionized gas in which collisions of the charge carriers (with one another) are negligibly small — a collisionless plasma provides an ideal setting to explore and understand the wave dispersion effects without having to deal with complications arising from losses and dissipation.

- A collisionless plasma is essentially a conducting medium with a purely imaginary conductivity σ (or, equivalently, a dielectric with a relative permittivity less than one, as we will see).

  - To develop the conductivity model for a collisionless plasma we envision a region of volume in free-space containing N free electrons per unit volume along with N positive ions (e.g., O^+ in the ionized portions of the upper atmosphere) which are also free. Each of these free charge carriers with charge q and mass m respond to an alternating electric field with a phasor $\tilde{\mathbf{E}}$ as dictated by Newton's first law:

$$m\frac{d\mathbf{v}}{dt} = q\mathbf{E} \quad \Rightarrow \quad mj\omega\tilde{\mathbf{v}} = q\tilde{\mathbf{E}}$$

where $\tilde{\mathbf{v}}$ denotes the phasor of particle velocity in sinusoidal steady-state. With N electrons per unit volume, each carrying a charge $q = -e$ with a phasor velocity $\tilde{\mathbf{v}}$, we then have a phasor current density

$$\tilde{\mathbf{J}} = Nq\tilde{\mathbf{v}} = Nq(\frac{q\tilde{\mathbf{E}}}{mj\omega}) = -j\frac{Nq^2}{m\omega}\tilde{\mathbf{E}}$$

carried by free electrons only. Positive ions with much larger mass than the electrons will also carry a similar current density, but with a much smaller magnitude given the inverse mass dependence of the expression for $\tilde{\mathbf{J}}$. Hence, a reasonable model for a current density in a collisionless plasma is

$$\tilde{\mathbf{J}} = \sigma\tilde{\mathbf{E}} \text{ with a plasma conductivity } \sigma = -j\frac{Ne^2}{m\omega}$$

where the contribution of ions is neglected.

The crucial result above is that conductivity is purely imaginary — the collisionless plasma is not a resistive but a reactive propagation medium!

In the absence of collisions, kinetic energy of the charge carriers acquired from the wave field is not dissipated (lost) into heat, but instead returned to the wave field much like energy exchange in a circuit consisting of a source and an inductor.

- Since TEM wave propagation in conducting media is the same as prop-

agation in a dielectric with an effective permittivity

$$\epsilon + \frac{\sigma}{j\omega},$$

a collisionless *plasma* with $\epsilon = \epsilon_o$ and conductivity

$$\sigma = -j\frac{Ne^2}{m\omega}$$

can be treated like a dielectric with a permittivity

$$\epsilon_o + \frac{-j\frac{Ne^2}{m\omega}}{j\omega} = \epsilon_o(1 - \frac{\frac{Ne^2}{m\epsilon_o}}{\omega^2}) = \epsilon_o(1 - \frac{\omega_p^2}{\omega^2}),$$

where

$$\omega_p \equiv \sqrt{\frac{Ne^2}{m\epsilon_o}}$$

is known as the **plasma cutoff frequency** (or "plasma frequency" for short).

With the above definitions,

$$\epsilon_r = 1 - \frac{\omega_p^2}{\omega^2} \text{ and } n = \sqrt{\epsilon_r} = \sqrt{1 - \frac{\omega_p^2}{\omega^2}},$$

are, respectively, the **relative permittivity** and **refractive index** in a plasma treated as a dielectric, where we also have $\mu = \mu_o$ (true because except for its free carriers, a plasma is essentially a vacuum).

Hence, in a plasma, TEM waves are described by a **wavenumber**

$$k = \omega\sqrt{\mu\epsilon} = \omega\sqrt{\mu_o\epsilon_o\epsilon_r} \equiv \omega\sqrt{\mu_o\epsilon_o}\sqrt{1-\frac{\omega_p^2}{\omega^2}} = \frac{\omega}{c}\sqrt{1-\frac{\omega_p^2}{\omega^2}},$$

and an **intrinsic impedance**

$$\eta = \sqrt{\frac{\mu}{\epsilon}} = \sqrt{\frac{\mu_o}{\epsilon_o\epsilon_r}} = \frac{\sqrt{\mu_o/\epsilon_o}}{\sqrt{1-\frac{\omega_p^2}{\omega^2}}} = \frac{\eta_o}{\sqrt{1-\frac{\omega_p^2}{\omega^2}}},$$

respectively.

- The collisionless **plasma dispersion relation** for TEM waves, namely

  $$k = \frac{\omega}{c}\sqrt{1-\frac{\omega_p^2}{\omega^2}}$$

  derived above governs the properties of TEM waves to be encountered in a plasma medium.

  It most significantly differs from the free-space dispersion relation

  $$k = \frac{\omega}{c}$$

  by exhibiting a non-linear relationship between ω and k.

  As a consequence,

  1. The **propagation velocity**

     $$v_p = \frac{\omega}{k} = \frac{c}{\sqrt{1-\frac{\omega_p^2}{\omega^2}}} = \frac{c}{n}$$

in a plasma is frequency dependent and a meaningful concept only for $\omega > \omega_p$ when k is real valued (see next).

2. For $\omega < \omega_p$, we find a purely imaginary k in which case e^{-jkz} describes not a propagating wave but an **evanescent** one (for which v_p is not a relevant concept).
3. The plasma **refractive index**

$$n = \frac{c}{v_p} = \sqrt{1 - \frac{\omega_p^2}{\omega^2}}$$

is real valued but less than unity in the propagation regime $\omega > \omega_p$ and it is imaginary in the evanescence region $\omega < \omega_p$.

- Topics that remain to be examined over the next two lectures:

1. The distinction between **phase** and **group velocity** concepts in the regime $\omega > \omega_p$
2. Evanescent plasma waves in the $\omega < \omega_p$ regime and related tunneling phenomena.

- We close this lecture with a brief discussion of plasma frequency ω_p.
- The parameter

$$\omega_p \equiv \sqrt{\frac{Ne^2}{m\epsilon_o}}$$

in the plasma dispersion relation has the dimension of frequency and grows with the square root of the electron density of the plasma. A useful formula for the plasma frequency is

$$\omega_p = 2\pi f_p \quad \text{with} \quad f_p \approx \sqrt{80.6N}$$

where f_p is quoted in Hz units when N is entered in m^{-3} units.

- For example, for $N = 10^{12}\,\mathrm{m}^{-3}$ as in the Earth's ionosphere,

$$f_p \approx \sqrt{80.6 \times 10^{12}} \approx 9 \times 10^6\,\mathrm{Hz} = 9\,\mathrm{MHz}$$

  and

$$\omega_p \approx 18\pi\,\mathrm{Mrad/s}.$$

- A plasma frequency of $f_p \approx 9$ MHz will have a severe impact on EM waves in the ionosphere when the wave frequency $f = \frac{\omega}{2\pi}$ is close to 9 MHz.
- The effect of the plasma on the EM wave will be negligible in the ionosphere only when f is many orders of magnitude larger than 9 MHz.

- The plasma frequency ω_p also has a direct interpretation in terms of plasma dynamics:

  - If all the electrons in a volume of plasma were pulled to one side of the volume, away from the positive ions within the same volume

— in analogy to a stretched spring — and then let go, the electron and ion populations would rush toward one another (because of electrostatic attraction) and then overshoot (because of inertia) and reverse their motions to establish a perpetual oscillation at the frequency f_p!

– The plasma frequency f_p is a resonance frequency of the plasma seen as an elastic body.

Example 1: Consider an infinite homogeneous plasma with a plasma frequency of $f_p = 10$ MHz. Determine the wavelength or the penetration depth — whichever is relevant — of a TEM wave in the plasma produced by an infinite current sheet located at $x = 0$ plane, if the oscillation frequency of the current density is (a) $f = 20$ MHz, and (b) $f = 5$ MHz. Also comment whether the TEM wave is propagating or evanescent.

Solution: (a) We have, for $f = 20$ MHz

$$\begin{aligned} k &= \frac{\omega}{c}\sqrt{1-\frac{\omega_p^2}{\omega^2}} = \frac{2\pi \times 20 \times 10^6}{3 \times 10^8}\sqrt{1-(\frac{10}{20})^2} \\ &= \frac{40\pi}{300}\sqrt{1-\frac{1}{4}} = \frac{\sqrt{3}40\pi}{600} = \frac{\sqrt{3}\pi}{15}\,\frac{\text{rad}}{\text{m}}. \end{aligned}$$

Hence, the TEM wave produced by the current sheet is "propagating" in that case (away from $x = 0$ surface on both sides) and its wavelength is

$$\lambda = \frac{2\pi}{k} = \frac{2\pi}{\frac{\sqrt{3}\pi}{15}} = \frac{30}{\sqrt{3}} = 10\sqrt{3}\,\text{m}.$$

(b) For $f = 5$ MHz

$$\begin{aligned} k &= \frac{\omega}{c}\sqrt{1-\frac{\omega_p^2}{\omega^2}} = \frac{2\pi \times 5 \times 10^6}{3\times 10^8}\sqrt{1-(\frac{10}{5})^2} \\ &= \frac{10\pi}{300}\sqrt{1-4} = \pm j\frac{\sqrt{3}10\pi}{600} = \pm j\frac{\sqrt{3}\pi}{60}\,\frac{\text{rad}}{\text{m}} \end{aligned}$$

where the sign giving rise to the decaying wave away from its source should be employed. In this case the TEM wave is "evanescent" and its attenuation constant is

$$|k| = \frac{\sqrt{3}\pi}{60}\,\frac{\text{Np}}{\text{m}}.$$

The corresponding penetration distance (distance over which the wave amplitude is reduced by an exponential factor of e^{-1}) is

$$\delta = \frac{1}{|k|} = \frac{60}{\sqrt{3}\pi} = \frac{\sqrt{3}}{\pi}20\,\text{m}.$$

Example 2: In Example 1, part (b) what is the attenuation rate of the evanescent wave in units of dB/m?

Solution: The evanescent field in Example 1 part (b) will vary as

$$\tilde{\mathbf{E}}(x) = \hat{p}E_o e^{-|k|x}$$

in $x > 0$ region, where $\hat{p}$ is a unit vector perpendicular to $\hat{x}$, E_o is the field strength at $x = 0$. Also,

$$|k| = \frac{\sqrt{3}\pi}{60}\,\frac{\text{Np}}{\text{m}}.$$

from the solution of part (b). Therefore, we have

$$\frac{|\tilde{\mathbf{E}}(x=0)|}{|\tilde{\mathbf{E}}(x=1)\,\mathrm{m}|} = \frac{1}{e^{-|k|}} = e^{|k|}$$

and

$$\begin{aligned} 20\log_{10}\frac{|\tilde{\mathbf{E}}(x=0)|}{|\tilde{\mathbf{E}}(x=1)\,\mathrm{m}|} &= 20\log_{10} e^{|k|} = |k|20\log_{10} e \\ &= \frac{\sqrt{3}\pi}{60}\,20\log_{10} e = \frac{\pi}{\sqrt{3}}\log_{10} e \approx \frac{\pi}{\sqrt{3}}0.434 \\ &\approx 0.788\,\frac{\mathrm{dB}}{\mathrm{m}} \end{aligned}$$

which is the attenuation rate of the evanescent field in dB/m units.

Attenuation in dB/m is expressed as the 20 log of the amplitude ratio across a one meter distance (as we have done above), or, equivalently, as 10 log of the "power" ratio across a one meter distance.

25 Phase and group velocities and delays

- Propagation velocity

$$v_p = \frac{\omega}{k}$$

of a co-sinusoid field component

$$\cos(\omega t - kz) \quad \Leftrightarrow \quad e^{-jkz}$$

is also known as **phase velocity**, because v_p as defined above, corresponds to the speed with which constant phase points (e.g., zero-crossings of the field) move.

- If the phase velocity is ω dependent — as in dispersive media — then field components (e.g., E_x, H_y, etc.), which are the superpositions of co-sinusoids with different frequencies (two, three, several, countless), can in general be described in terms of an **envelope** function and a **carrier** function (recall AM modulation from ECE 210), each having its own and distinct velocity.

  - The propagation velocity of the envelope is known as **group velocity** and it can be calculated as

$$v_g = \frac{\partial \omega}{\partial k}$$

once the dispersion relation relating k to ω is available.

– The propagation velocity of the carrier is simply a phase velocity

$$v_p = \frac{\omega}{k},$$

where we use the carrier frequency ω_o for frequency ω, and the carrier wavenumber k_o for wavenumber k as illustrated below.

A simple example: Consider the superposition

$$f(z,t) = \cos(\omega_1 t - k_1 z) + \cos(\omega_2 t - k_2 z)$$

where wavenumbers k_1 and k_2 depend on frequencies $\omega_1$and ω_2 as described by some dispersion relation (e.g., the plasma dispersion relation). Using some trig identities we can re-write $f(z,t)$ as

$$f(z,t) = 2\cos(\frac{\Delta\omega}{2}t - \frac{\Delta k}{2}z)\cos(\omega_o t - k_o z)$$

where

$$\omega_o \equiv \frac{\omega_1 + \omega_2}{2}, \;\; k_o \equiv \frac{k_1 + k_2}{2}, \;\; \Delta\omega = \omega_2 - \omega_1, \;\; \Delta k = k_2 - k_1.$$

– **Verification:** Given the above definitions,

$$\omega_{1,2} = \omega_o \mp \frac{\Delta\omega}{2} \;\text{ and }\; k_{1,2} = k_o \mp \frac{\Delta k}{2},$$

and so

$$f(z,t) \;\; = \;\; \cos(\omega_1 t - k_1 z) + \cos(\omega_2 t - k_2 z)$$

$$
\begin{aligned}
&= \cos(\omega_o t - k_o z + \frac{\Delta\omega t - \Delta k z}{2}) + \cos(\omega_o t - k_o z - \frac{\Delta\omega t - \Delta k z}{2}) \\
&= \cos(\omega_o t - k_o z)\cos(\frac{\Delta\omega t - \Delta k z}{2}) - \sin(\omega_o t - k_o z)\sin(\frac{\Delta\omega t - \Delta k z}{2}) \\
&+ \cos(\omega_o t - k_o z)\cos(\frac{\Delta\omega t - \Delta k z}{2}) + \sin(\omega_o t - k_o z)\sin(\frac{\Delta\omega t - \Delta k z}{2}) \\
&= \underbrace{2\cos(\frac{\Delta\omega}{2}t - \frac{\Delta k}{2}z)}_{\text{envelope}}\ \underbrace{\cos(\omega_o t - k_o z)}_{\text{carrier}}.
\end{aligned}
$$

In this simplest possible example of superpositioned co-sinusoids (simplest because we only used two components instead of many), both the envelope function and the carrier function are co-sinusoids.

Assuming that $\Delta\omega \ll \omega_1, \omega_2$, the carrier function $\cos(\omega_o t - k_o z)$ is a co-sinusoid within the same "frequency band" as the superposed co-sinusoids, while the envelope function $2\cos(\frac{\Delta\omega t - \Delta k z}{2})$ is a low-frequency co-sinusoid residing (in frequency space) outside the signal band.

With that distinction in mind, we identify the propagation velocities of the carrier and envelope functions as the phase and group velocities of composite waveform $f(z,t)$ — the phase velocity (describing the carrier motion) is

$$v_p = \frac{\omega_o}{k_o},$$

whereas the group velocity (describing the envelope motion) is

$$v_g = \frac{\Delta\omega}{\Delta k} = \frac{\omega_2 - \omega_1}{k_2 - k_1}.$$

Example 1: Consider the case

$$\Delta\omega = \frac{\omega_o}{10} \quad \text{and} \quad \Delta k = \frac{k_o}{5}.$$

In that case

$$v_g = \frac{\Delta\omega}{\Delta k} = \frac{\omega_o/10}{k_o/5} = \frac{1}{2}\frac{\omega_o}{k_o} = \frac{1}{2}v_p,$$

a waveform with half as large a group velocity as the phase velocity — in such a waveform, the zero-crossings of the carrier will march through the envelope as demonstrated by an animation on the web site.

Example 2: Determine the group velocity

$$v_g = \frac{\Delta\omega}{\Delta k}$$

of the sum of two co-sinusoidal waves propagating in z direction if $\omega_1 = 99$ rad/s, $\omega_2 = 101$ rad/s and the dispersion relation is

$$\omega = k^2.$$

Solution: We can solve this problem by first obtaining $k_{1,2} = \sqrt{\omega_{1,2}}$, and then dividing

$$\Delta\omega = \omega_2 - \omega_1$$

by

$$\Delta k = \sqrt{\omega_2} - \sqrt{\omega_1}.$$

Alternatively, we can approximate $\Delta\omega/\Delta k$ by the

$$\text{partial derivative } \partial\omega/\partial k \text{ evaluated at } \omega_o = 100\,\text{rad/s}$$

which is at the center of the frequency band flanked by ω_1 and ω_2.

Both approaches will give about the same result since $\Delta\omega \ll \omega_o$ and the slope $\partial\omega/\partial k$ of the ω versus k curve at $\omega = \omega_o$ is nearly the same as the ratio $\Delta\omega/\Delta k$.

Using the second method, we note

$$\omega = k^2 \;\;\Rightarrow\;\; \frac{\partial\omega}{\partial k} = 2k = 2\sqrt{\omega}.$$

Therefore, the group velocity of the sum is

$$v_g = \frac{\Delta\omega}{\Delta k} \approx \frac{\partial\omega}{\partial k} = 2\sqrt{\omega} = 20\,\frac{\text{m}}{\text{s}}$$

after evaluating at $\omega = \omega_o = 100$ rad/s. You should compare our result above with the exact value

$$\frac{\Delta\omega}{\Delta k} = \frac{\omega_2 - \omega_1}{\sqrt{\omega_2} - \sqrt{\omega_1}}$$

to convince yourself that both approaches give approximately the same result.

Example 3: What is the phase velocity of the sum signal in Example 2.

Solution: The phase velocity of the sum signal is

$$v_p = \frac{\omega_o}{k_o}$$

where

$$\omega_o \equiv \frac{\omega_1 + \omega_2}{2}, \;\; k_o \equiv \frac{k_1 + k_2}{2}, \;\; k = \sqrt{\omega}.$$

This is well approximated by

$$v_p = \frac{\omega_o}{\sqrt{\omega_o}} = \frac{100}{10} = 10 \, \frac{\mathrm{m}}{\mathrm{s}}.$$

The exact value can be obtained as

$$v_p = \frac{\omega_o}{k_o} = \frac{\omega_1 + \omega_2}{k_1 + k_2} = \frac{\omega_1 + \omega_2}{\sqrt{\omega_1} + \sqrt{\omega_2}}.$$

- Practical signals used in communication applications are more complicated than just the sum of two-co-sinusoids. In general, we are concerned with the superposition of a continuum of co-sinusoids over finite frequency bands $\Delta\omega$. How do we then define the wave envelope and the carrier in such cases and extend the notion of phase and group velocities introduced above? This question is addressed next:

- Consider a sum of many monochromatic waves of frequencies ω_m in a band $\Delta\omega$ centered about a frequency ω_o — such a sum can be represented as

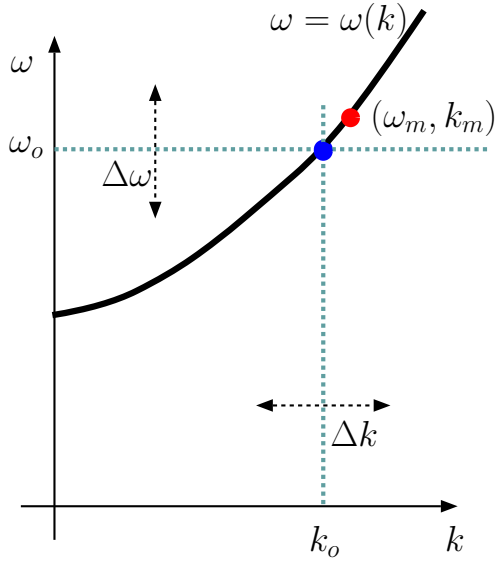
$$\sum_m \mathrm{Re}\{F_m e^{j(\omega_m t - k_m z)}\}$$

where F_m are the individual wave amplitudes. Introducing

$$\omega_m = \omega_o + \Delta\omega_m \;\text{ and }\; k_m = k_o + \Delta k_m,$$

we can re-write the same sum as

$$\mathrm{Re}\{e^{j(\omega_o t - k_o z)} \sum_m F_m e^{j\Delta\omega_m (t - \frac{z}{\Delta\omega_m/\Delta k_m})}\}.$$



Suppose that the band of frequencies $\Delta\omega$ containing all the components ω_m is sufficiently small so that the ratio

$$\frac{\Delta\omega_m}{\Delta k_m} \approx \frac{\partial\omega}{\partial k}_{|\omega=\omega_o}$$

is independent of index m.

In that case the sum above reduces to

$$\mathrm{Re}\{e^{j(\omega_o t - k_o z)} f(t - \frac{z}{v_g})\}$$

with

$$f(t) = \sum_m F_m e^{j\Delta\omega_m t} \quad \text{Envelope function}$$

and

$$v_g = \frac{\partial \omega}{\partial k}_{|\omega=\omega_o} \quad \text{Group velocity.}$$

- The above result indicates that a wave signal

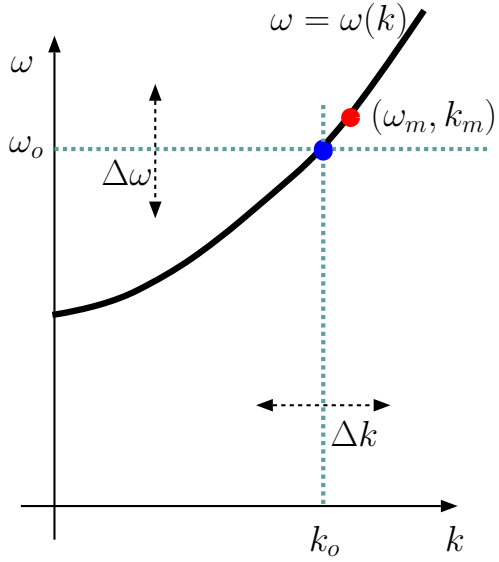
$$s(0,t) = \mathrm{Re}\{e^{j\omega_o t} f(t)\} = |f(t)| \cos(\omega_o t + \angle f(t))$$

observed at a location $z = 0$ will be observed at an arbitrary $z > 0$ as

$$s(z,t) = \mathrm{Re}\{e^{j(\omega_o t - k_o z)} f(t - \frac{z}{v_g})\} = |f(t - \frac{z}{v_g})| \cos(\omega_o t - k_o z + \angle f(t - \frac{z}{v_g})).$$

Such a signal[1] would be called an

1. **AM signal** for the case $\angle f(t) = 0$ — purely real $f(t)$, requiring $F_{-m} = F_m^*$ for $\Delta\omega_m = m\delta\omega$, with $m = 0, \pm 1, \pm 2, \cdots$,
2. **Phase modulated (PM) signal** for $|f(t)| = \text{const.}$— $f(t) \propto e^{j\phi(t)}$, with $|F_m| \ll F_0$ and $F_{-m} = -F_m^*$ for $m = \pm 1, \pm 2, \cdots$.



The important point is, the modulation $|f(t)|$ and/or $\angle f(t)$ of the AM and/or PM signal will travel with the group velocity

$$v_g = \frac{\partial \omega}{\partial k}_{|\omega=\omega_o}.$$

[1]Also, the same results apply in the continuum limit of $\delta\omega \to 0$ in which case the sum defining $f(t)$ in terms of Fourier coefficients F_m reduce to an integral defining $f(t)$ in terms its Fourier transform $F(\omega)$.

- The propagation of **narrowband signals** for which the above derivation of v_g is well justified — bandwidth $\Delta\omega \ll \omega_o$ — is therefore well described by the phase velocity (for the carrier) and the group velocity (for the modulation envelope and/or phase) concepts.

  - However, for **broadband signals** where $\Delta\omega \sim \omega_o$ the constancy of

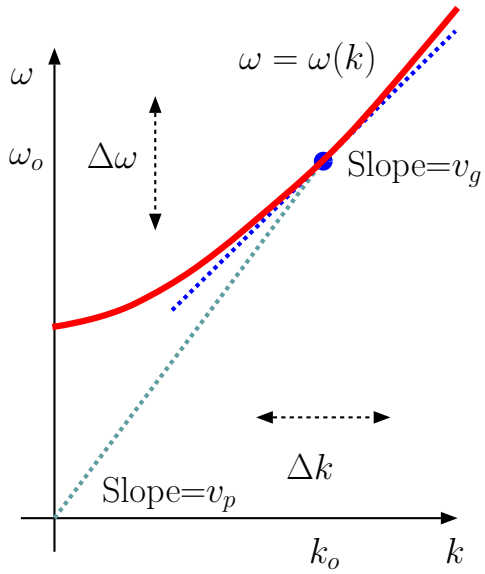
    $$\frac{\Delta\omega_m}{\Delta k_m}$$

    over the entire frequency band $\Delta\omega$ will not hold, and it may be necessary to define a set of frequency dependent group velocities associated with sub-bands of $\Delta\omega$ (see HW).

- In general, the computation of the phase and group velocities of narrowband signals requires the knowledge of pertinent dispersion relation, the algebraic relationship between the wave frequency ω and wavenumber k.

- When the dispersion relation is known, it is useful to display it in the form of a ω versus k plot as shown in the margin.

  - If the plot is a straight line then the waves are dispersionless and $v_g = v_p$.
  - However, if the plot is curved (like shown in the margin), then the waves are dispersive and the phase and group velocities v_p and v_g need to be computed separately.



Note that:

Phase velocity v_p **is the slope of the line from the origin to the dispersion curve at the band center.**

Group velocity v_g **is the slope of the dispersion curve itself at the band center.**

- The dispersion relation

$$k = \frac{\omega}{c}\sqrt{1 - \frac{\omega_p^2}{\omega^2}} \quad \Rightarrow \quad c^2k^2 = \omega^2 - \omega_p^2 \quad \Rightarrow \quad \omega = \sqrt{c^2k^2 + \omega_p^2}$$

for the collisionless plasma has a dispersion curve resembling the one shown in the margin — waves in a plasma are clearly dispersive.

To obtain the plasma group velocity we take the partial derivative of the plasma dispersion formula

$$c^2k^2 = \omega^2 - \omega_p^2$$

on both sides with respect to variable k, which leads to

$$\frac{\partial}{\partial k}(c^2k^2 = \omega^2 - \omega_p^2) \quad \Rightarrow \quad c^2 2k = 2\omega\frac{\partial\omega}{\partial k} \quad \Rightarrow \quad \frac{\omega}{k}\frac{\partial\omega}{\partial k} = c^2.$$

Since

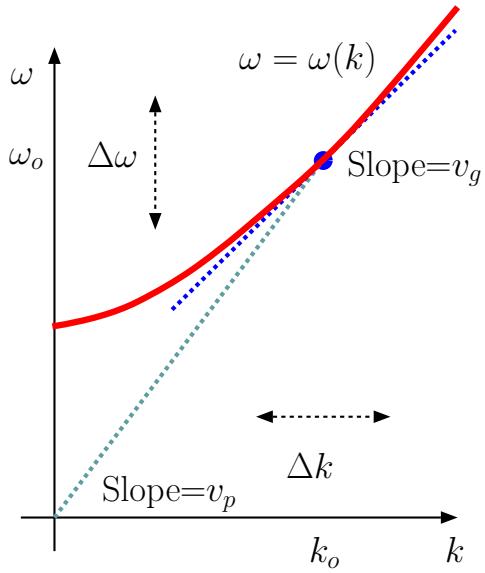
$$v_g = \frac{\partial\omega}{\partial k} \quad \text{and} \quad v_p \equiv \frac{\omega}{k},$$

this result indicates that in a plasma

$$v_g v_p = c^2$$

with the explicit formulas of

$$v_p = \frac{\omega}{k} = \frac{c}{\sqrt{1 - \frac{\omega_p^2}{\omega^2}}} \quad \text{and} \quad v_g = c\sqrt{1 - \frac{\omega_p^2}{\omega^2}}.$$



Note that:

Phase velocity v_p **is the slope of the line from the origin to the dispersion curve at the band center.**

Group velocity v_g **is the slope of the dispersion curve itself at the band center.**

Note that the phase velocity in the plasma exceeds c at all $\omega > \omega_p$, whereas the group velocity is bounded by c.

Einstein's speed limit of c for motions in the universe is only meant to apply to energy, mass, and information transport — that list does not include the phase velocity, since the phase velocity of an unmodulated carrier is not pertinent for the transfer of energy or mass or information across space.

  - A distant light bulb can be lit by sending an electric field pulse with an envelope which will travel the intervening distance at the group velocity of the propagation medium. The light bulb gets turned on only after the pulse envelope arrives at its location, independent of how fast (or slow) the pulse carrier moves. Energy moves with the group velocity[2].

- In general, depending on the dispersion relation, it is possible to have $v_g < v_p$ as well as $v_g > v_p$.

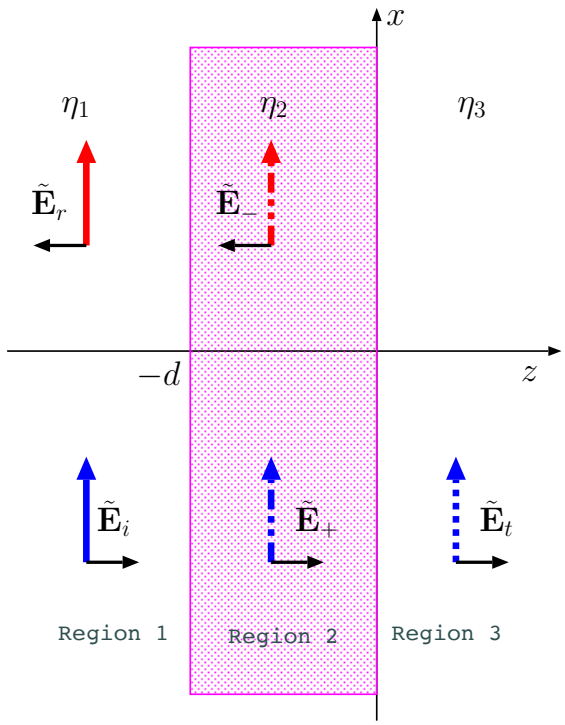
- Also, as just mentioned, it is possible to have $v_p < c$ as well as $v_p > c$ (as dictated by the relevant dispersion relation).

- However, $v_g > c$ is never possible for any wave motion — if a group velocity calculation indicates $v_g > c$ in some setting, you can be sure that the dispersion relation used for v_g calculation is invalid in that

[2]A rigorous proof that energy is transported with velocity v_g in linear and dispersive media can be found in Bers, *Am. J. Phys.*, **68**, 482 (2000).

setting and/or the dispersion curve has a shape that precludes the applicability of the narrowband signal model developed in this lecture.

26 Evanescent waves and tunneling



- In this lecture we will explore the tunneling phenomenon associated with evanescent waves established within finite-width regions.

  The multi-slab tunneling result to be derived in this lecture will:

  1. Enhance our qualitative understanding of the frustrated-TIR example shown back in Lecture 19,
  2. Illustrate a methodology based on ***transmission line analogies*** to be used in forthcoming lectures on waveguides.

- Consider the three-slab geometry depicted in the margin where a TEM wave field

$$\tilde{\mathbf{E}}_i = \hat{x} E_i e^{-jk_1 z}, \text{ accompanied by } \tilde{\mathbf{H}}_i = \hat{y} \frac{E_i}{\eta_1} e^{-jk_1 z},$$

  is incident from the left in the region $z < -d$ (region 1). As a response a reflected wave

$$\tilde{\mathbf{E}}_r = \hat{x} E_r e^{jk_1 z}, \text{ accompanied by } \tilde{\mathbf{H}}_r = -\hat{y} \frac{E_r}{\eta_1} e^{jk_1 z},$$

  is set up in the same region, as well as

$$\tilde{\mathbf{E}}_+ = \hat{x} E_+ e^{-jk_2 z}, \text{ accompanied by } \tilde{\mathbf{H}}_+ = \hat{y} \frac{E_+}{\eta_2} e^{-jk_2 z},$$

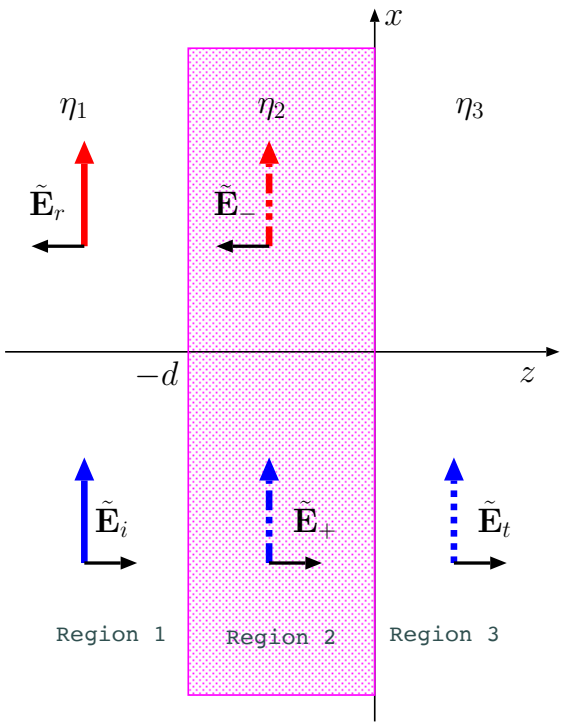
  and

$$\tilde{\mathbf{E}}_- = \hat{x} E_- e^{jk_2 z}, \text{ accompanied by } \tilde{\mathbf{H}}_- = -\hat{y} \frac{E_-}{\eta_2} e^{jk_2 z},$$

in the region $-d < z < 0$ (region 2). Finally, in region $z > 0$, we will have

$$\tilde{\mathbf{E}}_t = \hat{x} E_t e^{-jk_3 z}, \text{ accompanied by } \tilde{\mathbf{H}}_t = \hat{y} \frac{E_t}{\eta_3} e^{-jk_3 z}.$$

- Our aim is to determine the amplitudes E_t, E_+, E_-, E_r in terms of E_i using tangential boundary conditions at $z = -d$ and $z = 0$.
- We are in particular interested in the ratio of the transmitted power in region 3 to the incident power in region 1 as a function of slab width d as well as the refractive indices n_1, n_2, and n_3, including the case when n_2 is purely imaginary, the case corresponding to region 2 being in evanescent mode.



- Starting with the boundary at $z = 0$, the continuity of tangential $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{H}}$ across the boundary requires that

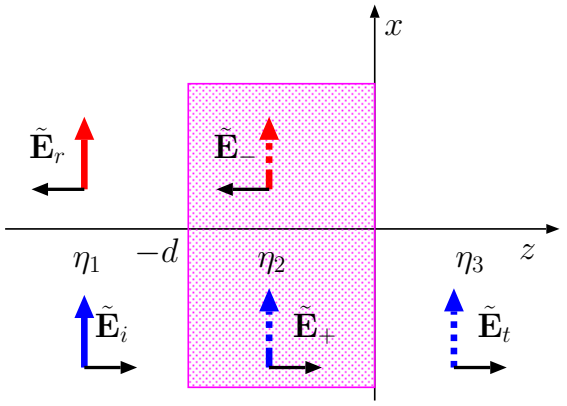
$$E_+ + E_- = E_t \text{ and } \frac{E_+ - E_-}{\eta_2} = \frac{E_t}{\eta_3}.$$

These equations can be solved for E_t and E_- in terms of E_+ to obtain

$$E_t = \underbrace{\frac{2\eta_3}{\eta_3 + \eta_2}}_{\tau_{32}} E_+ \text{ and } E_- = \underbrace{\frac{\eta_3 - \eta_2}{\eta_3 + \eta_2}}_{\Gamma_{32}} E_+.$$

Note that we have defined a pair of coefficients representing the interaction at $z = 0$ interface: a **transmission coefficient** τ_{32} and a

reflection coefficient Γ_{32} in terms of intrinsic impedances η_3 and η_2 in a manner analogous to similar relations seen in our studies of *transmission line* (TL) systems (in ECE 329).



- At the boundary on $z = -d$ plane the continuity of tangential $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{H}}$ requires that

$$E_i e^{jk_1 d} + E_r e^{-jk_1 d} = E_+ e^{jk_2 d} + E_- e^{-jk_2 d}$$

and

$$\frac{E_i e^{jk_1 d} - E_r e^{-jk_1 d}}{\eta_1} = \frac{E_+ e^{jk_2 d} - E_- e^{-jk_2 d}}{\eta_2}$$

respectively. To utilize these relations in a close analogy to TL problems we next define an effective **field impedance** $Z(-d)$ for the $z = -d$ plane as

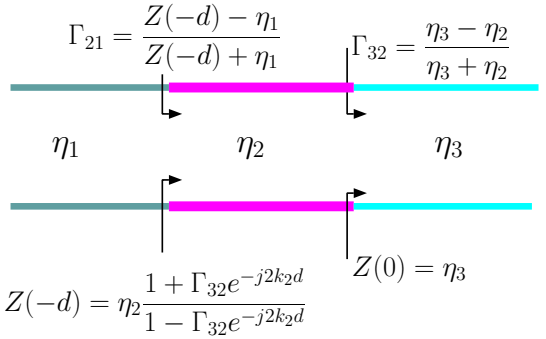
$$Z(-d) = \frac{E_i e^{jk_1 d} + E_r e^{-jk_1 d}}{\frac{E_i e^{jk_1 d} - E_r e^{-jk_1 d}}{\eta_1}} = \eta_1 \frac{1 + \Gamma_{21}}{1 - \Gamma_{21}} \quad \text{where} \quad \Gamma_{21} \equiv \frac{E_r e^{-jk_1 d}}{E_i e^{jk_1 d}}.$$

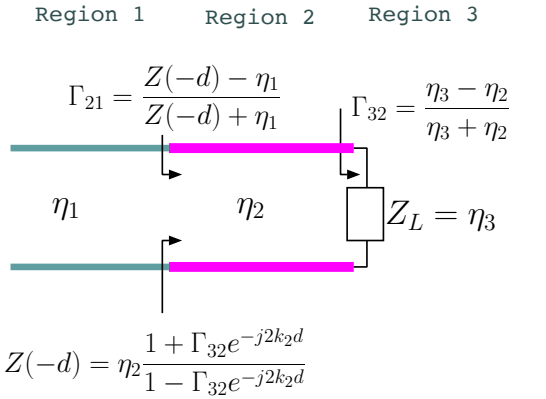
But, by the boundary condition equations above it is also true that

$$Z(-d) \equiv \frac{E_+ e^{jk_2 d} + E_- e^{-jk_2 d}}{\frac{E_+ e^{jk_2 d} - E_- e^{-jk_2 d}}{\eta_2}} = \eta_2 \frac{1 + \frac{E_- e^{-jk_2 d}}{E_+ e^{jk_2 d}}}{1 - \frac{E_- e^{-jk_2 d}}{E_+ e^{jk_2 d}}} = \eta_2 \frac{1 + \Gamma_{32} e^{-j2k_2 d}}{1 - \Gamma_{32} e^{-j2k_2 d}}.$$

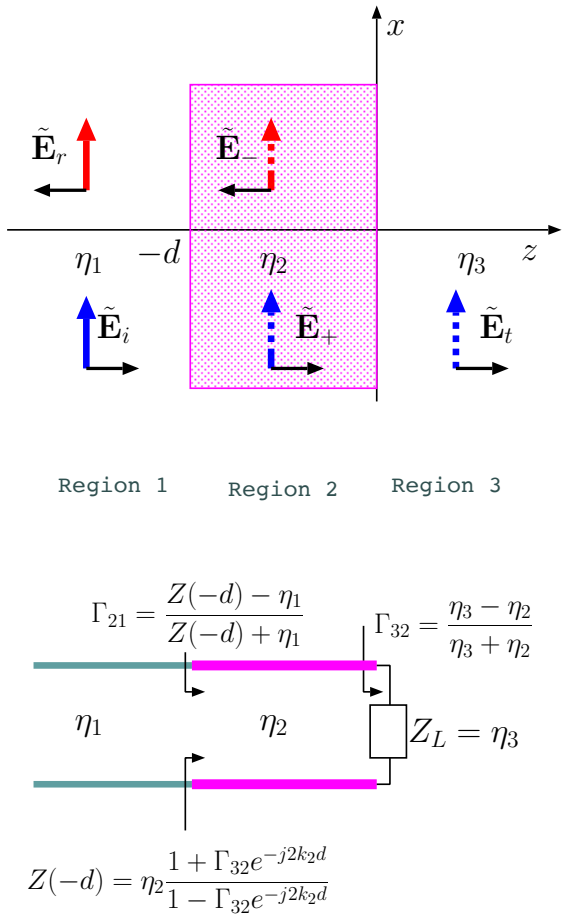
Solving the above expression for the **reflection coefficient** Γ_{21} at $z = -d$ plane in terms of impedance $Z(-d)$ we find that

$$\Gamma_{21} = \frac{Z(-d) - \eta_1}{Z(-d) + \eta_1}.$$





- The parameters Γ_{32}, $Z(-d)$, and Γ_{21} introduced above, bearing a strong analogy to an equivalent TL problem suggested in the margin, are sufficient to calculate the reflected and transmitted powers in our multiple slab problem as follows:



1. We first note that

$$|\Gamma_{21}|^2 = |\frac{E_r e^{-jk_1 d}}{E_i e^{jk_1 d}}|^2 \quad \Rightarrow \quad \frac{\langle S_r \rangle}{\langle S_i \rangle} = \frac{|E_r|^2/2\eta_1}{|E_i|^2/2\eta_1} = |\Gamma_{21}|^2$$

gives the **reflectance**, the fraction of the time-averaged incident power density reflected by the slab discontinuity back into region 1.

2. Assuming that the slab in region 2 is lossless, the **transmittance**, the time-averaged power density transmitted into the region 3 has to be

$$\langle S_t \rangle = \langle S_i \rangle - \langle S_r \rangle = \langle S_i \rangle (1 - |\Gamma_{21}|^2) \quad \Rightarrow \quad \frac{\langle S_t \rangle}{\langle S_i \rangle} = \frac{|E_t|^2/2\eta_3}{|E_i|^2/2\eta_1} = 1 - |\Gamma_{21}|^2.$$

The upshot is

$$\frac{\langle S_r \rangle}{\langle S_i \rangle} = |\Gamma_{21}|^2 \text{ and } \frac{\langle S_t \rangle}{\langle S_i \rangle} = 1 - |\Gamma_{21}|^2$$

where

$$\Gamma_{21} = \frac{Z(-d) - \eta_1}{Z(-d) + \eta_1}, \quad Z(-d) = \eta_2 \frac{1 + \Gamma_{32} e^{-j2k_2 d}}{1 - \Gamma_{32} e^{-j2k_2 d}}, \quad \Gamma_{32} = \frac{\eta_3 - \eta_2}{\eta_3 + \eta_2}$$

in analogy with an equivalent TL problem. An extension of these relations to an n-slab configuration is straightforward.

Example 1: Assume that regions 1 and 3 are free space whereas region 2 is a plasma slab of some width d and a plasma frequency f_p. Determine and plot the transmittance

$$\frac{\langle S_t \rangle}{\langle S_i \rangle} = 1 - |\Gamma_{21}|^2$$

as a function of d if (a) $f = \frac{5}{4} f_p$, and (b) $f = \frac{4}{5} f_p$.

Solution: (a) In this case the plasma refractive index in the slab is

$$n_2 = \sqrt{1 - \frac{f_p^2}{f^2}} = \sqrt{1 - (\frac{4}{5})^2} = \sqrt{1 - \frac{16}{25}} = \frac{3}{5}.$$

Hence, with $\eta_1 = \eta_3 = \eta_o$ and $\eta_2 = \eta_o / n_2 = 5\eta_o/3$, we have

$$\Gamma_{32} = \frac{\eta_o - \frac{5}{3}\eta_o}{\eta_o + \frac{5}{3}\eta_o} = \frac{3-5}{3+5} = -\frac{2}{8} = -0.25$$

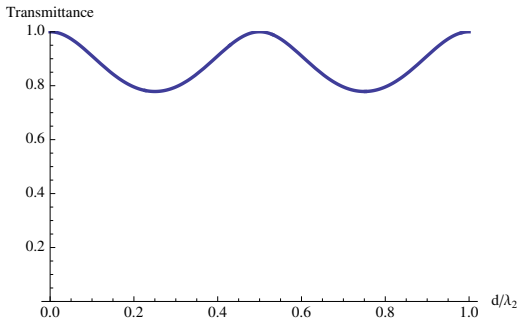
also, with real $k_2 = \frac{2\pi}{\lambda_2}$, we have

$$Z(-d) = \eta_2 \frac{1 + \Gamma_{32} e^{-j2k_2 d}}{1 - \Gamma_{32} e^{-j2k_2 d}} = \frac{5}{3}\eta_o \frac{1 - 0.25 e^{-j4\pi \frac{d}{\lambda_2}}}{1 + 0.25 e^{-j4\pi \frac{d}{\lambda_2}}}.$$

A plot of the transmittance $1 - |\Gamma_{21}|^2$ versus d/λ_2, where

$$\Gamma_{21} = \frac{Z(-d) - \eta_1}{Z(-d) + \eta_1}$$

Transmittance curve for part (a) when region 2 is in propagation mode:



and produced by a Mathematica code of these equations is shown in the margin. The transmittance shows a $\lambda_2/2$ periodicity in slab width d in consistency with the periodicity expected for lossless TL systems.

Solution: (b) In this case the plasma refractive index in the slab is

$$n_2 = \sqrt{1 - \frac{f_p^2}{f^2}} = \sqrt{1 - (\frac{5}{4})^2} = \sqrt{1 - \frac{25}{16}} = \sqrt{-\frac{9}{16}} = \pm j\frac{3}{4}$$

Hence, with $\eta_1 = \eta_3 = \eta_o$ and $\eta_2 = \eta_o/n_2 = j\frac{4}{3}\eta_o$, where we have used $n_2 = -j\frac{3}{4}$, we obtain

$$\Gamma_{32} = \frac{\eta_o - j\frac{4}{3}\eta_o}{\eta_o + j\frac{4}{3}\eta_o} = \frac{3-4j}{3+4j}$$

having a unity magnitude. Also, $k_2 = kn_2 = -j3k/4$, and we have

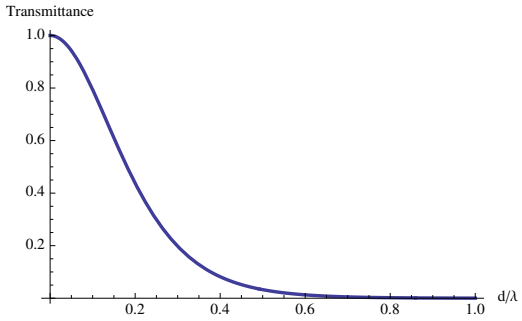
$$Z(-d) = \eta_2 \frac{1 + \Gamma_{32}e^{-j2k_2d}}{1 - \Gamma_{32}e^{-j2k_2d}} = j\frac{4}{3}\eta_o \frac{1 + \frac{3-4j}{3+4j}e^{-3\pi d/\lambda}}{1 - \frac{3-4j}{3+4j}e^{-3\pi d/\lambda}}.$$

A plot of transmittance $1 - |\Gamma_{21}|^2$ versus d/λ, where

$$\Gamma_{21} = \frac{Z(-d) - \eta_1}{Z(-d) + \eta_1}$$

and produced by a Mathematica code of these equations is shown in the margin. Note the strong **tunneling effect** at small d/λ. The choice of $n_2 = j\frac{3}{4}$ would also have resulted in the same plot even though the expressions for Γ_{32} and k_2 would have been changed.

Transmittance curve for part (b) when region 2 is in evanescence mode:



Note that adjusting d/λ to about 0.2 sets the transmittance as 1/2, creating in effect a "beam splitter" in reference to our discussions of prisms and tunneling in Lecture 24.

- A fascinating aspect of tunneling is:
  - even though the time-averaged Poynting vectors — i.e., the avg power densities — associated with the evanescent wave fields $\tilde{\mathbf{E}}_+$ and $\tilde{\mathbf{E}}_-$ in region 2 are individually *zero* because of the 90° phase shift between

$$\tilde{\mathbf{E}}_+ \text{ and } \tilde{\mathbf{H}}_+ \quad \text{as well as} \quad \tilde{\mathbf{E}}_- \text{ and } \tilde{\mathbf{H}}_-,$$

    the time-averaged Poynting vector associated with $\tilde{\mathbf{E}}_+ + \tilde{\mathbf{E}}_-$, i.e.,

$$\frac{1}{2}\text{Re}\{(\tilde{\mathbf{E}}_+ + \tilde{\mathbf{E}}_-) \times (\tilde{\mathbf{H}}_+ + \tilde{\mathbf{H}}_-)^*\}$$

    pertinent for region 2, is (as shown in HW) *non-zero* (and independent of position within region 2) because of the *non-zero cross term contributions* between

$$\tilde{\mathbf{E}}_+ \text{ and } \tilde{\mathbf{H}}_- \quad \text{as well as} \quad \tilde{\mathbf{E}}_- \text{ and } \tilde{\mathbf{H}}_+.$$

    - By contrast, in propagating regions (i.e., non-evanescent), the cross product terms cancel while "self product" terms determine the net average Poynting vector.

There are many practical implications and applications of tunneling:

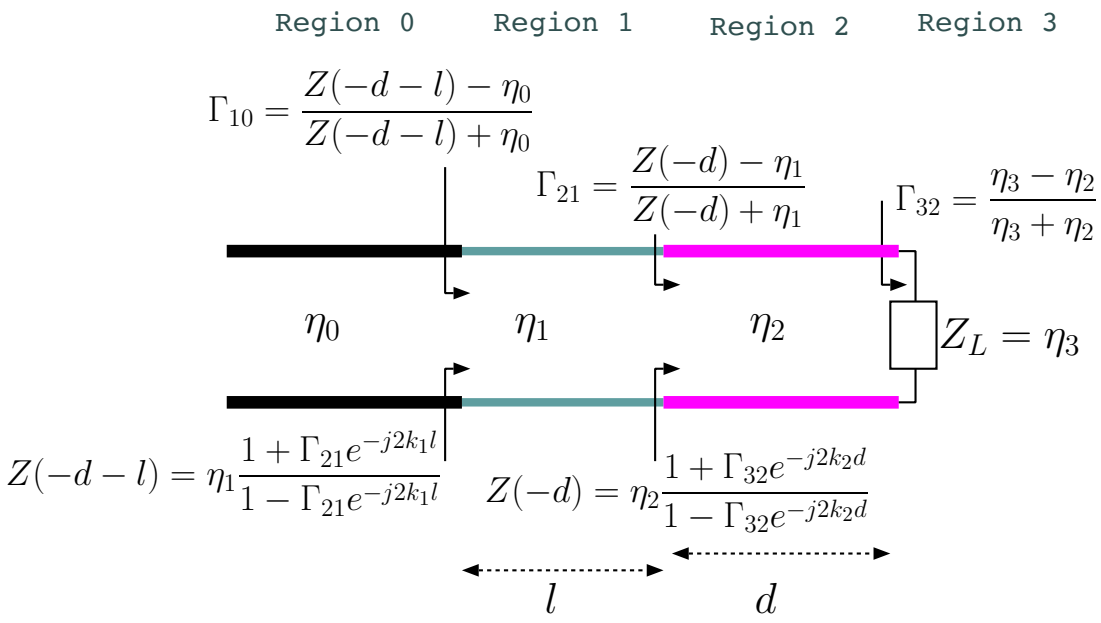
- Beam splitters, attenuators, (undesired) interference effect due to coupling of nearby systems ...

Quantum mechanical tunneling:

In quantum physics one talks about *probabilities* of encountering particles in a given physical system rather than the particle trajectories; furthermore, the probabilities are calculated as magnitude squares (like the average power) of "wave functions" obeying a wave equation (e.g., *Schroedinger's equation* in case of non-relativistic particles). Since waves in general (including Schrodinger waves) can exhibit tunneling properties across evanescent regions (as shown in this section), finite probabilities can be calculated in quantum mechanics for particles in regions separated from their source regions by classically impenetrable barriers (in which the wave function is evanescent).

Phenomena such as radioactive decay or Ohmic contacts (in metal semi-conductor junctions) can be explained in terms of quantum mechanical tunneling, a counterpart of electromagnetic tunneling studied in this section. Also, quantum mechanical tunneling is fundamental to the operation of "scanning tunneling microscopes" used to image atoms and crystals.

- The transmission line analogy to solve a four-slab problem:

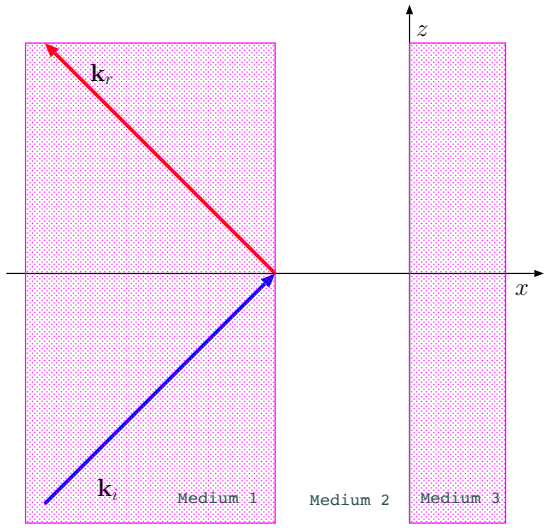


  - The relations shown on the diagram can be used to calculate the transmittance $1-|\Gamma_{10}|^2$ from region 0 to region 3 assuming that regions 1 and 2 are lossless.

Example 2: If in the above diagram region 3 is evanescent what would be the transmittance $1-|\Gamma_{10}|^2$?

Answer: In that case the transmittance should be zero (and reflectance unity) corresponding to a purely reactive $Z_L = \eta_3$!

Tunneling at oblique incidence



- Our analysis of tunneling and frustrated-TIR at oblique incidence will amount to analyzing the three-slab geometry shown in the margin with interfaces at $x = -d$ and $x = 0$ surfaces separating media with refractive indices n_1, n_2, and n_3, respectively.

- Assume that medium 1 has TE-polarized incident and reflecting electric fields superposing as

$$\hat{y}E_o(e^{-jk_1(\sin\theta_1 z+\cos\theta_1 x)} + Re^{-jk_1(\sin\theta_1 z-\cos\theta_1 x)});$$

  the field in medium 2 is

$$\hat{y}E_o(Pe^{-jk_2(\sin\theta_2 z+\cos\theta_2 x)} + Qe^{-jk_2(\sin\theta_2 z-\cos\theta_2 x)});$$

  and in medium 3 we have

$$\hat{y}E_oTe^{-jk_3(\sin\theta_3 z+\cos\theta_3 x)}.$$

- In medium 1 the $\hat{z}$ component of total $\tilde{\mathbf{H}}$ is

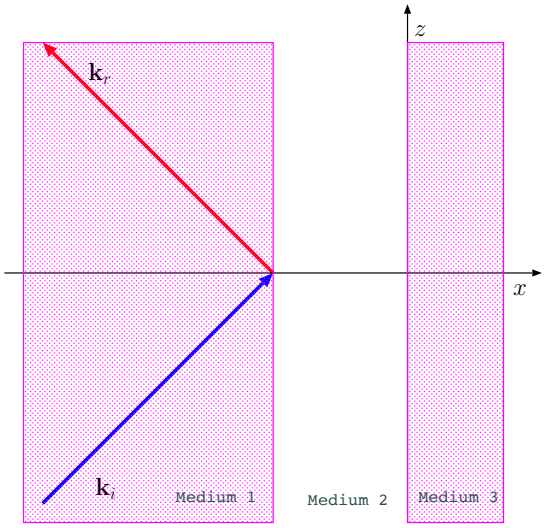
$$\frac{E_o}{\eta_1}\cos\theta_1(e^{-jk_1(\sin\theta_1 z+\cos\theta_1 x)} - Re^{-jk_1(\sin\theta_1 z-\cos\theta_1 x)});$$

  in medium 2 we have

$$\frac{E_o}{\eta_2}\cos\theta_2(Pe^{-jk_2(\sin\theta_2 z+\cos\theta_2 x)} - Qe^{-jk_2(\sin\theta_2 z-\cos\theta_2 x)});$$

and in medium 3

$$\frac{E_o}{\eta_3}\cos\theta_3 T e^{-jk_3(\sin\theta_3 z+\cos\theta_3 x)}.$$



- BC's applied at $x=-d$ and $x=0$ require a "phase matching", that is

$$k_1\sin\theta_1 = k_2\sin\theta_2 = k_3\sin\theta_3,$$

leading to Snell's law relations

$$n_2\sin\theta_2 = n_1\sin\theta_1 \quad \text{and} \quad n_3\sin\theta_3 = n_1\sin\theta_1.$$

- Matching the tangential $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{H}}$ at $x=0$ boundary yields

$$P+Q=T \ \text{ with } \ \frac{P-Q}{\eta_2/\cos\theta_2} = \frac{T}{\eta_3/\cos\theta_3}$$

implies

$$\frac{P-Q}{\eta_2/\cos\theta_2} = \frac{P+Q}{\eta_3/\cos\theta_3} \ \Rightarrow \ \frac{Q}{P} = \frac{\eta_3/\cos\theta_3 - \eta_2/\cos\theta_2}{\eta_3/\cos\theta_3 + \eta_2/\cos\theta_3} \equiv \Gamma_{32}$$

defining an effective "load reflection" coefficient for this problem.

- Transverse field matching at $x=-d$ requires for $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{H}}$

$$e^{jk_1\cos\theta_1 d} + Re^{-jk_1\cos\theta_1 d} = Pe^{jk_2\cos\theta_2 d} + Qe^{-jk_2\cos\theta_2 d}$$

and

$$\frac{1}{\eta_1}\cos\theta_1(e^{jk_1\cos\theta_1 d} - Re^{-jk_1\cos\theta_1 d}) = \frac{1}{\eta_2}\cos\theta_2(Pe^{jk_2\cos\theta_2 d} - Qe^{-jk_2\cos\theta_2 d}),$$

respectively. Now define transverse field impedance

$$Z(-d) = \frac{E_y}{H_z} = \frac{Pe^{jk_2\cos\theta_2 d} + Qe^{-jk_2\cos\theta_2 d}}{\frac{Pe^{jk_2\cos\theta_2 d} - Qe^{-jk_2\cos\theta_2 d}}{\eta_2/\cos\theta_2}} = \frac{\eta_2}{\cos\theta_2}\frac{1+\Gamma_{32}e^{-j2k_2\cos\theta_2 d}}{1-\Gamma_{32}e^{-j2k_2\cos\theta_2 d}}$$

and

$$Z(-d) = \frac{E_y}{H_z} = \frac{e^{jk_1\cos\theta_1 d} + Re^{-jk_1\cos\theta_1 d}}{\frac{e^{jk_1\cos\theta_1 d} - Re^{-jk_1\cos\theta_1 d}}{\eta_1/\cos\theta_1}} = \frac{\eta_1}{\cos\theta_1}\frac{1+\Gamma_{21}}{1-\Gamma_{21}}$$

where

$$\Gamma_{21} \equiv \frac{Re^{-jk_1\cos\theta_1 d}}{e^{jk_1\cos\theta_1 d}}.$$

- Clearly $|\Gamma_{21}|^2$ is the *reflectance* (fraction of incident power density in the reflected wave) and $1 - |\Gamma_{21}|^2$ is the *transmittance,* wherein

$$\Gamma_{21} = \frac{Z(-d) - \eta_1/\cos\theta_1}{Z(-d) - \eta_1/\cos\theta_1}.$$

These results suggest the use of transmission line analogy in terms of characteristic impedances $\eta_i/\cos\theta_i$ and $k_{xi} \equiv k_i\cos\theta_i$ in phase terms. Note that in evanescent regions $\cos\theta_i$ are purely imaginary. Also quarter-wave transformations can be used when $d\cos\theta_2 = \lambda_2/4$ and half-wave transformations when $d\cos\theta_2 = \lambda_2/2$.

- For the TM case, the use of $\eta_i \cos\theta_i$ in place of $\eta_i / \cos\theta_i$ in reflection coefficient and impedance calculations leads to the correct reflectance and transmittance values (same trick would also work in TE and TM Fresnel reflection coefficients in oblique reflections from a single interface as well as in guide impedance formulae where $\cos\theta$ is replaced by $\sqrt{1 - f_c^2/f^2}$).

  Here are the details:

  - Matching the tangential $\tilde{\mathbf{H}}$ and $\tilde{\mathbf{E}}$ at $x = 0$ boundary yields

    $$P{+}Q = T \;\text{ with }\; -\eta_2 \cos\theta_2 (P{-}Q) = -\eta_3 \cos\theta_3 T = -\eta_3 \cos\theta_3 (P{+}Q)$$

    implying

    $$\eta_2 \cos\theta_2 (P{-}Q) = \eta_3 \cos\theta_3 (P{+}Q) \;\Rightarrow\; \frac{Q}{P} = \frac{\eta_2 \cos\theta_2 - \eta_3 \cos\theta_3}{\eta_2 \cos\theta_2 + \eta_3 \cos\theta_3} \equiv -\Gamma_{32}$$

    defining an effective "load reflection" coefficient for this problem.
  - Transverse field matching at $x = -d$ requires, for $\tilde{\mathbf{H}}$ and $\tilde{\mathbf{E}}$,

    $$e^{jk_1 \cos\theta_1 d} + Re^{-jk_1 \cos\theta_1 d} = Pe^{jk_2 \cos\theta_2 d} + Qe^{-jk_2 \cos\theta_2 d}$$

    and

    $$-\eta_1 \cos\theta_1 (e^{jk_1 \cos\theta_1 d} - Re^{-jk_1 \cos\theta_1 d}) = -\eta_2 \cos\theta_2 (Pe^{jk_2 \cos\theta_2 d} - Qe^{-jk_2 \cos\theta_2 d}),$$

    respectively. Now define transverse field impedance

    $$Z(-d) = \frac{-E_z}{H_y} = \frac{\eta_2 \cos\theta_2 (Pe^{jk_2 \cos\theta_2 d} - Qe^{-jk_2 \cos\theta_2 d})}{Pe^{jk_2 \cos\theta_2 d} + Qe^{-jk_2 \cos\theta_2 d}} = \eta_2 \cos\theta_2 \frac{1 + \Gamma_{32} e^{-j2k_2 \cos\theta_2 d}}{1 - \Gamma_{32} e^{-j2k_2 \cos\theta_2 d}}$$

and

$$Z(-d) = \frac{-E_z}{H_y} = \frac{\eta_1 \cos\theta_1 (e^{jk_1 \cos\theta_1 d} - Re^{-jk_1 \cos\theta_1 d})}{e^{jk_1 \cos\theta_1 d} + Re^{-jk_1 \cos\theta_1 d}} = \eta_1 \cos\theta_1 \frac{1+\Gamma_{21}}{1-\Gamma_{21}}$$

where

$$\Gamma_{21} \equiv -\frac{Re^{-jk_1 \cos\theta_1 d}}{e^{jk_1 \cos\theta_1 d}}.$$

– The upshot is,

$$\Gamma_{21} = \frac{Z(-d) - \eta_1 \cos\theta_1}{Z(-d) + \eta_1 \cos\theta_1}$$

wherein we see the replacement of all $\eta_i / \cos\theta_i$ in TE-mode relations by $\eta_i \cos\theta_i$ to become the corresponding TM-mode relations.

27 Parallel-plate waveguides — TE_m modes

- Consider a TE polarized incident field

$$\tilde{\mathbf{E}}_i = \hat{y} E_o e^{-j(-k_x x + k_z z)}$$

reflecting from a conducting plate on $x = 0$ plane as depicted in the margin so that a reflected wave

$$\tilde{\mathbf{E}}_r = -\hat{y} E_o e^{-j(k_x x + k_z z)}$$

is produced to ascertain $\hat{x} \times (\tilde{\mathbf{E}}_i + \tilde{\mathbf{E}}_r) = 0$ on $x = 0$ plane. In these expressions

$$k_x^2 + k_z^2 = k^2 = \omega^2 \mu_o \epsilon_o$$

assuming that the plate is embedded in free space (otherwise use μ and ϵ, instead).

- The incident and reflected waves will then produce a **total field**

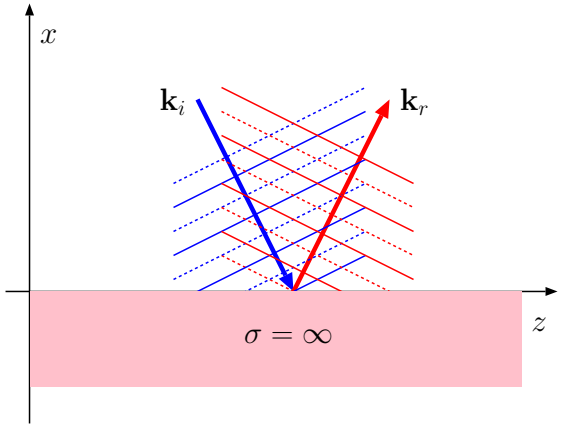
$$\tilde{\mathbf{E}} = \tilde{\mathbf{E}}_i + \tilde{\mathbf{E}}_r = \hat{y} E_o e^{-jk_z z}(e^{jk_x x} - e^{-jk_x x}) = 2j\hat{y} E_o e^{-jk_z z} \sin(k_x x)$$

in $x > 0$ region which propagates in the z-direction with a phase velocity

$$v_{pz} = \frac{\omega}{k_z}$$

and "stands" in the x-direction.

With this lecture we start our study of guided waves and resonant cavities.
In ECE 329 you were already exposed to guided TEM wave propagation in two-wire transmission line systems.
Here we will study TE and TM mode propagation on parallel-plate transmission-lines and hollow waveguides.
While guided TEM modes are dispersion-free and propagate at all frequencies, TE and TM modes are dispersive and exhibit frequency-dependent cutoff.



Intersections of solid and dashed wavefronts of the incident and reflected waves demark the locations of the nulls of the total y-directed electric field.

- The standing wave in the x-direction is characterized by a tangential electric field component

$$\tilde{E}_y(x,z) \propto \sin(k_x x)$$

which has nulls at all $x > 0$ satisfying

$$k_x x = m\pi \quad \Rightarrow \quad x = \frac{m\pi}{k_x} \text{ for integers } m = 1, 2, 3, \cdots.$$

- Likewise, the tangential component at $x = a$,

$$\tilde{E}_y(a,z) \propto \sin(k_x a),$$

will vanish for all k_x satisfying

$$k_x a = m\pi \quad \Rightarrow \quad k_x = \frac{m\pi}{a} \text{ for integers } m = 1, 2, 3, \cdots$$

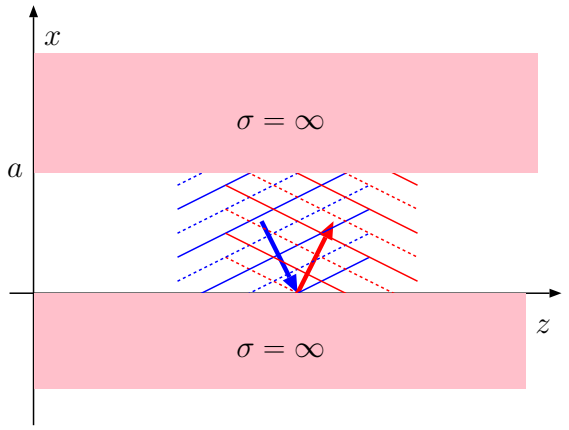
- Hence

$$\tilde{\mathbf{E}} = 2j\hat{y}E_o e^{-jk_z z}\sin(k_x x)$$

can be regarded as a steady-state solution of Maxwell's equations in between PEC plates located at $x = 0$ and $x = a$ so long as satisfies a "guidance condition"

$$k_x = \frac{m\pi}{a} \text{ for integers } m = 1, 2, 3, \cdots$$

This is true because



The y-directed total electric field is zero at x=0 and x=a surfaces of the guide formed by the conducting plates and exhibit maximum magnitude at the intersections of solid or dashed wavefront pairs.

Note the 180 degree reversals in the reflected phase fronts on both surfaces (top and bottom)as required by a reflection coefficient of -1.

– $\tilde{\mathbf{E}} \propto \hat{y}\sin(k_x x)$ satisfies Mazwell's boundary conditions of zero tangential electric field at the PEC surfaces $x = 0$ and $x = a$,

– $\tilde{\mathbf{E}} \propto \hat{y}\sin(k_x x)$ satisfies the Maxwell's equations in the channel $0 < x < a$

- These are designated as "TE_m mode" solutions and $m = 0$ case is excluded because $k_x = m\pi/a = 0$ for $m = 0$ leading to $\tilde{\mathbf{E}} \propto \hat{y}\sin(k_x x) = 0$.

- Guided TE_m mode waves with electric field phasor

$$\tilde{\mathbf{E}} = 2j\hat{y}E_o e^{-jk_z z}\sin(k_x x)$$

in region $0 < x < a$ with "quantized"

$$k_x = \frac{m\pi}{a}$$

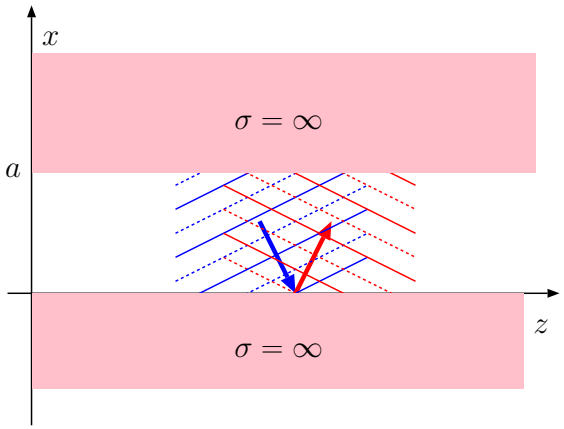
will have

$$k_z = \sqrt{k^2 - k_x^2} = k\sqrt{1 - \frac{k_x^2}{k^2}} = \frac{\omega}{c}\sqrt{1 - \frac{k_x^2 c^2}{\omega^2}} = \frac{\omega}{c}\sqrt{1 - \frac{\omega_c^2}{\omega^2}},$$

where

$$\omega_c \equiv k_x c = \frac{m\pi c}{a}$$

is known as cutoff frequency of TE_m mode.



The y-directed total electric field is zero at x=0 and x=a surfaces of the guide formed by the conducting plates and exhibit maximum magnitude at the intersections of solid or dashed wavefront pairs.

Note the 180 degree reversals in the reflected phase fronts on both surfaces (top and bottom)as required by a reflection coefficient of -1.

– The corresponding field in the time domain is obtained by multiplying the phasor with $e^{j\omega t}$ and taking the real part, yielding

$$\mathbf{E}(x,z,t) = -2\hat{y}E_o \sin(\omega t - k_z z)\sin(k_x x).$$

Since propagation of the TE_m mode field is controlled by k_z, the relationship

$$k_z = \frac{\omega}{c}\sqrt{1-\frac{\omega_c^2}{\omega^2}} = \frac{\omega}{c}\sqrt{1-\frac{f_c^2}{f^2}},$$

where

$$\omega_c = k_x c = \frac{m\pi c}{a} \text{ and } f_c \equiv \frac{\omega_c}{2\pi} = \frac{mc}{2a},$$

is the **dispersion relation** of the TE_m mode, from which it follows that:

Example:
for $a = 3$ cm, $m = 1$ implies

$$\begin{aligned} f_c &= \frac{mc}{2a} = \frac{3\times 10^8}{2\times 0.03} \\ &= 5\times 10^9 \,\mathrm{Hz} \\ &= 5\,\mathrm{GHz} \end{aligned}$$

for TE_1mode.
The cutoff frequency for TE_2mode is 10 GHz, for TE_3mode is 15 GHz, and so on.
A signal with 11 GHz frequency will propagate in TE_1 and TE_2mode, but will be evanescent in $\mathrm{TE}_3$and higher order modes.

1. **Propagation** takes place if $f > f_c = \frac{mc}{2a}$, and **evanescence** otherwise, and
2. Phase and group velocities

$$v_{pz} \equiv \frac{\omega}{k_z} = \frac{c}{\sqrt{1-\frac{f_c^2}{f^2}}} \text{ and } v_g \equiv \frac{\partial\omega}{\partial k_z} = c\sqrt{1-\frac{f_c^2}{f^2}}$$

in analogy with plasma dispersion (identical except for the interchange of ω_c with ω_p).

- The component TEM waves that satisfy the condition

$$\lambda f = c \quad \Leftrightarrow \quad \frac{\omega}{k} = c$$

and superpose to form the TE_m mode have a wavelength, for

$$f = f_c = \frac{mc}{2a},$$

given by

$$\lambda = \frac{c}{f} = \frac{c}{f_c} = \frac{c}{\frac{mc}{2a}} = \frac{2a}{m} \equiv \lambda_c.$$

Consequently, the TE_m mode dispersion relation can also be cast as

$$k_z = \frac{\omega}{c}\sqrt{1 - \frac{f_c^2}{f^2}} = \frac{\omega}{c}\sqrt{1 - \frac{c^2}{\lambda_c^2 f^2}} = \frac{\omega}{c}\sqrt{1 - \frac{\lambda^2}{\lambda_c^2}};$$

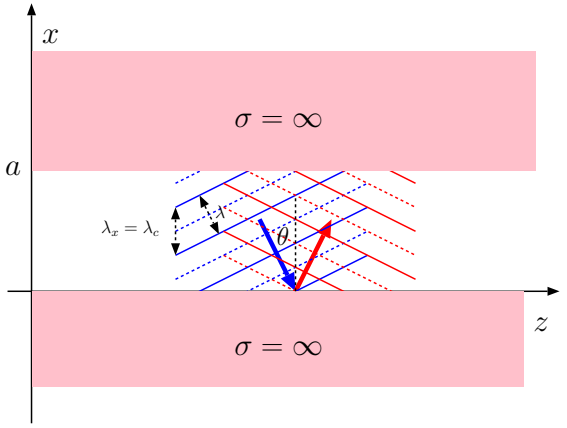
note that:

1. The "cutoff wavelength"

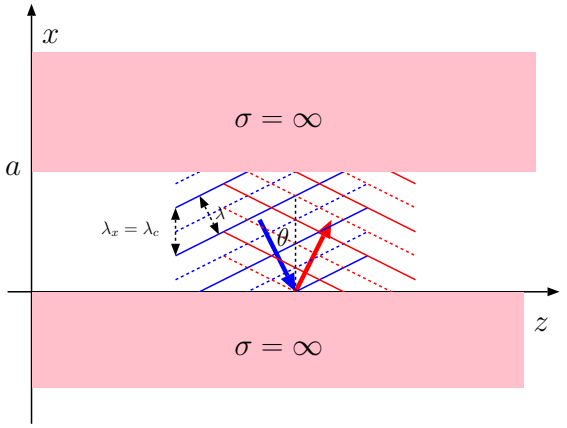
$$\lambda_c = \frac{2a}{m}$$

can be remembered to be "twice the guide width a divided by the mode number m",

2. The factor $\frac{\lambda}{\lambda_c}$ can be used to replace the factor $\frac{f_c}{f}$ that appears in all of the expressions given above (and below).



Note that cutoff-wavelength=2a/m is also the "trace wavelength" in x-direction.



Note that cutoff-wavelength=2a/m is also the "trace wavelength" in x-direction.

3. Finally, since $k_x = \frac{m\pi}{a}$, we have

$$\lambda_c = \frac{2a}{m} = \frac{2\pi}{k_x} = \lambda_x$$

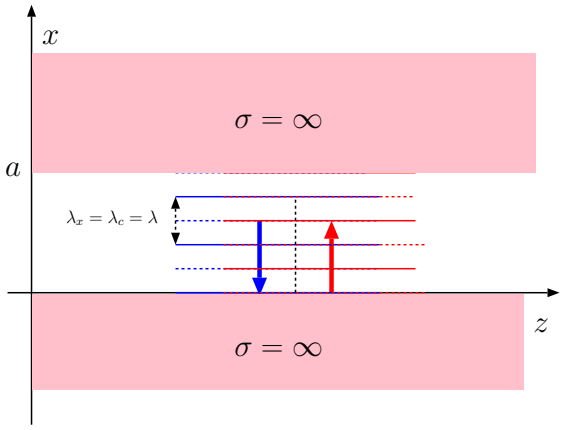
at any frequency f.

- Each permissible k_x (or mode) at a given operation frequency ω is associated with an incidence *and* reflection angle θ (see margin) of the component TEM waves superposing, which is given by

$$\cos\theta = \frac{k_x}{k} = \frac{\lambda}{\lambda_x} = \frac{\lambda}{\lambda_c} = \frac{f_c}{f} \quad \Rightarrow \theta = \cos^{-1}\frac{\lambda}{\lambda_c} = \cos^{-1}\frac{f_c}{f},$$

indicating that as $f \to f_c$, $\theta \to 0$, that is at $f = f_c$ (at cutoff), the field consists of plane waves bouncing back and forth between the plates at $x = 0$ and $x = a$ at normal incidence (see margin).

- Component waves superposing to constitute the guided field can be viewed as reflections of one another from the guide plates, produced (self-consistently) by surface currents induced on the conducting walls of the guide at $x = 0$ and $x = a$.



At cutoff (f=f_c) we have k_z=0 and thus TEM waves bouncing between the plates in exclusively x direction, carrying no energy in z-direction.

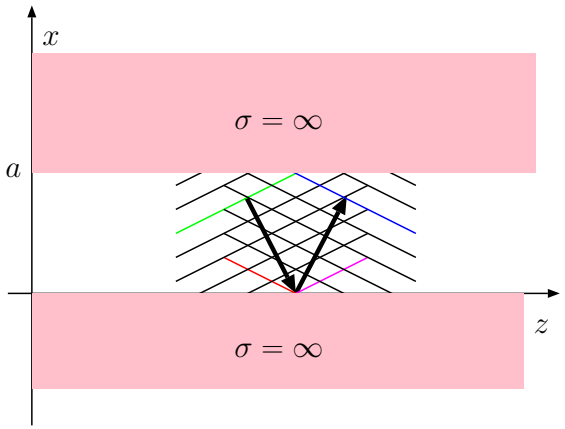
Guide wavelength: The component TEM waves that superpose to form the TE_m mode solutions have a wavelength

$$\lambda = \frac{2\pi}{k}$$

as usual. We define

$$\lambda_z \equiv \frac{2\pi}{k_z} = \frac{2\pi}{k\sqrt{1-\frac{f_c^2}{f^2}}} = \frac{\lambda}{\sqrt{1-\frac{f_c^2}{f^2}}}$$

to be the *guide wavelength* λ_g. Note the distinction between



Below, we re-derive the "guidance condition"

$$k_x = \frac{m\pi}{a}$$

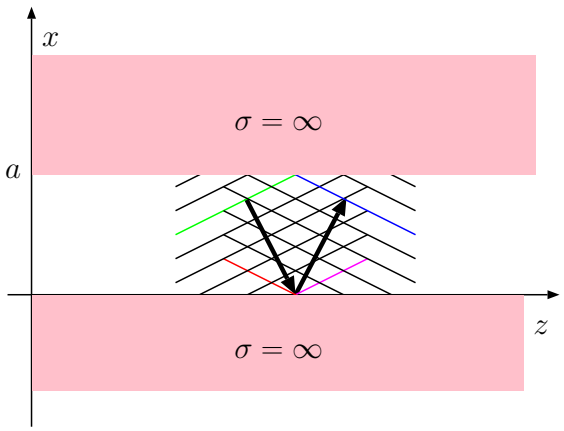
for TE_m modes — with reference to the diagram shown above — by requiring "a phase consistency" between the reflected wave-pairs "responsible for one another":

- Note the four of the phase fronts in the diagram above (repeated in the margin) belonging, in pairs, to the "upgoing" and "downgoing" component TEM waves of y-polarized electric field have been highlighted by the colors red and blue, and green and magenta, respectively. Indicating the phases on these singled out phase fronts as ϕ_r, ϕ_b, ϕ_g, and ϕ_m, we note in succession that

  –

  $$\phi_b = \phi_r - k_x a$$

  because of phase delay $k_x a$ in the upgoing wave over an excursion by a along the x-axis,



–

$$\phi_g = \phi_b + \angle\Gamma,$$

where Γ=-1 is the reflection coefficient on the upper plate converting the upgoing electric field wave into a downgoing wave (to which ϕ_g belongs),

–

$$\phi_m = \phi_g - k_x a$$

because of phase delay $k_x a$ in the downgoing wave over an excursion by a along the x-axis,

–

$$\phi_r = \phi_m + \angle\Gamma,$$

where Γ=-1 is the reflection coefficient on the lower plate converting the downgoing wave into an upgoing wave (to which ϕ_r belongs).

– Now back-substituting in the expression for ϕ_r the expressions for ϕ_m, ϕ_g, ϕ_b in succession, we find that

$$\phi_r = \phi_r - 2(k_x a - \angle\Gamma)$$

Requiring $\angle\Gamma$ to lie in the range $-\pi < \angle\Gamma \leq \pi$, and given the inherent 2π ambiguities permitted for a wave phase (because of

$$\begin{aligned} \phi_b &= \phi_r - k_x a, \\ \phi_g &= \phi_b + \angle\Gamma, \\ \phi_m &= \phi_g - k_x a, \\ \phi_r &= \phi_m + \angle\Gamma. \end{aligned}$$

the non-discriminatory response of co-sinusoids to phases differing by integer multiples of 2π), this condition implies that we can take

$$k_x a - \angle\Gamma = n\pi, \;\; n = 0, 1, 2, \cdots$$

since $k_x a$ is *by definition* non-negative. With

$$\Gamma = -1 \;\; \Rightarrow \;\; \angle\Gamma = \pi,$$

the condition reduces to

$$k_x a = (n+1)\pi = m\pi, \;\; m = 1, 2, 3, \cdots$$

for TE_m modes, which is the same condition we obtained by using uniqueness arguments in connection with the standing-wave field solution earlier on.

We will utilize the relation

$$k_x a - \angle\Gamma = n\pi, \;\; n = 0, 1, 2, \cdots,$$

with the constraint that $-\pi < \angle\Gamma \leq \pi$ later on when we will study dielectric-slab waveguides — in that case we will have a TIR ($\theta_1 > \theta_c$) related reflection coefficient Γ with a θ_1 dependent phase angle $\angle\Gamma$.

- There is an even easier way of obtaining the same guidance condition using the following steps: The guided TE_m modes propagating in z direction are superpositions of incident and reflected TEM plane waves

⇐ Pay close attention to this method.

Next lecture we will use it to

$$\tilde{\mathbf{E}}_i = \hat{y} E_o e^{-j(-k_x x + k_z z)}$$

and

$$\tilde{\mathbf{E}}_r = \hat{y} E_o \Gamma e^{-j(k_x x + k_z z)}$$

where $\Gamma = \Gamma_\perp = -1$ is the TE-mode reflection coefficient appropriate for an air-PEC interface. Since for the permissible guided modes, TEM wave $\tilde{\mathbf{E}}_r$ gets reflected (once more) *at* $x = a$ to become $\tilde{\mathbf{E}}_i$ at the same location, it is necessary that

$$(\hat{y} E_o \Gamma e^{-j(k_x a + k_z z)})\Gamma = \hat{y} E_o e^{-j(-k_x a + k_z z)} e^{-j2\pi n}$$

for any integer n, i.e.,

$$|\Gamma|^2 e^{j2\angle\Gamma} e^{-jk_x a} = e^{jk_x a} e^{-j2\pi n}.$$

This is possible if and only if $|\Gamma| = 1$ and

$$2\angle\Gamma - k_x a = k_x a - 2\pi n \;\; \Rightarrow \;\; k_x a = \angle\Gamma + n\pi = m\pi,$$

the same guidance condition that we obtained earlier on.

Example 1: TE_m mode fields have transverse electric field phasors

$$\tilde{\mathbf{E}} = 2j\hat{y} E_o e^{-jk_z z} \sin(k_x x)$$

satisfying the zero-tangential field conditions at $x = 0$ and $x = a$ planes with

$$k_x = \frac{m\pi}{a} \;\text{ and }\; k_z = \frac{\omega}{c}\sqrt{1 - \frac{f_c^2}{f^2}}$$

where

$$f_c = \frac{mc}{2a}.$$

(a) Determine the magnetic field intensity phasor $\tilde{\mathbf{H}}$ for TE_m mode waves. (b) Also determine $\eta_{TE} \equiv \frac{E_y}{-H_x}$, which is the effective guide impedance for TE_m mode.

Solution: (a) Using Faraday's law, we have

$$\begin{aligned}\tilde{\mathbf{H}} &= \frac{\nabla \times \tilde{\mathbf{E}}}{-j\omega\mu_o} = \frac{\begin{vmatrix} \hat{x} & \hat{y} & \hat{z} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ 0 & E_y & 0 \end{vmatrix}}{-j\omega\mu_o} = \frac{-\hat{x}\frac{\partial E_y}{\partial z} + \hat{z}\frac{\partial E_y}{\partial x}}{-j\omega\mu_o} \\ &= -\frac{2E_o}{\omega\mu_o}(\hat{x}(jk_z \sin(k_x x) + \hat{z}k_x \cos(k_x x))e^{-jk_z z}.\end{aligned}$$

(b) Using the result of part (a), we have

$$\begin{aligned}\eta_{TE} &= \frac{E_y}{-H_x} = \frac{2jE_o e^{-jk_z z}\sin(k_x x)}{\frac{2E_o}{\omega\mu_o}jk_z \sin(k_x x)e^{-jk_z z}} \\ &= \frac{1}{\frac{1}{\omega\mu_o}k_z} = \frac{\omega\mu_o}{\frac{\omega}{c}\sqrt{1-\frac{f_c^2}{f^2}}} = \frac{\eta_o}{\sqrt{1-\frac{f_c^2}{f^2}}}.\end{aligned}$$

28 Parallel-plate waveguides — TM_m modes

- Last lecture we discussed the TE_m modes of propagation in parallel-plate waveguides.
- These guided modes have y-polarized electric fields transverse to the propagation direction z and exhibit a standing wave pattern in x-direction with m half-wavelengths of variation between the guide plates at $x = 0$ and $x = a$.
  - More specifically, the TE_m modes have transverse electric field phasors

$$\tilde{\mathbf{E}} = 2j\hat{y}E_o e^{-jk_z z}\sin(k_x x)$$

where

$$k_x = \frac{m\pi}{a}, \quad m = 1, 2, \cdots$$

and

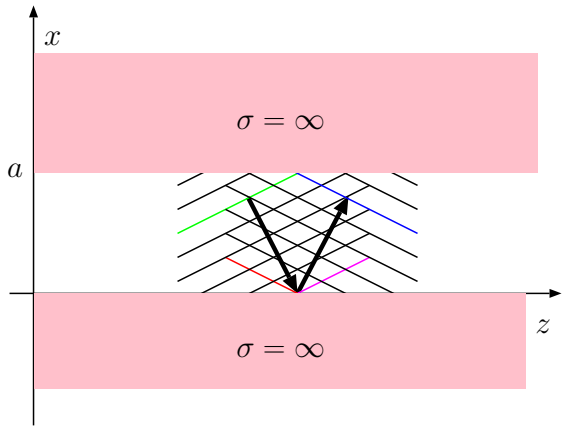
$$k_z = \sqrt{k^2 - k_x^2} = \frac{\omega}{c}\sqrt{1 - \frac{f_c^2}{f^2}}$$

with **cutoff frequencies**

$$f_c = \frac{mc}{2a}.$$

Alternatively (and equivalently),

$$k_z = \sqrt{k^2 - k_x^2} = \frac{\omega}{c}\sqrt{1 - \frac{\lambda^2}{\lambda_c^2}},$$



Superposing all TE_mmode fields with amplitudes E_m (to stand for $2jE_o$), $m \in [1, \infty]$, we can write

$$\tilde{\mathbf{E}}(x, y, 0) = \hat{y}\sum_{m=1}^{\infty} E_m \sin(\frac{m\pi}{a}x)$$

which is in the form of a Fourier series, implying that any $\tilde{\mathbf{E}}(x, y, 0)$ periodic in x with a period $2a$ can be expressed in this format as a superposition of TE_mmode fields.

Such fields can be generated within a waveguide using a small dipole or loop antenna — more on this later!

with **cutoff wavelengths**

$$\lambda_c = \frac{2a}{m}.$$

Above, the operation frequency f and operation wavelength λ satisfy $\lambda f = c$, and furthermore

$$k = \frac{2\pi}{\lambda} = \frac{\omega}{c}$$

is the operation wavenumber. The propagation characteristics of the guided mode, on the other hand, depends on k_z, with

$$v_p = \frac{\omega}{k_z} \text{ and } \lambda_g = \frac{2\pi}{k_z}$$

denoting the phase velocity and the wavelength of the guided mode when

$$f > f_c \text{ and, equivalently, } \lambda < \lambda_c,$$

corresponding to propagation condition for a given mode. When

$$f < f_c \text{ and, equivalently, } \lambda > \lambda_c,$$

the mode is evanescent.

– Since

$$k_z = \sqrt{k^2 - k_x^2} = \frac{\omega}{c}\sqrt{1 - \frac{f_c^2}{f^2}}$$

is effectively the dispersion relation of the guided modes having the same form as the plasma dispersion relation, it follows that the group velocity is

$$v_g = \frac{\partial \omega}{\partial k_z} = c\sqrt{1 - \frac{f_c^2}{f^2}} \quad \text{and} \quad v_g v_p = c^2$$

just like in plasmas.

– Finally TE_m mode fields have a guide impedance

$$\eta_{TE} = -\frac{E_y}{H_x} = \frac{\eta_o}{\sqrt{1 - \frac{f_c^2}{f^2}}}$$

relating the transverse field components of the wave.

- Next we turn our attention on TM_m mode fields which share most of the dispersion characteristics of the TE_m mode fields. However, they are essentially orthogonal to TE_m mode fields and furthermore support the $m = 0$ case which is absent for TE_m modes.

- TM_m mode guided waves propagating in z direction correspond to superpositions of incident and reflected TEM plane waves with

$$\tilde{\mathbf{H}}_i = \hat{y} H_o e^{-j(-k_x x + k_z z)}$$

and

$$\tilde{\mathbf{H}}_r = \hat{y} H_o \Gamma e^{-j(k_x x + k_z z)}$$

where $\Gamma = R = 1$ is the TM-mode reflection coefficient at an air-PEC interface.

For permissible TM_m modes $\tilde{\mathbf{H}}_r$ gets reflected (once more) *at* $x = a$ to become $\tilde{\mathbf{H}}_i$ at the same location, and thus it is necessary that

$$\underbrace{\hat{y} H_o \Gamma e^{-j(k_x a + k_z z)}}_{\tilde{\mathbf{H}}_r(x=a,z)} \Gamma = \underbrace{\hat{y} H_o e^{-j(-k_x a + k_z z)}}_{\tilde{\mathbf{H}}_i(x=a,z)} \overbrace{e^{-j2\pi m}}^{1}$$

for any integer m, i.e.,

$$|\Gamma|^2 e^{j2\angle\Gamma} e^{-jk_x a} = e^{jk_x a} e^{-j2\pi m}.$$

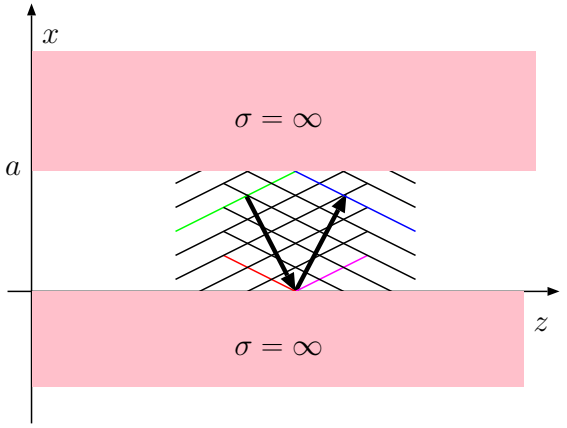
This is possible, since $|\Gamma| = |R| = 1$ and $\angle\Gamma = \angle R = 0$, with

$$-k_x a = k_x a - 2\pi m \quad \Rightarrow \quad k_x a = m\pi,$$

leading to

$$k_x = \frac{m\pi}{a}, \quad m = 0, 1, 2, \cdots$$

as the guiding condition for TM_m modes.



The same self-consistency condition of guided waves of any polarization by any type of channel of a width a can be posed as

$$|\Gamma|^2 e^{-j2k_x a} = e^{-j2(m\pi - \angle\Gamma)},$$

satisfied *iff*

$|\Gamma| = 1$ (TIR or reflection from PEC)

and

$$k_x = \frac{m\pi - \angle\Gamma}{a},$$

known as the *guidance condition*.

Permissible modes in dielectric waveguides can be identified using this guidance condition once $\angle\Gamma$ variation with k_x/k is specified.

- Since for TM_m modes the transverse field

$$\tilde{\mathbf{H}} = \tilde{\mathbf{H}}_i + \tilde{\mathbf{H}}_r = 2\hat{y}H_o e^{-jk_z z}\cos(k_x x)$$

does not vanish with vanishing k_x, the $m = 0$ mode is permitted. In fact, TM_0 mode corresponding to $m = 0$ is the TEM mode studied in EEC 329 in transmission line (TL) theory.

  - TM_0=TEM consists of wave fields

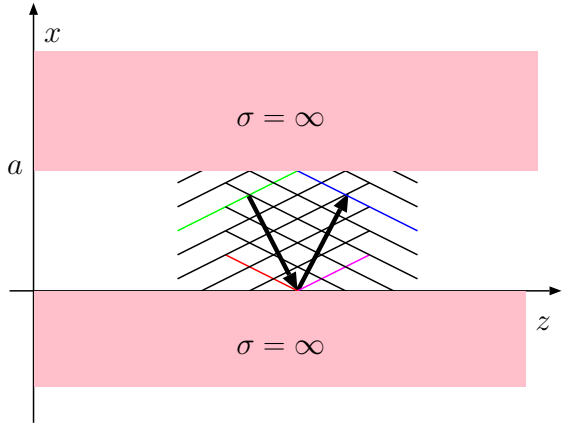
$$\tilde{\mathbf{E}} = \hat{x}E_o e^{-jkz} \quad \text{and} \quad \tilde{\mathbf{H}} = \hat{y}\frac{E_o}{\eta_o}e^{-jkz}$$

    which naturally satisfy the boundary conditions at $x = 0$ and $x = a$ planes of having zero tangential electric field.

  - Also, for this mode

$$k_x = 0 \text{ and } k_z = k,$$

    which follows when $m = 0$ is permitted in dispersion equations when applied for the case of TM_m modes.



Superposing all TM_mmode fields with amplitudes H_m (to stand for $2H_o$), $m \in [0, \infty]$, we can write

$$\tilde{\mathbf{H}}(x, y, 0) = \hat{y}\sum_{m=0}^{\infty} H_m \cos(\frac{m\pi}{a}x)$$

which is in the form of a Fourier series, implying that any $\tilde{\mathbf{H}}(x, y, 0)$ periodic in x with a period a can be expressed in this format as a superposition of TM_mmode fields.

Such fields can be generated within a waveguide using a small dipole or loop antenna — more on this later!

Example 1: TM_m mode fields have transverse magnetic intensity phasors

$$\tilde{\mathbf{H}} = 2\hat{y}H_o e^{-jk_z z}\cos(k_x x).$$

(a) Determine the electric field phasor $\tilde{\mathbf{E}}$ for TM_m mode waves. (b) Also determine $\eta_{TM} \equiv \frac{E_x}{H_y}$, the effective guide impedance for TM_m mode.

Solution: (a) Using Ampere's law, we have

$$\begin{aligned}\tilde{\mathbf{E}} &= \frac{\nabla \times \tilde{\mathbf{H}}}{j\omega\epsilon_o} = \frac{\begin{vmatrix} \hat{x} & \hat{y} & \hat{z} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ 0 & H_y & 0 \end{vmatrix}}{j\omega\epsilon_o} = \frac{-\hat{x}\frac{\partial H_y}{\partial z} + \hat{z}\frac{\partial H_y}{\partial x}}{j\omega\epsilon_o} \\ &= \frac{2H_o}{j\omega\epsilon_o}(\hat{x}(jk_z\cos(k_x x) - \hat{z}k_x\sin(k_x x))e^{-jk_z z}.\end{aligned}$$

(b) Using the result of part (a), we have

$$\begin{aligned}\eta_{TM} &= \frac{E_x}{H_y} = \frac{\frac{2H_o}{j\omega\epsilon_o}jk_z\cos(k_x x)e^{-jk_z z}}{2H_o e^{-jk_z z}\cos(k_x x)} \\ &= \frac{k_z}{\omega\epsilon_o} = \frac{\frac{\omega}{c}\sqrt{1-\frac{f_c^2}{f^2}}}{\omega\epsilon_o} = \eta_o\sqrt{1-\frac{f_c^2}{f^2}}.\end{aligned}$$

- Note that the results obtained in Example 1 give non-trivial results for $m = 0$ case with $k_x = 0$ and $k_z = k$.
- TM_0=TEM mode has no cutoff frequency and it is non-dispersive. It

has all the properties of the unguided TEM waves we are familiar with.

- Finally, regarding dispersive TE_m and TM_m modes with $m \geq 1$, all the equations derived above can also be used when the guiding plates are embedded in dielectric media (instead of air) by simply replacing

$$c = \frac{1}{\sqrt{\mu_o \epsilon_o}} \text{ with } v_p = \frac{1}{\sqrt{\mu\epsilon}}$$

in the dispersion equations.

- There is a straightforward geometrical interpretation of v_g obtained for guided TE and TM modes.

  - Clearly the component TEM waves which constitute the guided modes (TE and TM) propagate at angles $\pm\theta$ with a velocity c in air-filled waveguides. The projection along z of the velocity vectors pointing in $\pm\theta$ directions are

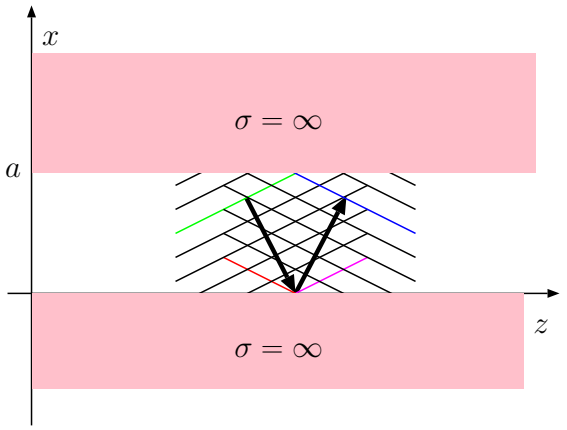
$$c\sin\theta = c\sqrt{1-\cos^2\theta} = c\sqrt{1-\frac{k_x^2}{k^2}} = c\sqrt{1-\frac{f_c^2}{f^2}},$$

    which is of course the group velocity of the guided modes as we have seen before.

This makes sense: in the component TEM waves — which are non-dispersive — of the guided modes, the phase fronts as well as any imposed modulations move with the same velocity, namely c.

– While the progress of modulation on the component waves along $\pm\theta$ occurs at a velocity c, the modulation covers a shorter distance along z than the corresponding slant distance along $\pm\theta$, and thus v_g measuring the progress of the modulation along z is smaller than c measuring the same progress along $\pm\theta$.

29 Parallel-plate waveguides: example problems



Summarizing the properties of guided modes of propagation in parallel-plate waveguides:

- TE_m and TE_m modes with the transverse field phasors

$$\tilde{\mathbf{E}} = 2j\hat{y}E_o e^{-jk_z z}\sin(k_x x) \text{ and } \tilde{\mathbf{H}} = 2\hat{y}H_o e^{-jk_z z}\cos(k_x x),$$

  respectively, where

$$k_x = \frac{m\pi}{a} \text{ and } k_z = \sqrt{k^2 - k_x^2} = \frac{\omega}{c}\sqrt{1 - \frac{f_c^2}{f^2}} = \frac{\omega}{c}\sqrt{1 - \frac{\lambda^2}{\lambda_c^2}},$$

  with **cutoff frequencies** and **wavelengths**

$$f_c = \frac{mc}{2a} \text{ and } \lambda_c = \frac{2a}{m},$$

  respectively, satisfy the zero tangential $\tilde{\mathbf{E}}$ boundary conditions on $x = 0$ and $x = a$ plates of the guide.

  - TE_0 mode does not exist but TM_0=TEM does and it is dispersionless.
  - All TE_m and TM_m modes are dispersive for $m \geq 1$, and propagate only if $f > f_c$, or, equivalently, $\lambda < \lambda_c$.
  - Non-propagating modes are evanescent and have an attenuation constant $|k_z|$.

- Also TE_m and TE_m mode fields have guide impedances

$$\eta_{TE} = \frac{E_y}{-H_x} = \frac{\eta_o}{\sqrt{1-\frac{f_c^2}{f^2}}} \text{ and } \eta_{TM} = \frac{E_x}{H_y} = \eta_o\sqrt{1-\frac{f_c^2}{f^2}}$$

relating the transverse field components of the guided modes.

All the results summarized above are for air-filled waveguides, but they can be readily modified, by replacing c and η_o with c/n and η, respectively, in the case of dielectric-filled waveguides.

Example 1: Consider a dielectric-filled parallel-plate waveguide with $a = 2$ cm. The permeability of the dielectric filling is μ_o and its refractive index is $n = 1.5$.

1. Which TE_m and TM_m modes can propagate a 12 GHz signal in the waveguide?
2. What would be the associated cutoff wavelengths in each case?
3. What would be the associated group velocities in each case? — here assume a modulated 12 GHz carrier with a narrow modulation bandwidth.

Solution: Unguided propagation velocity for the dielectric filling the waveguide is

$$v = \frac{c}{n} = \frac{3\times 10^8\,\mathrm{m/s}}{1.5} = 2\times 10^8\,\frac{\mathrm{m}}{\mathrm{s}}.$$

Using v in place of c in the cutoff frequency formula for TE_m and TM_m modes we find

$$f_c = \frac{mv}{2a} = \frac{m\times 2\times 10^{10}\,\mathrm{cm/s}}{2\times 2\,\mathrm{cm}} = m5\times 10^9\,\mathrm{Hz} = 5m\,\mathrm{GHz}.$$

1. $f = 12$ GHz exceeds the cutoff frequencies of TE_m and TM_m for $m = 1$ and 2, but not 3. Therefore, the propagating (i.e., non-evanescent) modes at $f = 12$ GHz are TM_0, TE_1, TM_1, TE_2, and TM_2.

2. Cutoff wavelength are given by the equation

$$\lambda_c = \frac{2a}{m}$$

and do not depend on the dielectric filling. They are, with $a = 2$ cm,

$$\lambda_c = 4\,\mathrm{cm} \text{ for } \mathrm{TE}_1{=}\mathrm{TM}_1 \text{ and } \lambda_c = 2\,\mathrm{cm} \text{ for } \mathrm{TE}_2{=}\mathrm{TM}_2.$$

The cutoff wavelength is ∞ for TEM mode (which does not have a cutoff condition).

3. Group velocities are given by the equation

$$v_g = v\sqrt{1 - \frac{f_c^2}{f^2}}$$

where $v = c/n$. For the non-dispersive TEM mode with $f_c = 0$ the group velocity is $v_g = 2 \times 10^8$ m/s. For TE_1 and TM_1 modes

$$v_g = v\sqrt{1 - \frac{5^2}{12^2}} = 1.82 \times 10^8\,\mathrm{m/s}.$$

For TE_2 and TM_2 modes

$$v_g = v\sqrt{1 - \frac{10^2}{12^2}} = 1.11 \times 10^8\,\mathrm{m/s}.$$

Example 2: Consider an air-filled parallel-plate waveguide with $a = 3$ cm. Calculate the guide wavelength λ_g or the attenuation rate in dB/cm of the TE_1 mode in the guide — whichever appropriate — if the operating wavelength of the mode is (a) $\lambda = 3$ cm, and (b) $\lambda = 12$ cm.

Solution: The cutoff wavelength of TE_1 mode in the guide is

$$\lambda_c = \frac{2a}{m} = \frac{2 \times 3\,\mathrm{cm}}{1} = 6\,\mathrm{cm}.$$

(a) For $\lambda = 3$ cm, $\lambda < \lambda_c$, and, therefore, the TE_1 mode is propagating. The propagation constant, that is k_z, is

$$k_z = k\sqrt{1 - \frac{\lambda^2}{\lambda_c^2}} = \frac{2\pi}{\lambda}\sqrt{1 - \frac{\lambda^2}{\lambda_c^2}}$$

and the guide wavelength is

$$\lambda_g = \frac{2\pi}{k_z} = \frac{\lambda}{\sqrt{1 - \frac{\lambda^2}{\lambda_c^2}}} = \frac{3\,\mathrm{cm}}{\sqrt{1 - \frac{3^2}{6^2}}} = \frac{3\,\mathrm{cm}}{\sqrt{1 - \frac{1}{4}}} = \frac{3\,\mathrm{cm}}{\sqrt{\frac{3}{4}}} = 2\sqrt{3}\,\mathrm{cm}.$$

(b) For $\lambda = 12$ cm, $\lambda > \lambda_c$, and, therefore, the TE_1 mode is evanescent. The attenuation constant is $|k_z|$, where

$$k_z = k\sqrt{1 - \frac{\lambda^2}{\lambda_c^2}} = \frac{2\pi}{\lambda}\sqrt{1 - \frac{\lambda^2}{\lambda_c^2}} = \frac{2\pi}{12}\sqrt{1 - \frac{12^2}{6^2}} = \frac{\pi}{6}\sqrt{-3} = \pm j\frac{\pi}{2\sqrt{3}}\,\mathrm{rad/cm}.$$

Therefore, the attenuation rate is

$$20\log_{10} e^{|k_z|} = |k_z| 20\log_{10} e = \frac{\pi}{2\sqrt{3}} \times 8.686 \approx 7.88\,\mathrm{dB/cm}.$$

Example 3: A parallel-plate waveguide with $a = 3$ cm is air filled for $z < 0$ but it is filled with a dielectric for $z > 0$ which has $\mu = \mu_o$ and a refractive index $n = 1.5$. If a TE_1 mode wave field with $\lambda = 3$ cm is incident from the air-filled region on the interface at $z = 0$, what fraction of the time-averaged incident power will be transmitted into the $z > 0$ region of the guide?

Solution: The cutoff wavelength of TE_1 mode in the guide is

$$\lambda_c = \frac{2a}{m} = \frac{2 \times 3\,\text{cm}}{1} = 6\,\text{cm}.$$

For $\lambda = 3$ cm, the intrinsic impedance of the TE_1 mode fields is therefore

$$\eta_{TE1} = \frac{\eta_o}{\sqrt{1 - \frac{\lambda^2}{\lambda_c^2}}} = \frac{120\pi}{\sqrt{1 - \frac{3^2}{6^2}}} = \frac{120\pi}{\sqrt{\frac{3}{4}}} = \frac{240\pi}{\sqrt{3}}\,\Omega$$

in the air filled section.

Within the dielectric region the operation wavelength is $\lambda_2 = \lambda/n = 3/1.5 = 2$ cm, and, therefore the intrinsic impedance is

$$\eta_{TE2} = \frac{\eta_o/n}{\sqrt{1 - \frac{\lambda_2^2}{\lambda_c^2}}} = \frac{120\pi/1.5}{\sqrt{1 - \frac{2^2}{6^2}}} = \frac{80\pi}{\sqrt{\frac{8}{9}}} = \frac{240\pi}{\sqrt{8}}\,\Omega.$$

Thus, using a transmission line analogy, the reflection coefficient at the interface is

$$\Gamma = \frac{\eta_{TE2} - \eta_{TE1}}{\eta_{TE2} + \eta_{TE1}} = \frac{\frac{1}{\sqrt{8}} - \frac{1}{\sqrt{3}}}{\frac{1}{\sqrt{8}} + \frac{1}{\sqrt{3}}} = \frac{\sqrt{3} - \sqrt{8}}{\sqrt{3} + \sqrt{8}} = -0.24,$$

which is the transverse electric field amplitude of the reflected wave in the air filled region divided by the incident electric field amplitude.

Consequently, the fraction of the incident time-averaged power reflected back from the interface is the reflectance

$$|\Gamma|^2 \approx 0.058,$$

and

$$1 - |\Gamma|^2 \approx 0.942$$

represents the transmittance, the fraction of the incident time-averaged power transmitted into the dielectric filled region.

30 Exciting and detecting waveguide modes

- Propagating guided-EM waves of any type that can be represented as some weighted superposition of $\mathrm{TE}_m$and TM_m mode fields can be "generated" in practice within waveguides by using dipole or loop antennas inserted within the waveguide, say, at $z = 0$ location, utilized in a "transmission" mode.

- The same waves can also be "detected" within the waveguide at some new location $z > 0$, by employing simple dipole or loop antennas in "reception" modes.

- The detecting or receiving antenna should be oriented to be "co-polarized" with the mode field and placed preferentially at some location where the mode field to be detected has the largest amplitude — e.g., a y-polarized dipole located at $x = a/2$ to optimally detect an incident TE_1 mode field.

- By the *antenna reciprocity theorem* applied to antennas within waveguides, we can infer that the optimal position and orientation of an antenna on the $z = 0$ plane to "excite" some particular waveguide mode such as TE_1 mode propagating in the z-direction would be same exact orientation and position of a receiving antenna in the xy-plane to optimally detect the very same mode field — e.g., a y-polarized dipole located at $x = a/2$ on $z = 0$ plane to optimally excite a propagating TE_1 mode field away from the $z = 0$ plane!

Example 1: Will a dipole antenna placed on the $z = 0$ plane, oriented in the y-direction at $x = a/2$ location in order to *optimally* generate the TE_1 mode fields, *also* be able to generate the TE_2 and TE_3 mode fields in the same waveguide?

Solution: The optimal dipole to excite the TE_1 mode is also optimal to excite the TE_3 mode field but it cannot excite the TE_2 mode field at all. The reason is, $\tilde{\mathbf{E}}$ for the TE_3 mode also maximizes at $x = a/2$, just as for the TE_1 mode, but it is zero for the TE_2 mode. More specifically, a dipole oriented in y-direction at $x = a/2$ location cannot excite the TE_2 mode at all! An optimal location for a y-polarized dipole to excite the TE_2 mode would be $x = a/4$ (why?) — such a dipole would also excite the TE_1 mode, but not optimally (why, again?).

Example 2: Let's consider exciting TE_m modes in a parallel plate waveguide with a current sheet antenna with a surface current density phasor $\tilde{\mathbf{J}}_s = S\hat{y}$ on the $z = 0$ plane. Which TE_m modes will be excited and which ones will not be excited?

Solution: All TE_m modes with odd $m = 1, 3, 5,...$ will be excited with finite amplitudes while no excitation is to be expected for modes with even $m = 2, 4, 6$ ... The reason for this is, for even valued m the mode electric fields $\propto \hat{y}\sin(\frac{m\pi}{a}x)$ will be 50-50 aligned and anti-aligned with $\tilde{\mathbf{J}}_s = S\hat{y}$, causing the cancellation of TE_m modes fields caused by the aligned and anti-aligned portions of $\tilde{\mathbf{J}}_s = S\hat{y}$. See the next example if this answer sounds ambiguous.

Example 3: Consider once more placing a surface current density $\tilde{\mathbf{J}}_s = J_{so}\hat{y}$ on the $z = 0$ plane to excite TE_m modes in a parallel plate waveguide. Let

$$\tilde{\mathbf{E}}_m \equiv \hat{y} 2j E_m e^{-jk_z z} \sin(\frac{m\pi}{a}x)$$

denote the electric field phasors of the TE_m modes with E_m amplitudes. Determine the amplitudes E_m for all m whether the mode is propagating or cut-off.

Solution: We have from Lecture 27

$$\tilde{\mathbf{H}}_m = -\frac{2E_m}{\omega\mu_o} e^{-jk_z z}\{\hat{x} j k_z \sin(k_x x) + \hat{z} k_x \cos(k_x x)\}$$

accompanying $\tilde{\mathbf{E}}_m$, where $k_x = \frac{m\pi}{a}$. Accordingly the total H_x on $z = 0$ plane needing to match half the y-directed surface current density S is the infinite sum

$$H_x(z = 0) = -\frac{2j}{\omega\mu_o} \sum_{m=1}^{\infty} E_m k_z \sin(k_x x).$$

Therefore the unknown mode amplitudes E_m — effectively Fourier series amplitudes of total E_y launched by the surface current J_{so} — can be deduced from the boundary condition equation

$$\sum_{m=1}^{\infty} E_m \sqrt{k^2 - (\frac{m\pi}{a})^2} \sin(\frac{m\pi}{a}x) = \frac{j\omega\mu_o J_{so}}{4}.$$

To extract the E_m's from this equation multiply both sides with $\sin(\frac{n\pi}{a})$ and integrate in x from 0 to a to obtain

$$\sum_{m=1}^{\infty} E_m \sqrt{k^2 - (\frac{m\pi}{a})^2} \underbrace{\int_0^a dx \sin(\frac{n\pi}{a}) \sin(\frac{m\pi}{a}x)}_{\frac{a}{2}\delta_{nm}} = \frac{j\omega\mu_o J_{so}}{4} \underbrace{\int_0^a dx \sin(\frac{n\pi}{a})}_{\frac{2a}{n\pi} \text{ if } n \text{ odd, } 0 \text{ else.}}.$$

Hence

$$E_m = \frac{j\omega\mu_o J_{so}}{m\pi\sqrt{k^2 - (\frac{m\pi}{a})^2}}$$

for odd m and zero for even m.

Notice that $E_m \propto \frac{1}{m^2}$ for large m, indicating that the excitation amplitudes of lower order modes (small m) will be larger than the amplitudes of higher order modes.

Also notice that evanescent modes (large m so that $k^2 - (\frac{m\pi}{a})^2 < 0$) can also have non-zero amplitudes E_m but they will be unable to transfer energy down the waveguide (unless the waveguide is of a finite length and energy tunneling becomes possible).

- The fact that the ω vs k dispersion curve for a plasma is the same as the ω vs k_z dispersion curve for a waveguide is interesting and convenient. We may examine plasmas to infer waveguide phenomena or vice versa, infer plasma phenomena from waveguide solutions.

- Take the problem of the excitation of evanescent modes within waveguides of finite length and the associated tunneling phenomena. We can get a grip on this phenomenon by modeling radiation from an infinite current sheet antenna operating in a finite width plasma slab sandwiched between a PEC reflector and vacuum (extending to infinity) — we examine this situation in the next example.

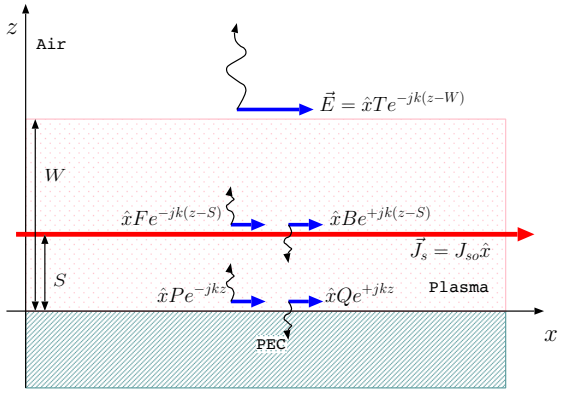
Example 4: An infinite current sheet antenna with current density $\tilde{\mathbf{J}}_s = J_{so}\hat{x}$ is positioned on $z = S$ surface a distance S above a PEC reflector and is at a distance $W - S$ below the plasma to air interface as shown in the margin. We wish to model the outgoing wave electric field amplitude T in terms of the source current strength J_{so} as well as f/f_p . We will do so by appropriately matching the tangential $\mathbf{E}$ and $\mathbf{H}$ at the boundaries $z = S$ and $z = W$, as well as requiring that tangential $\mathbf{E}$ vanishes at $z = 0$.

- In Region 1, $0 < z < S$, we have

$$\tilde{\mathbf{E}}_1(z) = \hat{x}(Pe^{-jkz} + Qe^{+jkz}) \text{ and } \tilde{\mathbf{H}}_1(z) = \hat{y}\frac{Pe^{-jkz} - Qe^{+jkz}}{\eta},$$

where $k = k_o n$ and $\eta = \frac{\eta_o}{n}$, with $k_o \equiv \frac{\omega}{c}$ and $n = \sqrt{1 - \frac{\omega_p^2}{\omega(\omega - j\nu)}}$ in terms of the plasma frequency ω_p and electron collision frequency ν. Since $\tilde{\mathbf{E}}_1(0) = \hat{x}(P + Q) = 0$ we need $Q = -P$ and consequently

$$\tilde{\mathbf{E}}_1(z) = \hat{x}P(e^{-jkz} - e^{+jkz}) \text{ and } \tilde{\mathbf{H}}_1(z) = \hat{y}\frac{P}{\eta}(e^{-jkz} + e^{+jkz}).$$



A plasma slab containing a current sheet radiator — we study the outgoing wave amplitude T and internal field distributions of the plasma as a function of f/f_p ...

- In Region 2, $S < z < W$, we have

$$\tilde{\mathbf{E}}_2(z) = \hat{x}(Fe^{-jk(z-S)} + Be^{+jk(z-S)}) \text{ and } \tilde{\mathbf{H}}_2(z) = \hat{y}\frac{Fe^{-jk(z-S)} - Be^{+jk(z-S)}}{\eta}$$

and we require $\tilde{\mathbf{E}}_2(S) = \tilde{\mathbf{E}}_1(S)$ as well as and $\hat{z} \times (\tilde{\mathbf{H}}_2(S) - \tilde{\mathbf{H}}_1(S)) = J_{so}\hat{x}$ leading to

$$F + B = P(e^{-jkS} - e^{+jkS}) \text{ and } \frac{F-B}{\eta} - \frac{P}{\eta}(e^{-jkS} + e^{+jkS}) = -J_{so}$$

to be utilized after we use $B = \Gamma_g F$ obtained by using B.C.'s on $z = W$ surface next.

- In Region 3, $z > W$, we have

$$\tilde{\mathbf{E}}_3(z) = \hat{x}Te^{-jk_o(z-W)} \text{ and } \tilde{\mathbf{H}}_3(z) = \hat{y}\frac{T}{\eta_o}e^{-jk_o(z-W)}$$

and we require that $\tilde{\mathbf{E}}_3(W) = \tilde{\mathbf{E}}_2(W)$ as well as $\tilde{\mathbf{H}}_3(W) = \tilde{\mathbf{H}}_2(W)$ leading to

$$T = Fe^{-jk(W-S)} + Be^{+jk(W-S)} \text{ and } \frac{T}{\eta_o} = \frac{Fe^{-jk(W-S)} - Be^{+jk(W-S)}}{\eta}.$$

Combining using the last two equations we get

$$B = \underbrace{\frac{\eta_o - \eta}{\eta_o + \eta}}_{\Gamma} e^{-j2k(W-S)}F \equiv \Gamma_g F$$

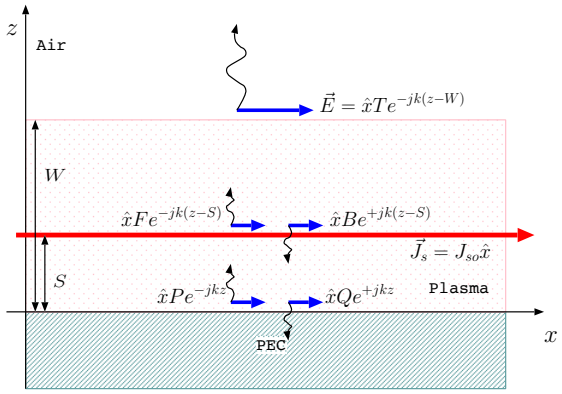
and using $B = \Gamma_g F$ in $-J_{so}$ expression we get

$$F = \frac{-\eta J_{so}}{(1-\Gamma_g) - (1+\Gamma_g)\frac{1+e^{+j2kS}}{1-e^{+j2kS}}}.$$

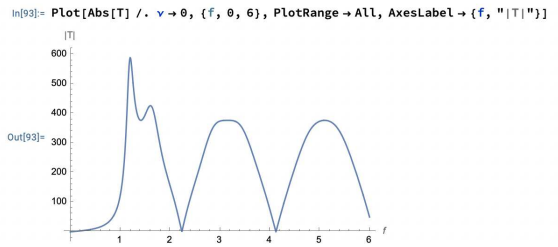
Having expressed F in terms of J_{so}, the remaining wave amplitudes can be computed using

$$B = \Gamma_g F,\ P = \frac{F+B}{e^{-jkS} - e^{+jkS}},\ \text{and}\ T = Fe^{-jk(W-S)} + Be^{+jk(W-S)}.$$

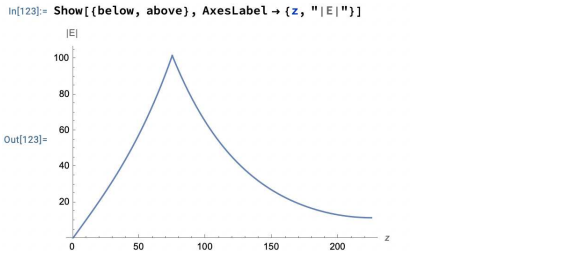
- The results from Example 4 were used in a Mathematica notebook to produce the plots shown in the margin.
  - The results are singular at $\omega = \omega_p$ unless a small collision frequency ν is employed in the computation of n. Real plasmas always have some small ν which can be ignored in most applications when $|\omega - \omega_p| \gg \nu$ to work in the *collisionless plasma approximation*.
  - To be consistent with the field definitions we made above, the root of complex n^2 with positive real part is used as n unless n is purely imaginary and in that case negative imaginary n is employed in computing k and η. This way F and Q-wave aplitudes decay away from the $z = S$ source plane when the plasma is in cutoff, that is $\omega < \omega_p$.
  - The total field in Region 2 is decaying with increasing $z > S$ when $\omega < \omega_p$ when F and B waves are evanescent since B amplitude is constrained as $\Gamma_g F$ (with $|\Gamma| = 1$) when $\omega < \omega_p$, that is, Channel 2 is "cut off" but tunnels some energy out to Region 3!!



A plasma slab containing a current sheet radiator — we study the outgoing wave amplitude T and internal field distributions of the plasma as a function of f/f_p ...

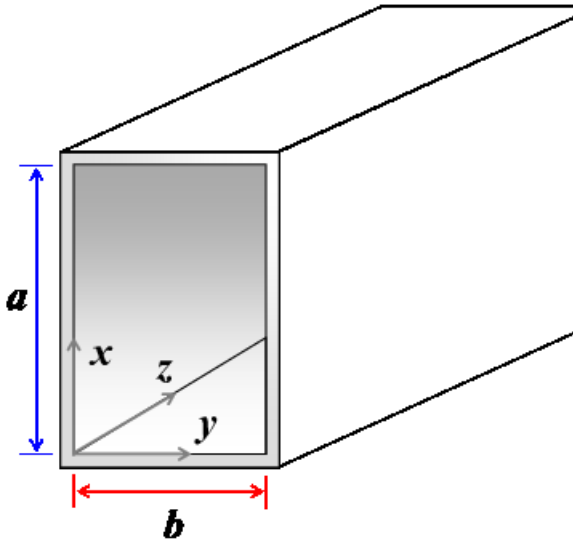


Plot of $|T|$ and as a function of f in MHz for $f_p = 1$ MHz and $J_{so} = 1$ A/m.



Total field amplitudes within the plasma slab as a fucction of z when $f = 0.5f_p$, $S = 75$ m, and $W = 3S$.

31 TM_{mn} modes in rectangular waveguides



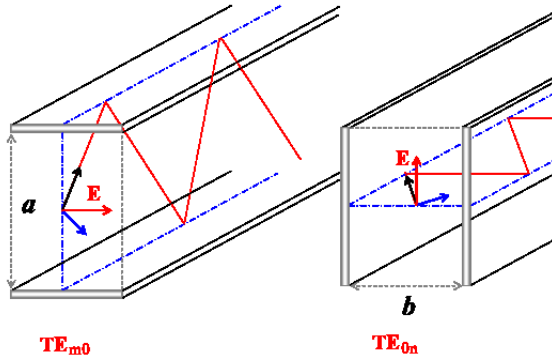
When the operation frequency f in a parallel-plate waveguide exceeds the cutoff frequency $f_c = \frac{c}{2a}$ of the TE_1 mode, dual- or multi-mode operations become unavoidable in the guide.

Single-mode operation at high frequencies can be attained by turning off the guided TEM(=TM_0) mode by introducing a pair of new plates on, say, $y = 0$ and $y = b$ planes as shown in the margin. This configuration is known as the "rectangular waveguide", which is the subject of the next set of lectures.

- Briefly, the guided TEM mode is suppressed in the rectangular waveguide, and propagation is only possible in terms of TM_{mn} and TE_{mn} modes. By definition:

  1. $H_z = 0$ for TM_{mn} mode, for which the mode properties can be derived from a non-zero $E_z(x, y, z) = f(x, y)e^{-jk_z z}$;
  2. $E_z = 0$ for TE_{mn} mode, for which the mode properties can be derived from a non-zero $H_z(x, y, z) = f(x, y)e^{-jk_z z}$;

     where the constraints on $f(x, y)$ and k_z are to be determined from Maxwell's equations and the relevant boundary conditions.



- Both TM_{mn} and TE_{mn} modes consist of the superposition of free-propagating TEM wave fields reflecting from the guide walls and satis-

fying the well-known vector wave equations

$$\nabla^2\tilde{\mathbf{E}} + \omega^2\mu_o\epsilon_o\tilde{\mathbf{E}} = 0 \quad \text{and} \quad \nabla^2\tilde{\mathbf{H}} + \omega^2\mu_o\epsilon_o\tilde{\mathbf{H}} = 0$$

derived from (see margin) Maxwell's equations.

Vector wave equation in phasor form: Taking the curl of Faraday's law

$$\nabla \times \tilde{\mathbf{E}} = -j\omega\mu_o\tilde{\mathbf{H}},$$

and using

$$\begin{aligned} \nabla \times \nabla \times \tilde{\mathbf{E}} &= \nabla(\nabla \cdot \tilde{\mathbf{E}}) - \nabla^2\tilde{\mathbf{E}}, \\ \nabla \cdot \tilde{\mathbf{E}} &= 0, \\ \nabla \times \tilde{\mathbf{H}} &= j\omega\epsilon_o\tilde{\mathbf{E}}, \end{aligned}$$

it follows that

$$\nabla^2\tilde{\mathbf{E}} + \omega^2\mu_o\epsilon_o\tilde{\mathbf{E}} = 0.$$

Likewise,

$$\nabla^2\tilde{\mathbf{H}} + \omega^2\mu_o\epsilon_o\tilde{\mathbf{H}} = 0.$$

TM_{mn} modes:

- To examine the TM_{mn} mode with

$$H_z = 0 \quad \text{and} \quad E_z(x, y, z) = f(x, y)e^{-jk_z z}$$

consider the z-component of the wave-equation for the electric field, namely

$$\nabla^2 E_z + k^2 E_z = 0,$$

where

$$k^2 \equiv \omega^2\mu_o\epsilon_o \quad \text{and} \quad \nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

Substituting E_z into the wave-equation component we have

$$(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2})f(x, y)e^{-jk_z z} + k^2 f(x, y)e^{-jk_z z} = 0,$$

from which

$$(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2})f(x, y)e^{-jk_z z} + (-jk_z)^2 f(x, y)e^{-jk_z z} + k^2 f(x, y)e^{-jk_z z} = 0$$

or

$$(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2})f(x, y) + (k^2 - k_z^2)f(x, y) = 0.$$

- We will next solve this 2D pdf using the method of **separation of variables**. In this method we assume that

$$f(x,y) = X(x)Y(y),$$

that is, we assume[1] that 2D function $f(x,y)$ of variables x and y is a product of 1D functions $X(x)$ and $Y(y)$ of x and y, respectively. With this assumption, the pdf above takes the form

$$YX'' + XY'' + (k^2 - k_z^2)XY = 0 \;\; \Rightarrow \;\; \frac{X''}{X} + \frac{Y''}{Y} + (k^2 - k_z^2) = 0$$

where

$$X'' \equiv \frac{\partial^2 X}{\partial x^2} \;\text{ and }\; Y'' \equiv \frac{\partial^2 Y}{\partial y^2}$$

  - Since $(k^2 - k_z^2)$ is independent x and y, it follows from the above pdf that X''/X as well as Y''/Y are constants independent of spatial coordinates. Thus we can write

$$\frac{X''}{X} = -k_x^2 \;\; \Rightarrow \;\; \frac{\partial^2 X}{\partial x^2} + k_x^2 X = 0$$

where k_x is some constant. Also, by the same argument,

$$\frac{Y''}{Y} = -k_y^2 \;\; \Rightarrow \;\; \frac{\partial^2 Y}{\partial y^2} + k_y^2 Y = 0,$$

$$(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2})f+(k^2-k_z^2)f = 0$$

2.29 cm by 1.012 cm Standard X-band (8.2-12.4 GHz) waveguide in which only TE_{10} mode is non-evanescent within X-band.

[1]This may *appear* to be a restricting assumption but if it leads to a *complete* set of solutions (modes) — as it does, as we will see — to be able to express all possible solutions as a weighted sum then the method is fully justified — see footnote 2.

where k_y is some other constant. Furthermore, utilizing both of these conditions within

$$\frac{X''}{X} + \frac{Y''}{Y} + (k^2 - k_z^2) = 0$$

we get

$$-k_x^2 - k_y^2 + (k^2 - k_z^2) = 0 \quad \Rightarrow \quad k_z = \sqrt{k^2 - k_x^2 - k_y^2}.$$

$$\frac{\partial^2 X}{\partial x^2} + k_x^2 X = 0$$

$$\frac{\partial^2 Y}{\partial y^2} + k_y^2 Y = 0,$$

- We continue by noting that the 2nd order ODEs for $X(x)$ and $Y(y)$ above are solved by

$$X(x) = A\cos k_x x + B\sin k_x x \quad \text{and} \quad Y(y) = C\cos k_y y + D\sin k_y y.$$

These general solutions with constants A, B, C, D simplify when we apply the boundary conditions that $X(x)Y(y) = 0$ at $x = 0$ and $y = 0$ as follows:

- $X(0) = 0$ implies $A = 0$, and in turn $X(x) = B\sin k_x x$;
- $Y(0) = 0$ implies $C = 0$, and in turn $Y(y) = D\sin k_y y$;

Furthermore,

- $X(a) = 0$ implies $k_x a = m\pi$, $m = 1, 2, 3, \cdots$
- $Y(b) = 0$ implies $k_y b = n\pi$, $n = 1, 2, 3, \cdots$

– Combining the above results, we get

$$f(x, y) = X(x)Y(y) = E_o \sin(k_x x)\sin(k_y y)$$

and consequently

$$E_z(x,y,z) = E_o \sin(k_x x)\sin(k_y y)e^{-jk_z z},$$

with

$$k_x = \frac{m\pi}{a},\ \ k_y = \frac{n\pi}{b},\ \ k_z = \frac{\omega}{c}\sqrt{1-\frac{k_x^2+k_y^2}{k^2}} = \frac{\omega}{c}\sqrt{1-\frac{f_c^2}{f^2}},$$

where

$$f_c = \sqrt{(\frac{mc}{2a})^2 + (\frac{nc}{2b})^2}$$

is the pertinent cutoff frequency of the TM_{mn} mode with $m, n \geq 0$.

Cutoff wavelength: As usual we have

$$\frac{\lambda_c}{\lambda} = \frac{f}{f_c}$$

and hence

$$\lambda_c = \frac{\lambda f}{\sqrt{(\frac{mc}{2a})^2 + (\frac{nc}{2b})^2}} = \frac{1}{\sqrt{(\frac{m}{2a})^2 + (\frac{n}{2b})^2}}.$$

– Note that neither $m = 0$ nor $n = 0$ are permitted with non-zero E_z. Thus TM_{m0} and TM_{0n} modes don't exist[2].

[2]The most general TM solution for the waveguide can be expressed as a weighted infinite sum, which takes the form, for $z = 0$,

$$E_z(x,y,0) = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} E_{m,n}\sin(\frac{m\pi}{a}x)\sin(\frac{n\pi}{b}y),$$

which is a 2D *Fourier series* representation with *Fourier coefficients* $E_{m,n}$ of an arbitrary periodic $f(x,y)$ over the xy-plane matching the boundary conditions $f(0,y) = f(a,y) = f(x,0) = f(x,b) = 0$. This attests to the *completeness* of the separation of variables solution method employed in this section since *all* periodic functions can be represented in Fourier series form as we learned in ECE 210.

Transverse field components:

Above, we have obtained the dispersion relation for TM_{mn} mode in rectangular waveguides. The dispersion characteristics of these modes are identical to those we have discussed in connection with parallel-plate waveguides except for the generalized expression for f_c.

- Given $E_z(x, y, z)$ determined above as well as the fact that $H_z = 0$ (by assumption), transverse field components of TM_{mn} mode waves can be inferred from Faraday's and Ampere's laws as shown next:

  - With field components varying with z according to $e^{-jk_z z}$, Faraday's law implies

$$\nabla \times \tilde{\mathbf{E}} = \begin{vmatrix} \hat{x} & \hat{y} & \hat{z} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & -jk_z \\ E_x & E_y & E_z \end{vmatrix} = -j\omega\mu_o(H_x, H_y, H_z),$$

    from which

$$H_x = \frac{\frac{\partial E_z}{\partial y} + jk_z E_y}{-j\omega\mu_o}, \quad H_y = \frac{\frac{\partial E_z}{\partial x} + jk_z E_x}{j\omega\mu_o}, \quad H_z = \frac{\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}}{-j\omega\mu_o}.$$

  - Likewise, Ampere's law implies

$$\nabla \times \tilde{\mathbf{H}} = \begin{vmatrix} \hat{x} & \hat{y} & \hat{z} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & -jk_z \\ H_x & H_y & H_z \end{vmatrix} = j\omega\epsilon_o(E_x, E_y, E_z),$$

from which

$$E_x = \frac{\frac{\partial H_z}{\partial y} + jk_z H_y}{j\omega\epsilon_o}, \quad E_y = \frac{\frac{\partial H_z}{\partial x} + jk_z H_x}{-j\omega\epsilon_o}, \quad E_z = \frac{\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y}}{j\omega\epsilon_o}.$$

- Now (as confirmed in HW),

$$H_x = \frac{\frac{\partial E_z}{\partial y} + jk_z E_y}{-j\omega\mu_o} \text{ and } E_y = \frac{\frac{\partial H_z}{\partial x} + jk_z H_x}{-j\omega\epsilon_o}$$

from above imply that

$$H_x = -\frac{jk_z\frac{\partial H_z}{\partial x} - j\omega\epsilon_o\frac{\partial E_z}{\partial y}}{k^2 - k_z^2} \text{ and } E_y = -\frac{jk_z\frac{\partial E_z}{\partial y} - j\omega\mu_o\frac{\partial H_z}{\partial x}}{k^2 - k_z^2}$$

and, likewise,

$$H_y = \frac{\frac{\partial E_z}{\partial x} + jk_z E_x}{j\omega\mu_o} \text{ and } E_x = \frac{\frac{\partial H_z}{\partial y} + jk_z H_y}{j\omega\epsilon_o}$$

imply that

$$H_y = -\frac{jk_z\frac{\partial H_z}{\partial y} + j\omega\epsilon_o\frac{\partial E_z}{\partial x}}{k^2 - k_z^2} \text{ and } E_x = -\frac{jk_z\frac{\partial E_z}{\partial x} + j\omega\mu_o\frac{\partial H_z}{\partial y}}{k^2 - k_z^2}.$$

- The expressions above provide the transverse field components in terms of transverse derivatives of longitudinal components E_z and H_z.
  - By setting $H_z = 0$, they yield the transverse field components for TM_{mn} modes shown in the margin.

TM mode fields:

$$H_x = \frac{j\omega\epsilon_o\frac{\partial E_z}{\partial y}}{k^2 - k_z^2},$$

$$H_y = \frac{-j\omega\epsilon_o\frac{\partial E_z}{\partial x}}{k^2 - k_z^2},$$

$$E_x = \frac{-jk_z\frac{\partial E_z}{\partial x}}{k^2 - k_z^2},$$

$$E_y = \frac{-jk_z\frac{\partial E_z}{\partial y}}{k^2 - k_z^2}.$$

TE mode fields:

$$E_x = \frac{-j\omega\mu_o\frac{\partial H_z}{\partial y}}{k^2 - k_z^2},$$

$$E_y = \frac{j\omega\mu_o\frac{\partial H_z}{\partial x}}{k^2 - k_z^2},$$

$$H_x = \frac{-jk_z\frac{\partial H_z}{\partial x}}{k^2 - k_z^2},$$

$$H_y = \frac{-jk_z\frac{\partial H_z}{\partial y}}{k^2 - k_z^2}.$$

Also,

- By setting $E_z = 0$, they yield the transverse field components for TE_{mn} modes also shown in the margin.

32 TE_{mn} modes in rectangular waveguides

- The analysis of TE_{mn} modes starts with the wave equation for H_z, that is

$$\nabla^2 H_z + k^2 H_z = 0.$$

In analogy with the TM_{mn} case, and using separation of variables, we have

$$H_z(x,y,z) = (A\cos k_x x + B\sin k_x x)(C\cos k_y y + D\sin k_y y)e^{-jk_z z}.$$

Pertinent boundary conditions need to be applied in terms of E_y and E_x on waveguide walls at $x = 0$ and a, and $y = 0$ and b, respectively:

1. $E_y = 0$ at $x = 0$ and a requires $\frac{\partial H_z}{\partial x} = 0$ at the same locations, implying $B = 0$ and $k_x a = m\pi$.
2. $E_x = 0$ at $y = 0$ and b requires $\frac{\partial H_z}{\partial y} = 0$ at the same locations, implying $D = 0$ and $k_y b = n\pi$.

Hence,

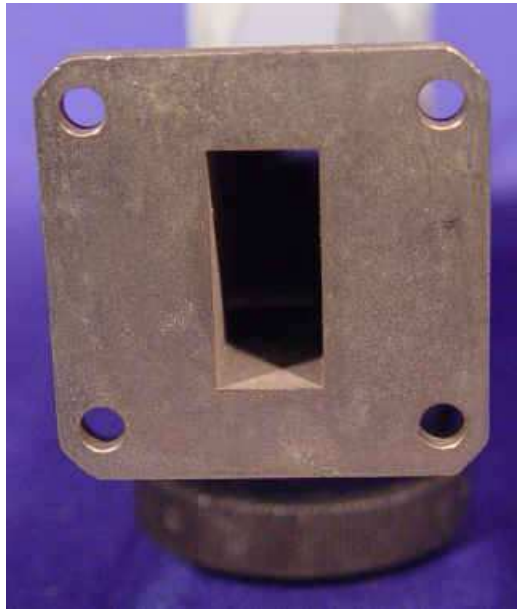
$$H_z(x,y,z) = H_o\cos(k_x x)\cos(k_y y)e^{-jk_z z},$$

with

$$k_x = \frac{m\pi}{a},\quad k_y = \frac{n\pi}{b},\quad k_z = \frac{\omega}{c}\sqrt{1-\frac{k_x^2+k_y^2}{k^2}} = \frac{\omega}{c}\sqrt{1-\frac{f_c^2}{f^2}},$$

TE mode fields:

$$\begin{aligned}
E_x &= \frac{-j\omega\mu_o\frac{\partial H_z}{\partial y}}{k^2-k_z^2},\\
E_y &= \frac{j\omega\mu_o\frac{\partial H_z}{\partial x}}{k^2-k_z^2},\\
H_x &= \frac{-jk_z\frac{\partial H_z}{\partial x}}{k^2-k_z^2},\\
H_y &= \frac{-jk_z\frac{\partial H_z}{\partial y}}{k^2-k_z^2}.
\end{aligned}$$



where

$$f_c = \sqrt{(\frac{mc}{2a})^2 + (\frac{nc}{2b})^2}$$

is the pertinent cutoff frequency of the TE_{mn} mode.

- Note that $m = 0$ or $n = 0$ — but not both zero — are permitted since these choices do not lead to trivial H_z.
- However, $m = n = 0$ is not permitted, because in that case H_z becomes independent of x and y, and leads to zero transverse fields (see Example 1 for the full reason).

TE mode fields:

$$E_x = \frac{-j\omega\mu_o \frac{\partial H_z}{\partial y}}{k^2 - k_z^2},$$

$$E_y = \frac{j\omega\mu_o \frac{\partial H_z}{\partial x}}{k^2 - k_z^2},$$

$$H_x = \frac{-jk_z \frac{\partial H_z}{\partial x}}{k^2 - k_z^2},$$

$$H_y = \frac{-jk_z \frac{\partial H_z}{\partial y}}{k^2 - k_z^2}.$$

Example 1: Determine the transverse field components for the TE_{mn} mode explicitly by differentiating

$$H_z(x, y, z) = H_o \cos(k_x x) \cos(k_y y) e^{-jk_z z}$$

and using the relations in the margin. Show that the fields for TE_{00} are trivial while the fields TE_{m0} are finite.

Solution: We have,

$$H_x = \frac{-jk_z \frac{\partial H_z}{\partial x}}{k^2 - k_z^2} = \frac{jk_z H_o k_x \sin(k_x x) \cos(k_y y) e^{-jk_z z}}{k_x^2 + k_y^2},$$

$$H_y = \frac{-jk_z \frac{\partial H_z}{\partial y}}{k^2 - k_z^2} = \frac{jk_z H_o k_y \cos(k_x x) \sin(k_y y) e^{-jk_z z}}{k_x^2 + k_y^2},$$

$$E_y = \frac{j\omega\mu_o \frac{\partial H_z}{\partial x}}{k^2 - k_z^2} = \frac{-j\omega\mu_o H_o k_x \sin(k_x x)\cos(k_y y)e^{-jk_z z}}{k_x^2 + k_y^2},$$

$$E_x = \frac{-j\omega\mu_o \frac{\partial H_z}{\partial y}}{k^2 - k_z^2} = \frac{j\omega\mu_o H_o k_y \cos(k_x x)\sin(k_y y)e^{-jk_z z}}{k_x^2 + k_y^2}.$$

For TE_{m0} we have $k_y = \frac{n\pi}{b} = 0$ and, therefore,

$$H_x = \frac{jk_z H_o \sin(k_x x)e^{-jk_z z}}{k_x}, \;\; H_y = 0, \;\; E_y = \frac{-j\omega\mu_o H_o \sin(k_x x)e^{-jk_z z}}{k_x}, \;\; E_x = 0.$$

Now, we obtain the TE_{00} field from these by setting $k_x = 0$ using L'Hospital's law, leading to

$$H_x = jk_z H_o x e^{-jk_z z}, \;\; H_y = 0, \;\; E_y = -j\omega\mu_o H_o x e^{-jk_z z}, \;\; E_x = 0;$$

these violate the boundary condition of zero E_y at $x = a$ unless $H_o = 0$, which is of course the trivial solution.

Detecting and exciting waveguide modes within rectangular waveguides:

- Guided EM waves of any type that can be represented as some weighted superposition of $\text{TE}_{mn}$and TM_{mn} mode fields can be "generated" in practice within rectangular waveguides by using simple dipole or loop antennas inserted within the waveguide, say, at $z = 0$ location, utilized in a "transmission" mode.

- The same waves can also be "detected" within the waveguide at some new location $z > 0$, by employing simple dipole or loop antennas in "reception" modes.

- The *detecting* or *receiving* antenna should be oriented to be "co-polarized" with the mode field and placed preferentially at some location where the mode field to be *detected* has the largest amplitude — e.g., a y-polarized dipole located at $x = a/2$ to optimally detect an incident TE_{10} mode field.

- By **antenna reciprocity** the *transmitting* antenna should also be oriented to be "co-polarized" with the mode field and placed preferentially at some location where the mode field to be *excited* has the largest amplitude — e.g., a y-polarized dipole located at $x = a/2$ to optimally excite and launch a TE_{10} mode field.

- In general orient and position the **dipole** to be used for "exciting" the waveguide — usually such dipoles are called "probes" — in a particular

$\text{TE}_{mn}$and TM_{mn} mode so that it lies parallel to the strongest **E**-field vector on the xy-plane of the mode.

  - The application of this rule requires the knowledge of transverse **E**-field components E_x and E_y, which can be inferred from the transverse field equations derived in Lecture 31.
  - Alternatively, use the TE and TM mode transverse-plane field plots from *Lee et al. [1985]* — see HW!

- Loop antennas are *magnetic* dipole antennas!

  - A loop centered about the origin, with the current flowing clockwise around the z-axis (in the xy-plane) when looking in the z-direction, is a z-polarized magnetic dipole.

When using **loop antennas**, orient and position the loop within a waveguide so that *as a magnetic dipole* it lies parallel to the strongest **H**-field vector of the desired $\text{TE}_{mn}$and TM_{mn} mode on the xy-plane.

  - The application of this rule requires the knowledge of transverse **H**-field components H_x and H_y, which can be inferred from the transverse field equations derived in Lecture 31.
  - Alternatively, use the TE and TM mode transverse-plane field plots from *Lee et al. [1985]* :-)

Waveguide design and application examples:

Example 2: Design a rectangular air-filled wave guide for single-mode transmission of the frequency band 3.75 GHz – 4.25 GHz in TE_{10} mode. That is, select the dimensions a and $b \leq a$ of the waveguide so that only the TE_{10} mode is propagating in the guide within the specified frequency band while cross sectional area ab is as large as possible for purposes of the power transmission capacity of the guide.

Solution: First, to make sure that TE_{10} mode is propagating in the band for $f > 3.75$ GHz, we need

$$f_c = \sqrt{(\frac{mc}{2a})^2 + (\frac{nc}{2b})^2}|_{\substack{m=1 \\ n=0}} = \frac{c}{2a} < 3.75 \times 10^9 \,\mathrm{Hz}$$

from which we get

$$a > \frac{3 \times 10^{10} \,\mathrm{cm/s}}{2 \times 3.75 \times 10^9 \,/\mathrm{s}} = 4 \,\mathrm{cm}.$$

With $a = 4$ cm, the cutoff frequency of TE_{20} mode will be 7.5 GHz, which is safely outside our band of interest. Of course with $a > 4$ cm TE_{20} cutoff frequency will be less than 7.5 GHz, and we can afford reducing it to as small as 4.25 GHz by selecting

$$a = \frac{cm}{2f_{cm0}}|_{m=2} = \frac{3 \times 10^{10} \times 2}{2 \times 4.25 \times 10^9} = \frac{30}{4.25} = 7.06 \,\mathrm{cm}.$$

To ensure single mode operation in 3.75 GHz – 4.25 GHz band we also need for TE_{01} mode a cutoff frequency

$$f_c = \sqrt{(\frac{mc}{2a})^2 + (\frac{nc}{2b})^2}|_{\substack{m=0 \\ n=1}} = \frac{c}{2b} > 4.25 \times 10^9 \,\mathrm{Hz}$$

yielding

$$b < \frac{3 \times 10^{10}\,\text{cm/s}}{2 \times 4.25 \times 10^{9}\,/\text{s}} = 3.53\,\text{cm}.$$

Hence, a design with maximum possible ab for the specified band works out to have $a = 7.06$ cm and $b = a/2 = 3.53$ cm.

Example 3: Re-design the waveguide in Example 2 for the frequency band 3.75 GHz – 4.25 GHz to include some safety margins as follows: Select the dimensions of the wave guide such that the lowest frequency of the band is at least 20% above the cutoff frequency of the fundamental mode (TE_{10}), and the highest frequency of the band is at least 20% lower than the cutoff frequency of the next higher-order mode. **This provision is for *preventing* the upper sideband frequencies of a modulated TE10 carrier at 4.25 GHz frequency from launching undesired TE20 and TE01 propagation!**

Solution: We already have the lowest frequency of the band, 3.75 GHz, more than 20% above the TE_{10} cutoff frequency $c/2a = 2.125$ GHz — therefore at first it appears that $a = 7.06$ cm can remain as is.

But b clearly has to change. To select b, let 4.25 GHz be 0.8 times the cutoff frequency of the TE_{01} mode. Hence

$$4.25 \times 10^{9} = 0.8\frac{c}{2b} \Rightarrow \quad b = \frac{0.8 \times 3 \times 10^{10}\,\text{cm/s}}{2 \times 4.25 \times 10^{9}\,/\text{s}} = 2.82\,\text{cm}.$$

But then we realize that with $a = 7.06$ cm, 4.25 GHz is still the cutoff frequency the of the TE_{20} mode, which is no longer permissible because a safety margin is

needed — TE_{20} cutoff frequency also needs to be moved up by the same margin as TE_{01}; this can be realized by taking the new modified a as $a = 2b = 5.64$ cm. The corresponding TE_{10} mode cutoff frequency is

$$f_c = \frac{mc}{2a}|_{m=1} = \frac{c}{2a} = \frac{c}{4b} = \frac{30 \times 10^9}{4 \times 2.82} = 2.65\,\text{GHz}$$

and 3.75 GHz is still more than 20% above this.

In conclusion, with $a = 5.64$ cm and $b = 2.82$ cm we have the required modified dimensions and safety margins.

Example 4: The waveguide of Example 3 is to be used as an attenuator for the next (non-propagating) higher-order mode. What is the minimum attenuation rate for the mode in dB/cm over the band 3.75 GHz – 4.25 GHz?

Solution: The next higher-order modes are TE_{01} and TE_{20} having equal cutoff frequencies because $a = 2b$.
The attenuation of these modes will be less severe at $f = 4.25$ GHz than at 3.75 GHz. We have, for these modes, at $f = 4.25$ GHz,

$$k_z = k\sqrt{1 - (\frac{f_c}{f})^2} = k\sqrt{1 - (\frac{1}{0.8})^2} = k\sqrt{1 - (\frac{5}{4})^2} = -jk\frac{3}{4}.$$

Since, at $f = 4.25$ GHz,

$$\lambda = \frac{3 \times 10^{10}}{4.25 \times 10^9} = \frac{30}{4.25}\ \mathrm{cm} \ \Rightarrow\ k = \frac{2\pi}{\lambda} = \frac{4.25\pi}{15}\,\mathrm{rad/cm},$$

we have

$$|k_z| = \frac{3}{4}k = \frac{3}{4}\frac{4.25\pi}{15} = \frac{4.25\pi}{20}\,\frac{\mathrm{Np}}{\mathrm{cm}}.$$

Consequently, the attenuation rate is

$$20\log_{10} e^{|k_z|} = 4.25\pi \log_{10} e = 5.7986\,\mathrm{dB/cm}.$$

The attenuation rate will be larger at frequencies smaller than 4.25 GHz within the band.

Example 5: In the examples above we considered the propagation of signals within the 3.75 GHz – 4.25 GHz frequency band in TE_{10} mode. In this example we focus on a 4 GHz signal at the band center and contrast its properties within the rectangular waveguide designed in Example 2 and inside a parallel plate waveguide (a transmission line in effect) of equal dimensions, $a = 7.06$ cm and $b = w = a/2$, and later with suitably modified dimensions.

Discussion: The cutoff frequency of TE_1 and TM_1 modes in the parallel plate waveguide will be the same as TE_{10} mode cutoff frequency in the designed rectangular waveguide, namely $\frac{c}{2a} = \frac{3\times10^{10}\,\text{cm/s}}{2\times7.06\,\text{cm}} = 2.125$ GHz and therefore at 4 GHz frequency the parallel plate waveguide will be operating with all of TM_0, TM_1 and TE_1 in propagating state, causing symbol ambiguity with modulated input.

The ambiguity can be eliminated by adjusting a so that $\frac{c}{2a} > 4$ GHz and hence $a < \frac{3\times10^{10}\,\text{cm/s}}{2\times4\times10^{9}\,1/\text{s}} = 3.75$ cm, which is about half the size for the rectangular waveguide designed for TE_{10} mode. But now we have a two-wire transmission line with a wire separation of $a = 3.75$ cm which is exactly $\lambda/2$ at 4 GHz. With equal and opposite currents on two wires $\lambda/2$ apart, the system will radiate rather than guide at 4 GHz and is really to be used at frequencies $\ll 4$ GHz.

To avoid radiation at 4 GHz you can reduce a of the parallel plate TL to $\ll 3.75$ cm, but then the power transported will be reduced $\ll$ than the power transported by the TE_{10} mode waveguide with $a = 7.06$ cm. This is because power transported is proportional to ab, the cross-sectional area of the guide.

To avoid radiation at 4 GHz a better solution is to convert your parallel plate TL with $a = 3.75$ cm size into a coax with an outer radius of the same scale, i.e., $b = 3.75$ cm outer radius, to transport about the same power as the TE_{10} mode waveguide with $a = 7.06$ cm. But then the coax will be more lossy than the waveguide because of the smaller surface area, hence larger resistance, of the inner conductor of the coax (outer conductor and waveguide wall losses will be comparable).

If/when the loss-causing inner conductor of the coax is removed to reduce the losses the coax becomes a cylindrical waveguide! The dominant mode bandwidth of the waveguide can subsequently be inreased by reshaping the waveguide as a 2:1 rectangle, which is a nice configuration.

Overall, the larger the f the better is the waveguide solution in a microwave circuit given EMI/EMC as well as power loss concerns.

33 Guide impedance and TL analogies

TE mode fields:

$$E_x = \frac{-j\omega\mu_o \frac{\partial H_z}{\partial y}}{k^2 - k_z^2},$$

$$E_y = \frac{j\omega\mu_o \frac{\partial H_z}{\partial x}}{k^2 - k_z^2},$$

$$H_x = \frac{-jk_z \frac{\partial H_z}{\partial x}}{k^2 - k_z^2},$$

$$H_y = \frac{-jk_z \frac{\partial H_z}{\partial y}}{k^2 - k_z^2}.$$

TM mode fields:

$$H_x = \frac{j\omega\epsilon_o \frac{\partial E_z}{\partial y}}{k^2 - k_z^2},$$

$$H_y = \frac{-j\omega\epsilon_o \frac{\partial E_z}{\partial x}}{k^2 - k_z^2},$$

$$E_x = \frac{-jk_z \frac{\partial E_z}{\partial x}}{k^2 - k_z^2},$$

$$E_y = \frac{-jk_z \frac{\partial E_z}{\partial y}}{k^2 - k_z^2}.$$

The above relations between the transverse components of TE and TM mode fields imply that

TE case:

$$\frac{E_x}{H_y} = \frac{E_y}{-H_x} = \frac{\omega\mu_o}{k_z}$$

$$= \frac{\omega\mu_o/k}{\sqrt{1 - \frac{f_c^2}{f^2}}} = \frac{\eta_o}{\sqrt{1 - \frac{f_c^2}{f^2}}} \equiv \eta_{TE}.$$

TM case:

$$\frac{E_x}{H_y} = \frac{E_y}{-H_x} = \frac{k_z}{\omega\epsilon_o}$$

$$= \frac{k\sqrt{1 - \frac{f_c^2}{f^2}}}{\omega\epsilon_o} = \eta_o\sqrt{1 - \frac{f_c^2}{f^2}} \equiv \eta_{TM}.$$

The guide impedances defined above can be used to set up transmission line models for waveguide circuits in which the parameters η_{TE} and η_{TM} for each mode play the same role as the characteristic impedance Z_o in TL theory.

- For example, two waveguides in cascade with different values of η_{TE} can be quarter-wave matched by inserting a quarter-wave section having a guide impedance equal to the geometric means of the two guides.

- For dielectric-field guides replace η_o by the appropriate η, and also in calculating the length of the quarter-wave section use $\lambda_g = \frac{2\pi}{k_z}$ appropriate for that section (see HW).

Note that, using the cutoff wavelength, we have

TE case:

$$\eta_{TE} = \frac{\eta_o}{\sqrt{1-\frac{f_c^2}{f^2}}} = \frac{\eta_o}{\sqrt{1-\frac{\lambda^2}{\lambda_c^2}}}$$

TM case:

$$\eta_{TM} = \eta_o\sqrt{1-\frac{f_c^2}{f^2}} = \eta_o\sqrt{1-\frac{\lambda^2}{\lambda_c^2}}.$$

Example 2: Consider an air-filled rectangular waveguide with $a = 3$ cm and $b = 1$ cm. Determine the TE_{10} mode fields for the guide from the results of Example 1 of Lect 29 assuming that at the operation frequency the free-space wavelength is $\lambda = 3$ cm.

Solution: By setting $k_y = 0$, $k_x = \frac{m\pi}{a} = \frac{2\pi}{\lambda_c}$, and $k_z = k\sqrt{1-(\frac{\lambda}{\lambda_c})^2} = \frac{2\pi}{\lambda_z}$ in the results of Example 1 (in Lect 29) we find for TE_{m0} mode

$$\begin{aligned}\tilde{\mathbf{H}}(x,y,z) &= H_o[\hat{x}\frac{jk_z}{k_x}\sin(k_x x) + \hat{z}\cos(k_x x)]e^{-jk_z z} \\ &= H_o[\hat{x}\frac{j\lambda_c\sqrt{1-(\frac{\lambda}{\lambda_c})^2}}{\lambda}\sin(\frac{2\pi}{\lambda_c}x) + \hat{z}\cos(\frac{2\pi}{\lambda_c}x)]e^{-jk\sqrt{1-(\frac{\lambda}{\lambda_c})^2}z}\end{aligned}$$

and

$$\begin{aligned}\tilde{\mathbf{E}}(x,y,z) &= -H_o\hat{y}\frac{j\omega\mu_o}{k_x}\sin(k_x x)e^{-jk_z z} \\ &= -H_o\hat{y}\eta_{TE}\frac{j\lambda_c\sqrt{1-(\frac{\lambda}{\lambda_c})^2}}{\lambda}\sin(\frac{2\pi}{\lambda_c}x)e^{-jk\sqrt{1-(\frac{\lambda}{\lambda_c})^2}z} \\ &= -H_o\hat{y}\eta_o\frac{j\lambda_c}{\lambda}\sin(\frac{2\pi}{\lambda_c}x)e^{-jk\sqrt{1-(\frac{\lambda}{\lambda_c})^2}z}.\end{aligned}$$

With $a = 3$ cm and $b = 1$ cm, the cutoff wavelength for TE_{10} mode is

$$\lambda_c = \frac{2a}{m} = 6\,\text{cm}.$$

Thus, with $\lambda = 3$ cm

$$\sqrt{1-(\frac{\lambda}{\lambda_c})^2} = \sqrt{1-\frac{1}{4}} = \frac{\sqrt{3}}{2}\ ,\frac{\lambda_c}{\lambda} = 2,\ \ \frac{\lambda_z}{\lambda} = \frac{2}{\sqrt{3}}.$$

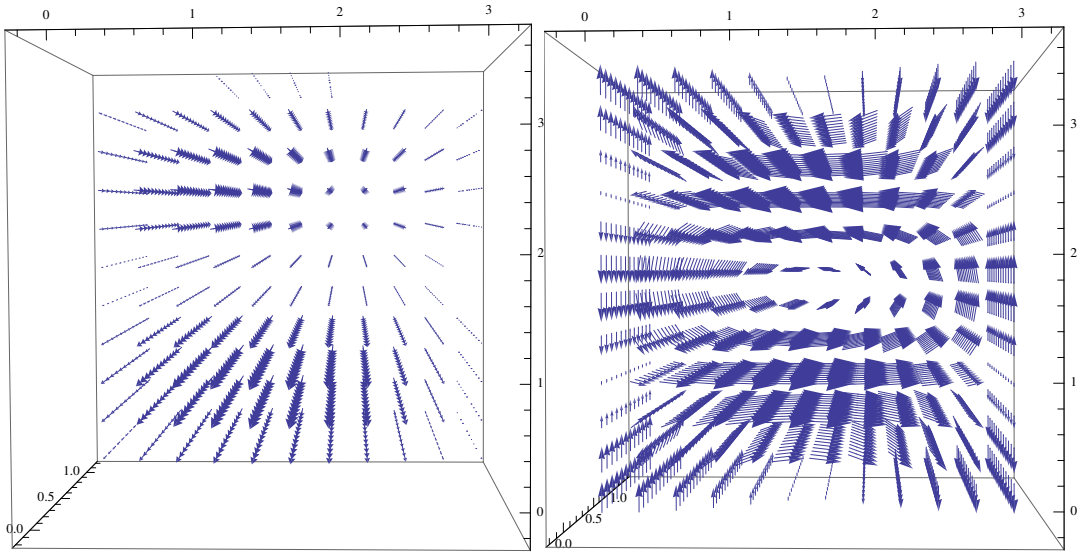
Then, for TE_{10} mode we have

$$\tilde{\mathbf{H}}(x,y,z) = H_o[\hat{x}j\sqrt{3}\sin(\frac{\pi}{3}x) + \hat{z}\cos(\frac{\pi}{3}x)]e^{-j\pi z/\sqrt{3}}$$

and

$$\tilde{\mathbf{E}}(x,y,z) = -H_o\hat{y}\eta_o j2\sin(\frac{\pi}{3}x)e^{-j\pi z/\sqrt{3}}.$$

The real part of these phasors would yield the field vectors inside the waveguide at time $t = 0$, as depicted below.



- In the 3D plots shown above we depict $\mathbf{E}(x,y,z,0)$ vectors from Example 2 on the left, and $\mathbf{H}(x,y,z,0)$ on the right; the horizontal axis is x, vertical is z, and y axis is into the page (all labelled in cm units) —note that
  - there is no field variation in y-direction because this is the TE_{10} mode,
  - $\mathbf{E} \times \mathbf{H}$ is predominantly in $\hat{z}$ direction.

Example 3: Repeat Example 2 for the case of TE_{20} mode and $\lambda = 2$ cm.

Solution: For the TE_{m0} mode we have

$$\tilde{\mathbf{H}}(x,y,z) = H_o[\hat{x}\frac{j\lambda_c\sqrt{1-(\frac{\lambda}{\lambda_c})^2}}{\lambda}\sin(\frac{2\pi}{\lambda_c}x) + \hat{z}\cos(\frac{2\pi}{\lambda_c}x)]e^{-jk\sqrt{1-(\frac{\lambda}{\lambda_c})^2}z}$$

and

$$\tilde{\mathbf{E}}(x,y,z) = -H_o\hat{y}\eta_o\frac{j\lambda_c}{\lambda}\sin(\frac{2\pi}{\lambda_c}x)e^{-jk\sqrt{1-(\frac{\lambda}{\lambda_c})^2}z}.$$

With $a = 3$ cm and $b = 1$ cm, the cutoff wavelength for TE_{20} mode is

$$\lambda_c = \frac{2a}{m} = 3\,\text{cm}.$$

Thus, with $\lambda = 2$ cm

$$\sqrt{1-(\frac{\lambda}{\lambda_c})^2} = \sqrt{1-\frac{4}{9}} = \frac{\sqrt{5}}{3}\ , \frac{\lambda_c}{\lambda} = 1.5,\ \ \frac{\lambda_z}{\lambda} = \frac{3}{\sqrt{5}}.$$

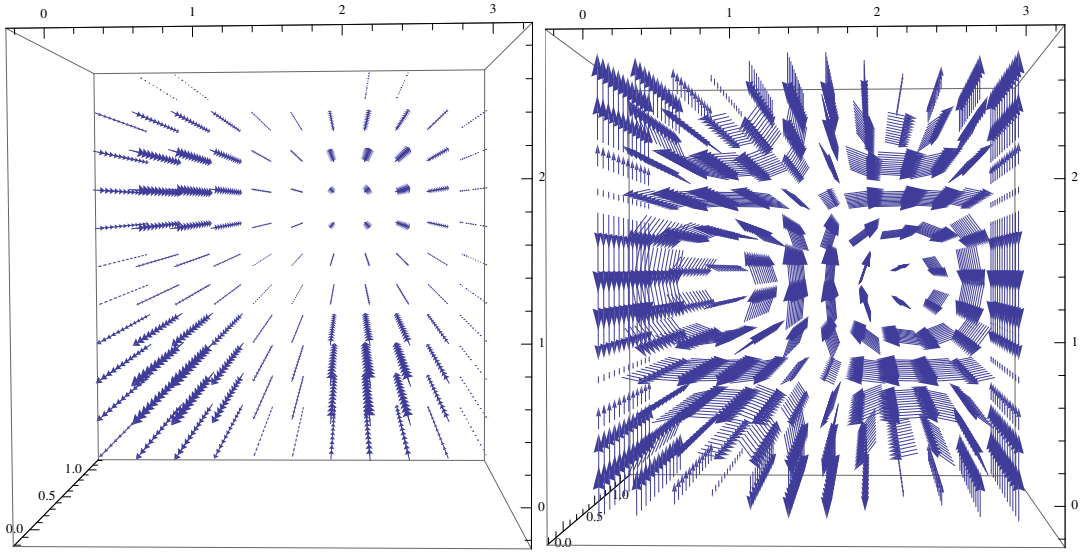
Then, for TE_{10} mode we have

$$\tilde{\mathbf{H}}(x,y,z) = H_o[\hat{x}j\frac{\sqrt{5}}{2}\sin(\frac{2\pi}{3}x) + \hat{z}\cos(\frac{2\pi}{3}x)]e^{-j\pi z\sqrt{5}/3}$$

and

$$\tilde{\mathbf{E}}(x,y,z) = -H_o\hat{y}\eta_o j\frac{3}{2}\sin(\frac{2\pi}{3}x)e^{-j\pi z\sqrt{5}/3}.$$

The real part of these phasors would yield the field vectors inside the waveguide at time $t = 0$, as depicted below.



- In the 3D plots shown above we depict $\mathbf{E}(x, y, z, 0)$ vectors from Example 3 on the left, and $\mathbf{H}(x, y, z, 0)$ on the right; the horizontal axis is x, vertical is z, and y axis is into the page (all labelled in cm units).

- Imagine the vector patterns depicted above sliding upwards in the z-axis direction at the speed $v_{pz} = \frac{\omega}{k_z}$, with each feature of the pattern passing by a stationary observer who experiences a monochromatic oscillation.

  - that would be the proper way of visualizing the propagation of an unmodulated TE_{20} mode.

Example 4: For the TE_{m0} mode we have the wave fields

$$\tilde{\mathbf{H}}(x,y,z) = H_o[\hat{x}\frac{j\lambda_c\sqrt{1-(\frac{\lambda}{\lambda_c})^2}}{\lambda}\sin(\frac{2\pi}{\lambda_c}x)+\hat{z}\cos(\frac{2\pi}{\lambda_c}x)]e^{-jk\sqrt{1-(\frac{\lambda}{\lambda_c})^2}z}$$

and

$$\tilde{\mathbf{E}}(x,y,z) = -H_o\hat{y}\eta_o\frac{j\lambda_c}{\lambda}\sin(\frac{2\pi}{\lambda_c}x)e^{-jk\sqrt{1-(\frac{\lambda}{\lambda_c})^2}z}.$$

Express the time-averaged power transmitted by the mode in $\hat{z}$ direction in terms of

$$E_o \equiv H_o\eta_o\frac{\lambda_c}{\lambda}$$

representing the amplitude of the electric field wave.

Solution: We start with the time-averaged Poynting vector

$$\begin{aligned}\langle\mathbf{E}\times\mathbf{H}\rangle &= \frac{1}{2}\mathrm{Re}\{\tilde{\mathbf{E}}\times\tilde{\mathbf{H}}^*\}\\ &= \frac{|H_o|^2\eta_o}{2}(\frac{\lambda_c}{\lambda})^2\sqrt{1-(\frac{\lambda}{\lambda_c})^2}\sin^2(\frac{2\pi}{\lambda_c}x)\hat{z}\\ &= \frac{|E_o|^2}{2\eta_o}\sqrt{1-(\frac{\lambda}{\lambda_c})^2}\sin^2(\frac{2\pi}{\lambda_c}x)\hat{z} = \frac{|E_o|^2}{2\eta_{TE}}\sin^2(\frac{2\pi}{\lambda_c}x)\hat{z}.\end{aligned}$$

Now, integrating $\langle\mathbf{E}\times\mathbf{H}\rangle\cdot\hat{z}$ across the guide cross section we get the time-average power

$$\begin{aligned}P &= \int_{x=0}^{a}\int_{y=0}^{b}\langle\mathbf{E}\times\mathbf{H}\rangle\cdot\hat{z}\,dx\,dy\\ &= \frac{|E_o|^2}{2\eta_{TE}}b\int_0^a\sin^2(\frac{2\pi}{\lambda_c}x)dx = \frac{|E_o|^2}{2\eta_{TE}}\frac{ab}{2}\end{aligned}$$

since the integral of

$$\sin^2(\frac{2\pi}{\lambda_c}x) = \frac{1}{2}(1-\cos(\frac{4\pi}{2a/m}x)) = \frac{1}{2}(1-\cos(2\pi mx/a))$$

yields 1/2. It can be shown that in the case of TE_{mn} modes with non-zero n, the above result for P is still valid provided $ab/2$ is replaced by $ab/4$ (see HW).

Example 5: A rectangular waveguide with $a = 2$ cm and $b = 1$ cm is air filled for $z < 0$, but is is filled with a dielectric in $z > 0$ region with a refractive index $n = 1.5$ and $\mu_r = 1$. For $f = 12.5$ GHz and TE_{10} mode operation design a $\lambda/4$ transformer to match the two sections of the waveguide. Use transmission-line analogy to solve this problem (as in Lecture 24).

Solution: To solve this problem using a transmission-line analogy we first need the impedances η_{TE} for the two sections of the guide. Since

$$\eta_{TE} = \frac{\eta}{\sqrt{1 - \frac{f_c^2}{f^2}}}$$

we need to find f_c and η in the two sections of the guide.
The cutoff frequency is

$$f_c = \frac{mc}{2a} = \frac{3 \times 10^{10}\,\text{cm/s}}{2 \times 2\,\text{cm}} = 7.5\,\text{GHz} \text{ in air,}$$

and

$$f_c = \frac{mc/n}{2a} = \frac{7.5\,\text{GHz}}{n} = \frac{7.5\,\text{GHz}}{1.5} = 5\,\text{GHz} \text{ in dielectric.}$$

Hence

$$\eta_{TE} = \frac{\eta}{\sqrt{1 - \frac{f_c^2}{f^2}}} = \frac{120\pi}{\sqrt{1 - (\frac{7.5}{12.5})^2}} = 150\pi\,\Omega \text{ in air,}$$

and

$$\eta_{TE} = \frac{\eta}{\sqrt{1 - \frac{f_c^2}{f^2}}} = \frac{120\pi/1.5}{\sqrt{1 - (\frac{5}{12.5})^2}} = \frac{400\pi}{\sqrt{21}}\,\Omega \text{ in dielectric.}$$

Since $\eta_{TE,air} \neq \eta_{TE,diel}$, we will certainly have reflections at the interface at $z = 0$ unless a matching section is inserted.

Consider a $\lambda/4$ long section of a waveguide with identical dimensions as above but filled with some dielectric having a refractive index n_x. Then, transmission-line analogy would indicate that an impedance match can be achieved if

$$\eta_{TE,air}\eta_{TE,diel} = \eta_{TE,x}^2$$

where $\eta_{TE,x}$is the impedance of the matching segment. In view of the above relations, this can be written as

$$(150\pi)(\frac{400\pi}{\sqrt{21}}) = \left[\frac{120\pi/n_x}{\sqrt{1-\left(\frac{7.5/n_x}{12.5}\right)^2}}\right]^2,$$

which yields

$$n_x^2 - \left(\frac{7.5}{12.5}\right)^2 = \frac{120^2\sqrt{21}}{150\times 400} \;\Rightarrow\; n_x^2 = 1.459.$$

To determine the actual length of the $\lambda/4$ long section we need to find out λ, which is really the guide wavelength λ_g for the TE_{10} mode, i.e.,

$$\begin{aligned}\lambda_g &= \frac{2\pi}{k_z} = \frac{2\pi/k}{\sqrt{1-\frac{f_c^2}{f^2}}} = \frac{c/n_x}{12.5\times 10^9\sqrt{1-\left(\frac{7.5/n_x}{12.5}\right)^2}} = \frac{30}{12.5n_x\sqrt{1-\left(\frac{7.5/n_x}{12.5}\right)^2}} \\ &= \frac{30}{\sqrt{(12.5n_x)^2-7.5^2}} = 2.28\,\mathrm{cm}.\end{aligned}$$

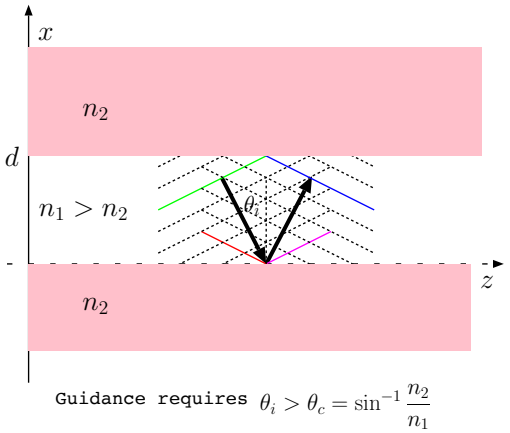
Thus, the matching section has a physical length of

$$d = \frac{\lambda_g}{4} = 0.572\,\mathrm{cm}.$$

34 TE modes in dielectric slab waveguides

- As frequency f increases well beyond the microwave range, the cutoff wavelength $\lambda_c = \frac{2a}{1} = \frac{c}{f}$ of the TE_{10} of mode will dip towards μm scales. Guiding structures with μm scales can be more naturally implemented as dielectric slabs as opposed to hollow waveguides. Optical integrated circuits contain many such channels of **dielectric slab waveguides**.

  - In this lecture we will examine briefly the guidance conditions and dispersion characteristics encountered in dielectric slab waveguides.



- Consider a slab of dielectric material of refractive index $n_1 = \sqrt{\epsilon_{1r}}$ of a width d embedded in a dielectric with a smaller refractive index $n_2 = \sqrt{\epsilon_{2r}}$. Propagating modes of frequency f can be trapped and guided in the slab with the refractive index $n_1 > n_2$,

  - so long as the mode can be represented as a superposition of unguided TEM waves reflected from plane boundaries of regions with index n_1 and n_2 with an incidence angle θ_i larger that θ_c, where

$$\theta_c = \sin^{-1}\frac{n_2}{n_1}$$

    is the **critical angle** for **total internal reflection** (TIR).

  - Recall that when TIR occurs, the reflected wave has the same amplitude as the incident wave, while an evanescent transmitted

wave is found in the second region. If $\theta_i < \theta_c$ no guidance can occur since the transmitted fields in that case would be propagating rather than evanescent.

- Guided modes not only require

$$\theta_i > \sin^{-1}\frac{n_2}{n_1},$$

but also

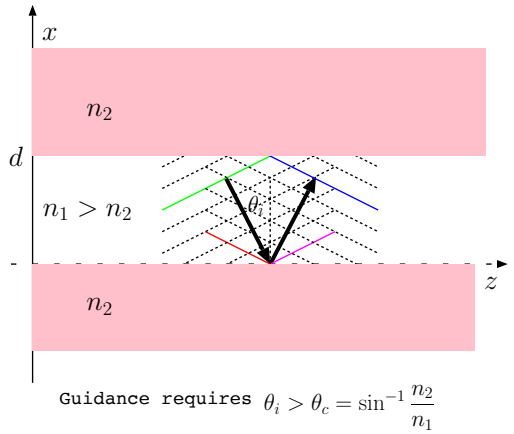
$$k_1 d\cos\theta_i = \angle\Gamma + m\pi, \quad m = 0, 1, 2, \cdots$$

where Γ denotes the reflection coefficient at the interfaces between the regions of n_1 and n_2. This guidance condition ensures the self-consistency of free TEM components of the guided modes reflected from the planar interfaces separated by distance d.

$$\begin{aligned}\phi_b &= \phi_r - k_{1x}d,\\ \phi_g &= \phi_b + \angle\Gamma,\\ \phi_m &= \phi_g - k_{1x}d,\\ \phi_r &= \phi_m + \angle\Gamma\end{aligned}$$

where

$$k_{1x} = k_1\cos\theta_i.$$



$$k_1 d\cos\theta_i = \angle\Gamma + m\pi, \quad m = 0, 1, 2, \cdots$$

- For the TE mode case where the incident and reflected fields taken as

$$\tilde{\mathbf{E}}_i = \hat{y}E_o e^{-jk_1(-\cos\theta_i x + \sin\theta_i z)} \text{ and } \tilde{\mathbf{E}}_r = \hat{y}E_o\Gamma_{TE}e^{-jk_1(\cos\theta_i x + \sin\theta_i z)}$$

the reflection coefficient is given as

$$\Gamma_{TE} = \frac{\eta_2\cos\theta_i - \eta_1\cos\theta_2}{\eta_2\cos\theta_i + \eta_1\cos\theta_2} = \frac{n_1\cos\theta_i - n_2\cos\theta_2}{n_1\cos\theta_i + n_2\cos\theta_2}$$

since

$$\eta = \sqrt{\frac{\mu_o}{\epsilon_r\epsilon_o}} = \frac{\eta_o}{\sqrt{\epsilon_r}} = \frac{\eta_o}{n}.$$

- Above,

$$\cos\theta_2 = \sqrt{1-\sin^2\theta_2} = \sqrt{1-\frac{n_1^2}{n_2^2}\sin^2\theta_i}$$

since according to Snell's law

$$k_1\sin\theta_i = k_2\sin\theta_2 \;\text{ and }\; \sin\theta_2 = \frac{k_1}{k_2}\sin\theta_1 = \frac{\omega\sqrt{\mu_o\epsilon_1}}{\omega\sqrt{\mu_o\epsilon_2}}\sin\theta_1 = \frac{n_1}{n_2}\sin\theta_1.$$

Clearly, for

$$1 < \frac{n_1^2}{n_2^2}\sin^2\theta_i \;\Leftrightarrow\; \theta_i > \theta_c = \sin^{-1}\frac{n_2}{n_1}$$

we have

$$\cos\theta_2 = \pm j\sqrt{\frac{n_1^2}{n_2^2}\sin^2\theta_i - 1}$$

and (using the root that causes the decay of the fields $\propto e^{\mp jk_2\cos\theta_2 x}$ above and below the slab)

$$\Gamma_{TE} = \frac{n_1\cos\theta_i + j\sqrt{n_1^2\sin^2\theta_i - n_2^2}}{n_1\cos\theta_i - j\sqrt{n_1^2\sin^2\theta_i - n_2^2}}$$

with

$$\angle\Gamma_{TE} = 2\tan^{-1}\frac{\sqrt{\sin^2\theta_i - n_2^2/n_1^2}}{\cos\theta_i}.$$

Now, substituting $\angle\Gamma_{TE}$ in the guidance condition shown in the margin, we obtain

$$\frac{k_1 d}{2}\cos\theta_i - m\frac{\pi}{2} = \tan^{-1}\frac{\sqrt{\sin^2\theta_i - n_2^2/n_1^2}}{\cos\theta_i},\;\; m = 0,1,2,3,\cdots$$

Guidance conditions:

$$\theta_i > \sin^{-1}\frac{n_2}{n_1}$$

and

$$k_1 d\cos\theta_i = \angle\Gamma + m\pi,\;\; m = 0,1,2,\cdots$$

Another way to obtain the guidance conditions:

Since

$$\tilde{\mathbf{E}}_i = \hat{y}E_o e^{-jk_1(-\cos\theta_i x + \sin\theta_i z)}$$

and

$$\tilde{\mathbf{E}}_r = \hat{y}E_o\Gamma_{TE}e^{-jk_1(\cos\theta_i x + \sin\theta_i z)},$$

and $\tilde{\mathbf{E}}_r$ gets reflected at $x = d$ (once again) to become $\tilde{\mathbf{E}}_i$, it is then necessary that

$$(\Gamma_{TE}e^{-jk_1\cos\theta_i d})\Gamma_{TE} = e^{jk_1\cos\theta_i d}e^{-j2\pi m}$$

for integers m. This is possible iff $|\Gamma_{TE}| = 1$, i.e.,

$$\theta_i > \sin^{-1}\frac{n_2}{n_1},$$

and

$$k_1 d\cos\theta_i = \angle\Gamma_{TE} + m\pi,\;\; m = 0,1,2,\cdots$$

which is only valid for θ_i satisfying

$$\theta_i \geq \sin^{-1} \frac{n_2}{n_1} \equiv \theta_c.$$

Since

$$k_1 = \frac{\omega}{v_1}, \text{ where } v_1 = \frac{1}{\sqrt{\mu_o \epsilon_1}} = \frac{c}{n_1},$$

the guidance condition can also be cast as

$$\frac{d}{v_1/f} \cos\theta_i - \frac{m}{2} = \frac{1}{\pi} \tan^{-1} \frac{\sqrt{\sin^2\theta_i - n_2^2/n_1^2}}{\cos\theta_i}, \quad m = 0, 1, 2, 3, \cdots$$

which is known as the **characteristic equation** for TE modes.

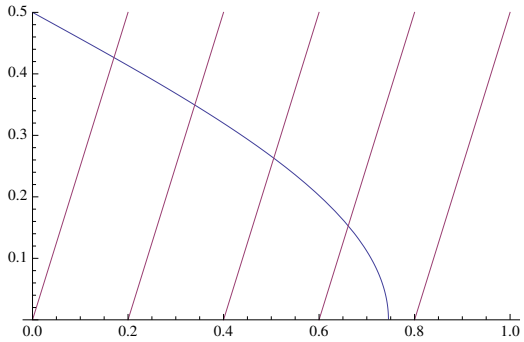
- The above equations constrain the number of propagating modes at a given frequency f and the associated angle θ_i for each TE mode m.

  - Each propagating mode for a given d is associated with a cutoff frequency f_c, and propagation is possible only if $f > f_c$ for the given mode.
  - At $f = f_c$ we have $\theta_i = \theta_c$ for the given mode, in which case

    $$\frac{d}{v_1/f_c} \cos\theta_c - \frac{m}{2} = \frac{1}{\pi} \tan^{-1} \frac{\sqrt{\sin^2\theta_c - n_2^2/n_1^2}}{\cos\theta_c} = 0,$$

    from which we obtain the cutoff frequencies

    $$f_c = \frac{m v_1}{2d \cos\theta_c}, \quad m = 0, 1, 2, 3 \cdots$$



Graphical solution of the characteristic equation for TE modes $m = 0, 1, 2, 3$ in propagation and mode $m = 4$ in evanescence. The blue curve depicts

$$\frac{1}{\pi} \tan^{-1} \frac{\sqrt{\sin^2\theta_i - n_2^2/n_1^2}}{\cos\theta_i}$$

as a function of $\cos\theta_i$ for $n_2 = 1$ and $n_1 = 1.5$ while the straight lines depict

$$\frac{d}{v_1/f} \cos\theta_i - \frac{m}{2}$$

with $\frac{d}{v_1/f} = 2.5$ and m increasing from left to right in steps of one.

Since

$$v_1 = \frac{c}{n_1} \text{ and } \cos\theta_c = \sqrt{1-\sin^2\theta_c} = \sqrt{1-\frac{n_2^2}{n_1^2}}$$

it follows that

$$f_c = \frac{mc}{2d\sqrt{n_1^2-n_2^2}}, \quad m = 0, 1, 2, 3\cdots$$

for TE modes.

Example 1: Consider a dielectric slab waveguide with $d = 3$ mm, $n_1 = 1.5$, and $n_2 = 1$. (a) Determine the cutoff frequency for the TE_1 mode in the guide. (b) Determine the frequency f of a TE_0 mode signal in the waveguide if $\theta_i = 60°$. (c) Determine the phase velocity of the mode described in part (b).

Solution: (a) The cutoff frequency for TE_1 mode is

$$f_c = \frac{mc}{2d\sqrt{n_1^2-n_2^2}} = \frac{3\times10^{10}\,\text{cm/s}}{2\times0.3\,\text{cm}\sqrt{1.5^2-1}} = \frac{5\times10^{10}}{\sqrt{1.25}}\,\text{Hz} \approx 44.72\,\text{GHz}.$$

(b) Evaluating the characteristic equation

$$\frac{d}{v_1/f}\cos\theta_i - \frac{m}{2} = \frac{1}{\pi}\tan^{-1}\frac{\sqrt{\sin^2\theta_i - n_2^2/n_1^2}}{\cos\theta_i}, \quad m = 0, 1, 2, 3, \cdots$$

with $m = 0$, $n_2/n_1 = 2/3$, $\sin\theta_i = \sqrt{3}/2$, and $\cos\theta_i = 1/2$, we find

$$\frac{fd}{v_1 2} = \frac{1}{\pi}\tan^{-1}\frac{\sqrt{3/4-(2/3)^2}}{1/2} = 0.266 \;\Rightarrow\; f = 0.266\times2\times\frac{v_1}{d}.$$

Since

$$v_1 = \frac{c}{n_1} = \frac{3 \times 10^{10}\,\mathrm{cm/s}}{3/2} = 2 \times 10^{10}\,\mathrm{cm/s},$$

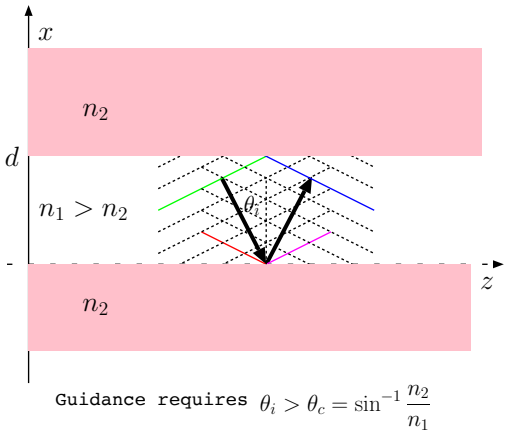
we find

$$f = 0.266 \times 2 \times \frac{v_1}{d} = 0.266 \times 2 \times \frac{2 \times 10^{10}\,\mathrm{cm/s}}{0.3\,\mathrm{cm}} = 3.54 \times 10^{10}\,\mathrm{Hz} = 35.4\,\mathrm{GHz}.$$

(c) The phase velocity of the mode is given by

$$v_{pz} = \frac{\omega}{k_z} = \frac{\omega}{k_1 \sin\theta_i} = \frac{v_1}{\sin\theta_i} = \frac{2 \times 10^{10}\,\mathrm{cm/s}}{\sqrt{3}/2} = 2.31 \times 10^{10}\,\mathrm{cm/s}.$$

35 TM modes in dielectric waveguides



- Last lecture we examined the **characteristic equation** and the **cutoff frequencies** of TE mode of propagation in dielectric slab waveguides.

  Guided TE_m mode fields consisting of the superposition of transverse polarized electric fields

  $$\tilde{\mathbf{E}}_i = \hat{y}E_o e^{-jk_1(-\cos\theta_i x + \sin\theta_i z)} \text{ and } \tilde{\mathbf{E}}_r = \hat{y}E_o\Gamma_{TE}e^{-jk_1(\cos\theta_i x + \sin\theta_i z)},$$

  where

  $$\Gamma_{TE} = \frac{n_1\cos\theta_i + j\sqrt{n_1^2\sin^2\theta_i - n_2^2}}{n_1\cos\theta_i - j\sqrt{n_1^2\sin^2\theta_i - n_2^2}},$$

  have

  1. propagation angles

     $$\theta_i > \theta_c = \sin^{-1}\frac{n_2}{n_1}, \text{ (critical angle)}$$

  2. satisfying a characteristic equation

     $$\frac{d}{v_1/f}\cos\theta_i - \frac{m}{2} = \frac{1}{\pi}\tan^{-1}\frac{\sqrt{\sin^2\theta_i - n_2^2/n_1^2}}{\cos\theta_i}, \quad m = 0, 1, 2, 3, \cdots$$

  3. for frequencies f exceeding the cutoff frequency

     $$f_c = \frac{mc}{2d\sqrt{n_1^2 - n_2^2}}, \quad m = 0, 1, 2, 3\cdots$$

- For a given $f = \frac{2\pi}{\omega}$, the characteristic equation can be solved (typically by using graphical techniques) for θ_i, from which we can calculate the propagation constant

$$k_z = k_1 \sin\theta_i$$

where

$$k_1 = \frac{\omega}{v_1} = \frac{\omega}{c} n_1$$

is the wavenumber in the core region of the guide at the operation frequency $\omega = 2\pi f$. It follows that the

$$\text{guide wavelength } \lambda_g = \frac{2\pi}{k_z}$$

and [1]

$$\text{phase velocity } v_{pz} = \frac{\omega}{k_z} = \frac{\omega}{k_1 \sin\theta_i} = \frac{v_1}{\sin\theta_i}$$

can be obtained once θ_i is calculated from the characteristic equation.

- Given the k_z in the core region, k_x and k_z outside the core region (with index n_2) can be obtained by using the fact that k_z is identical in both regions (why?).

- In this lecture we will continue our study of dielectric slab waveguides by examining the TM modes.

[1] Note that group velocity $v_g = v_1 \sin\theta_i$ (in analogy with parallel plate waveguides) if and only if $\omega \gg \omega_c = 2\pi f_c$ because of the effect of the cladding region that contains a substatial fraction on the wave energy unless $\omega \gg \omega_c$ — in fact for TE_0 mode $v_g \approx v_2$ at frequencies much less than the cutoff frequency of TE_1 mode.

TM modes

- For the TM mode where the incident and reflected fields are taken as

$$\tilde{\mathbf{H}}_i = \hat{y} H_o e^{-jk_1(-\cos\theta_i x + \sin\theta_i z)} \text{ and } \tilde{\mathbf{H}}_r = \hat{y} H_o R e^{-jk_1(\cos\theta_i x + \sin\theta_i z)}$$

the reflection coefficient is given as

$$R = \frac{\eta_1 \cos\theta_i - \eta_2 \cos\theta_2}{\eta_1 \cos\theta_i + \eta_2 \cos\theta_2} = \frac{n_2 \cos\theta_i - n_1 \cos\theta_2}{n_2 \cos\theta_i + n_1 \cos\theta_2}.$$

This leads to

$$R = \frac{n_2 \cos\theta_i + j n_1 \sqrt{\frac{n_1^2}{n_2^2}\sin^2\theta_i - 1}}{n_2 \cos\theta_i - j n_1 \sqrt{\frac{n_1^2}{n_2^2}\sin^2\theta_i - 1}}$$

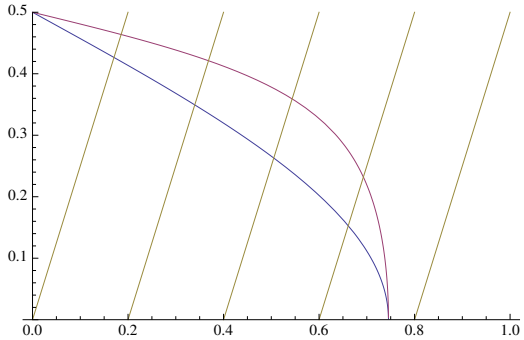
with

$$\angle R = 2\tan^{-1} \frac{\sqrt{\sin^2\theta_i - n_2^2/n_1^2}}{\frac{n_2^2}{n_1^2}\cos\theta_i}.$$

Hence, in this case the guidance condition leads to the characteristic equation

$$\frac{d}{v_1/f}\cos\theta_i - \frac{m}{2} = \frac{1}{\pi}\tan^{-1}\frac{\sqrt{\sin^2\theta_i - n_2^2/n_1^2}}{\frac{n_2^2}{n_1^2}\cos\theta_i}, \quad m = 0, 1, 2, 3, \cdots$$

Note that this result for the TM mode leads to the same f_c expression as in TE modes.



Graphical solution of the characteristic equations for TE and TM modes $m = 0, 1, 2, 3$ in propagation and mode $m = 4$ in evanescence. The blue and red curves depict

$$\frac{1}{\pi}\tan^{-1}\frac{\sqrt{\sin^2\theta_i - n_2^2/n_1^2}}{\cos\theta_i}$$

and

$$\frac{1}{\pi}\tan^{-1}\frac{\sqrt{\sin^2\theta_i - n_2^2/n_1^2}}{\frac{n_2^2}{n_1^2}\cos\theta_i}$$

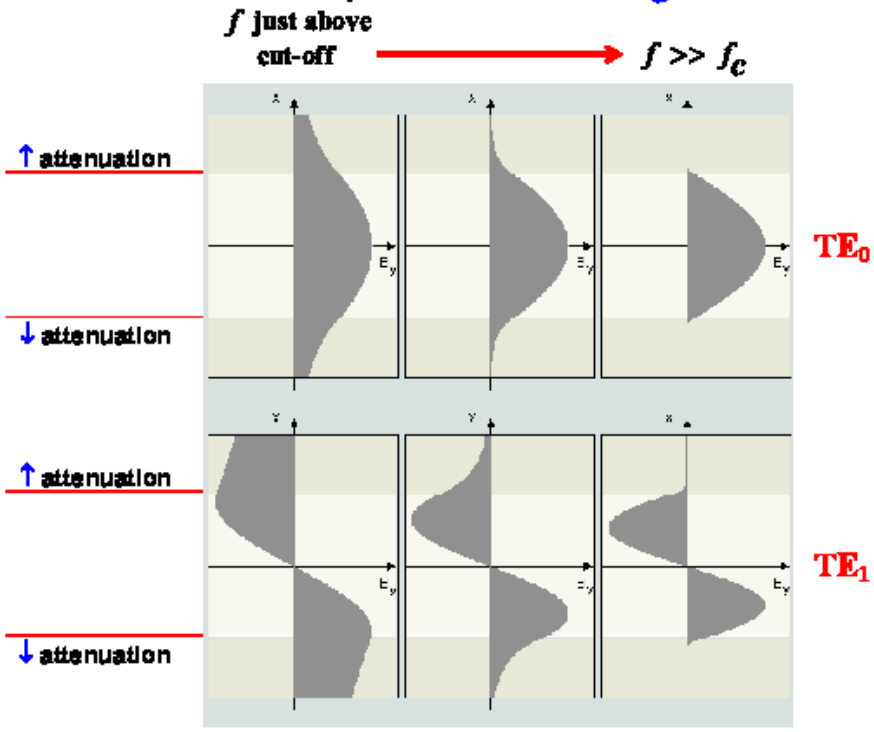
as a function of $\cos\theta_i$ for $n_2 = 1$ and $n_1 = 1.5$, while the straight lines depict

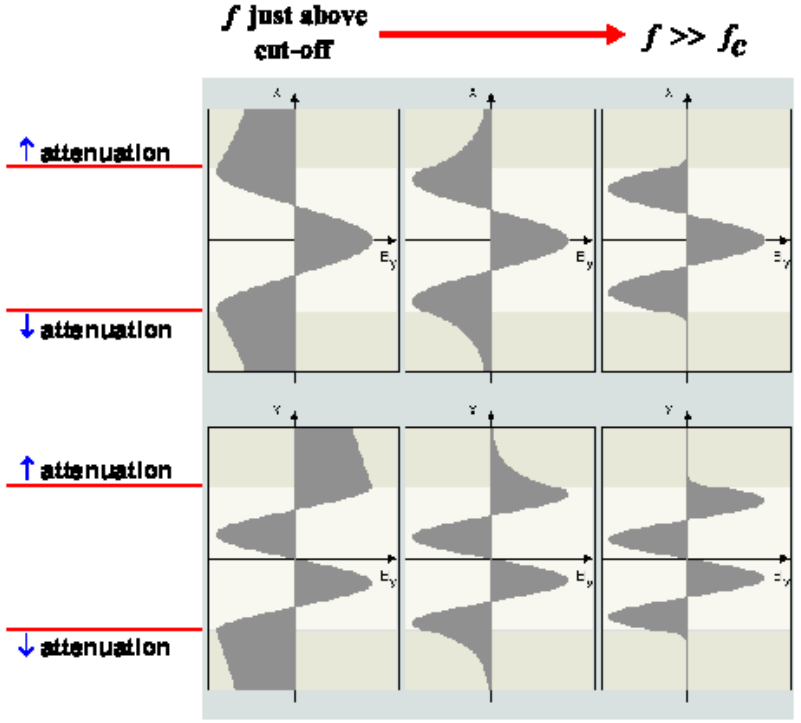
$$\frac{d}{v_1/f}\cos\theta_i - \frac{m}{2}$$

with $\frac{d}{v_1/f} = 2.5$ and m increasing from left to right in steps of one.

Mode structures:

Examples of profiles for the transverse electric field of TE modes.
TM modes have similar profiles for the magnetic field.





Acceptance cone and numerical aperture:

Guidance requires

$$\theta_i > \theta_c = \sin^{-1}\frac{n_2}{n_1} = \sin^{-1}\sqrt{\frac{\epsilon_{r2}}{\epsilon_{r1}}}$$

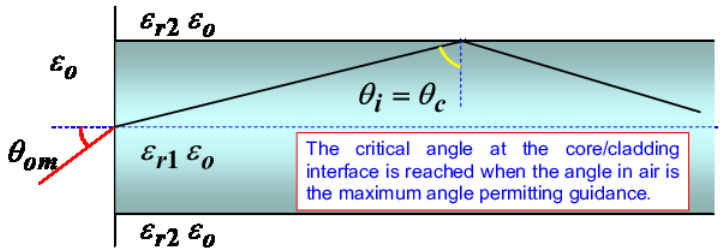
and therefore "acceptance angles" from air (see the diagram below)

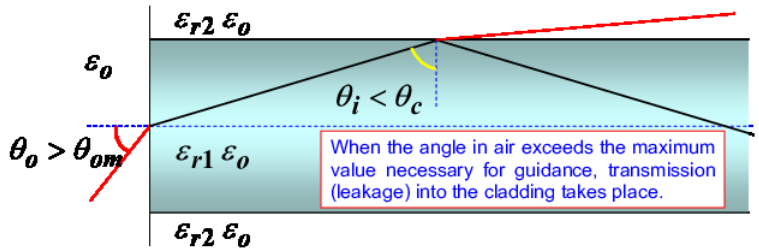
$$\theta_o < \theta_{om} = \sin^{-1}\sqrt{n_1^2 - n_2^2} = \sin^{-1}\sqrt{\epsilon_{r1} - \epsilon_{r2}}$$

where

$\sin\theta_{om} = \sqrt{n_1^2 - n_2^2} = \sqrt{\epsilon_{r1} - \epsilon_{r2}}$ is called **numerical aperture**.

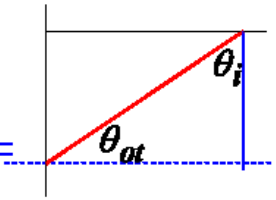
The maximum acceptance angle θ_{om} defines the so-called "acceptance cone" that includes all the external signals incident on the dielectric waveguide that can couple to the waveguide at the air/core interface on a constant-z plane.





At the air–core interface

$$\sin\theta_{ot} = \sqrt{\frac{\varepsilon_{rair}\varepsilon_o}{\varepsilon_{r1}\varepsilon_o}}\sin\theta_o = \sqrt{\frac{1}{\varepsilon_{r1}}}\sin\theta_o$$



$$\theta_i + \theta_{ot} = 90^\circ \quad \Rightarrow \quad \cos\theta_{ot} = \sin\theta_i \quad \Leftarrow$$

At the critical angle

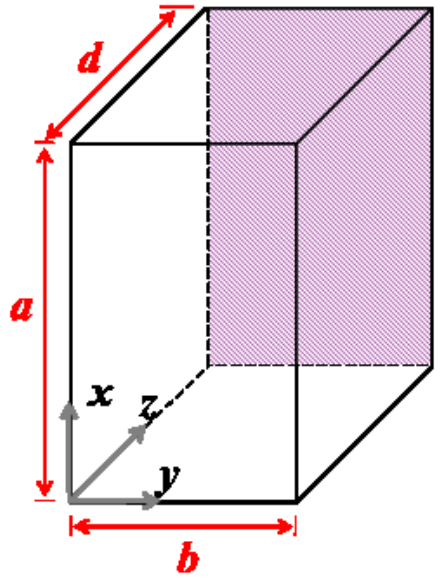
$$\sin\theta_i = \sin\theta_c = \sqrt{\frac{\varepsilon_{r2}}{\varepsilon_{r1}}}$$

$$\sin^2\theta_{otm} = 1 - \cos^2\theta_{otm} = 1 - \sin^2\theta_c = 1 - \frac{\varepsilon_{r2}}{\varepsilon_{r1}} = \frac{\sin^2\theta_{om}}{\varepsilon_{r1}}$$

$$\Rightarrow \quad \sin\theta_{om} = \sqrt{\varepsilon_{r1} - \varepsilon_{r2}} = \text{numerical aperture}$$

$$\theta_{om} = \sin^{-1}\sqrt{\varepsilon_{r1} - \varepsilon_{r2}}$$

36 Rectangular cavities



- Consider a rectangular waveguide propagating some TE_{mn} mode having a longitudinal magnetic field component

$$H_z^+ \propto \cos(k_x x)\cos(k_y y)e^{-jk_z z},$$

where

$$k_x = \frac{m\pi}{a}, \quad k_y = \frac{n\pi}{b}, \quad \text{and} \quad k_z = \sqrt{k^2 - k_x^2 - k_y^2}.$$

In principle, the same guide can also propagate a TE_{mn} mode field with

$$H_z^- \propto \cos(k_x x)\cos(k_y y)e^{+jk_z z}$$

in reverse direction, and if both waves were present in the guide, we would have a total field

$$H_{zt} = \cos(k_x x)\cos(k_y y)(Fe^{-jk_z z} + Re^{+jk_z z})$$

where F and R denote the amplitudes of the forward and reverse waves depending on sources and/or boundaries in z.

  - Of course $E_{zt} = 0$ for TE_{mn} modes, while
  - transverse field components can be obtained using the equations shown in the margin (derived in Lecture 29) where the sign of $\mp jk_z$ is taken in accordance with the order implied in $e^{\mp jk_z z}$ for forward and reverse propagating waves.

TE mode fields:

$$H_x^\pm = \frac{\mp jk_z \frac{\partial H_z}{\partial x}}{k^2 - k_z^2},$$

$$H_y^\pm = \frac{\mp jk_z \frac{\partial H_z}{\partial y}}{k^2 - k_z^2},$$

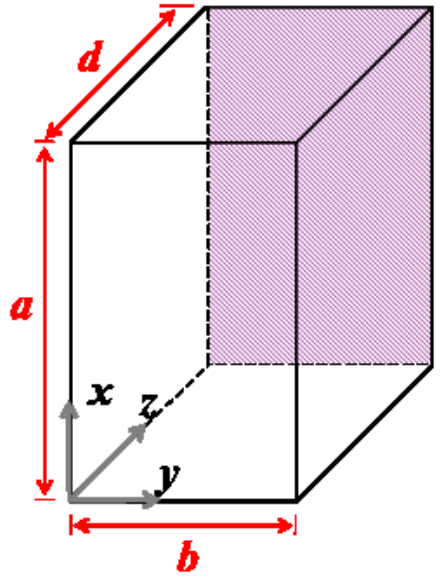
$$E_y^\pm = \frac{j\omega\mu_o \frac{\partial H_z}{\partial x}}{k^2 - k_z^2},$$

$$E_x^\pm = \frac{-j\omega\mu_o \frac{\partial H_z}{\partial y}}{k^2 - k_z^2}.$$

- For a **rectangular cavity** formed by introducing conducting walls at $z = 0$ and $z = d$ within a rectangular waveguide, the pertinent boundary conditions to be imposed on H_{zt} become

$$H_{zt}(x, y, 0) = 0 \text{ and } H_{zt}(x, y, d) = 0$$

since **H** cannot be perpendicular to a conducting plate. Accordingly,



1. $H_{zt}(x, y, 0) = 0$ requires $R = -F$, in which case we can write (taking F=1 for simplicity)

$$\begin{aligned} H_{zt} &= \cos(k_x x)\cos(k_y y)(e^{-jk_z z} - e^{+jk_z z}) \\ &= -j2\cos(k_x x)\cos(k_y y)\sin(k_z z). \end{aligned}$$

2. $H_{zt}(x, y, d) = 0$, in turn, requires

$$k_z d = l\pi, \quad l = 1, 2, 3 \cdots$$

excluding $l = 0$ for non-zero H_{zt}.

- $H_{zt}(x, y, z)$ now describes a standing wave pattern within the rectangular cavity, having a periodicity in z where

$$|H_{zt}(x, y, z)| \propto |\sin(k_z z)|$$

repeats over z by integer multiples of $\lambda_z/2$, where

$$\lambda_z = \frac{2\pi}{k_z} \quad \text{and} \quad k_z = \frac{l\pi}{d}, \quad l = 1, 2, 3 \cdots$$

- These standing waves are termed TE_{mnl} modes and oscillate with characteristic frequencies

$$f = \frac{\omega}{2\pi} = \sqrt{(\frac{mc}{2a})^2 + (\frac{nc}{2b})^2 + (\frac{lc}{2d})^2} = \frac{c}{2}\sqrt{(\frac{m}{a})^2 + (\frac{n}{b})^2 + (\frac{l}{d})^2} \equiv f_{mnl}$$

that follow from

$$k_x = \frac{m\pi}{a}, \;\; k_y = \frac{n\pi}{b}, \;\; k_z = \frac{l\pi}{d}$$

implying

$$k^2 = \frac{\omega^2}{c^2} = k_x^2 + k_y^2 + k_z^2 = (\frac{m\pi}{a})^2 + (\frac{n\pi}{b})^2 + (\frac{l\pi}{d})^2.$$

Characteristic frequencies f_{nml} are also known as **resonance frequencies** of the cavity, since they represent a discrete set of frequencies for which source-free field variations are possible within the cavity, in analogy with

1. having source-free voltage variations in an ideal LC circuit at its resonance frequency $\omega = \frac{1}{\sqrt{LC}}$, and also in analogy with
2. TL resonators studied in ECE 329 (short or open circuited TL segments).

- Transverse field components of TE_{mnl} resonances can be obtained by superposing the transverse derivatives of

$$H_z^{\pm} = \pm \cos(k_x x)\cos(k_y y)e^{\mp jk_z z}$$

as specified in the margin. We that find

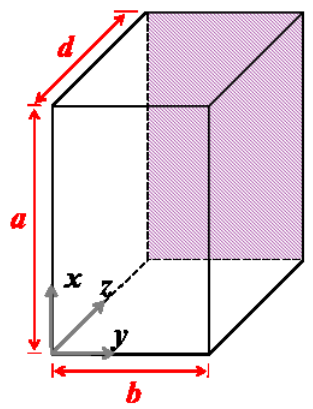
$$\begin{aligned}
H_x^{\pm} &= \frac{jk_z k_x \sin(k_x x)\cos(k_y y)e^{\mp jk_z z}}{k^2 - k_z^2} \\
H_y^{\pm} &= \frac{jk_z k_y \cos(k_x x)\sin(k_y y)e^{\mp jk_z z}}{k^2 - k_z^2} \\
E_y^{\pm} &= \frac{\mp j\omega\mu_o k_x \sin(k_x x)\cos(k_y y)e^{\mp jk_z z}}{k^2 - k_z^2} \\
E_x^{\pm} &= \frac{\pm j\omega\mu_o k_y \cos(k_x x)\sin(k_y y)e^{\mp jk_z z}}{k^2 - k_z^2}
\end{aligned}$$

which in turn lead to

$$\begin{aligned}
H_{zt} &= H_z^+ + H_z^- = -j2\cos(k_x x)\cos(k_y y)\sin(k_z z) \\
H_{xt} &= H_x^+ + H_x^- = \frac{j2k_z k_x \sin(k_x x)\cos(k_y y)\cos(k_z z)}{k^2 - k_z^2} \\
H_{yt} &= H_y^+ + H_y^- = \frac{j2k_z k_y \cos(k_x x)\sin(k_y y)\cos(k_z z)}{k^2 - k_z^2} \\
E_{yt} &= E_y^+ + E_y^- = \frac{-2\omega\mu_o k_x \sin(k_x x)\cos(k_y y)\sin(k_z z)}{k^2 - k_z^2} \\
E_{xt} &= E_x^+ + E_x^- = \frac{2\omega\mu_o k_y \cos(k_x x)\sin(k_y y)\sin(k_z z)}{k^2 - k_z^2}.
\end{aligned}$$

TE mode fields:

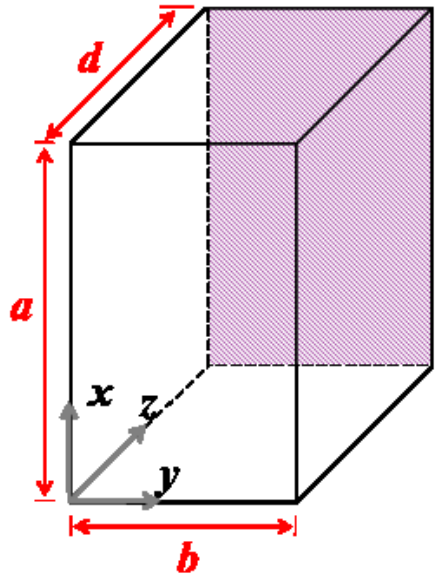
$$\begin{aligned}
H_x^{\pm} &= \frac{\mp jk_z \frac{\partial H_z}{\partial x}}{k^2 - k_z^2}, \\
H_y^{\pm} &= \frac{\mp jk_z \frac{\partial H_z}{\partial y}}{k^2 - k_z^2}, \\
E_y^{\pm} &= \frac{j\omega\mu_o \frac{\partial H_z}{\partial x}}{k^2 - k_z^2}, \\
E_x^{\pm} &= \frac{-j\omega\mu_o \frac{\partial H_z}{\partial y}}{k_z^2}.
\end{aligned}$$



- Standing waves formed with superposed TM_{mn} mode fields having z-components

$$E_z^{\pm} = \sin(k_x x)\sin(k_y y)e^{\mp jk_z z}$$

will likewise produce TM_{mnl} mode resonances in rectangular cavities of dimensions $a > b$ and d having identical resonant frequencies as TE_{mnl} modes.



- For TM_{mnl} modes with $H_{zt} = 0$, a longitudinal standing wave field

$$E_{zt} = E_z^+ + E_z^- = \sin(k_x x)\sin(k_y y)(e^{-jk_z z} + e^{+jk_z z})$$

leads to transverse field components satisfying the boundary conditions at $z = 0$ and $z = d$ provided that

$$k_z = \frac{l\pi}{d}, \quad l = 0, 1, 2, 3 \cdots$$

  - $l = 0$ is allowed in this case since $k_z = 0$ does not lead to "incompatible" boundary conditions (normal E_z and tangential $H_{x,y}$ are allowed on conducting walls at $z = 0$ and d)
  - on the other hand, it is required that m and n are *both* non-zero, a property inherited from propagating TM_{mn} modes.

- TM_{mn} transverse field components accompanying

$$E_z^{\pm} = \sin(k_x x)\sin(k_y y)e^{\mp jk_z z}$$

can be found from the relations given in the margin. They lead to

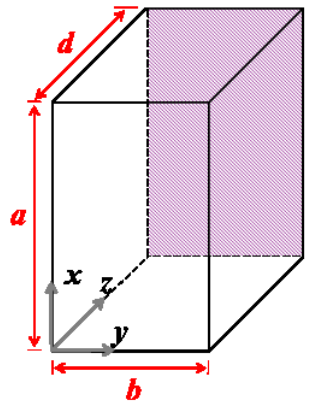
$$\begin{aligned}
H_x^{\pm} &= \frac{j\omega\epsilon_o k_y \sin(k_x x)\cos(k_y y)e^{\mp jk_z z}}{k^2 - k_z^2} \\
H_y^{\pm} &= \frac{-j\omega\epsilon_o k_x \cos(k_x x)\sin(k_y y)e^{\mp jk_z z}}{k^2 - k_z^2} \\
E_y^{\pm} &= \frac{\mp jk_z k_y \sin(k_x x)\cos(k_y y)e^{\mp jk_z z}}{k^2 - k_z^2} \\
E_x^{\pm} &= \frac{\mp jk_z k_x \cos(k_x x)\sin(k_y y)e^{\mp jk_z z}}{k^2 - k_z^2}
\end{aligned}$$

from which

$$\begin{aligned}
E_{zt} &= E_z^+ + E_z^- = 2\sin(k_x x)\sin(k_y y)\cos(k_z z) \\
H_{xt} &= H_x^+ + H_x^- = \frac{j2\omega\epsilon_o k_y \sin(k_x x)\cos(k_y y)\cos(k_z z)}{k^2 - k_z^2} \\
H_{yt} &= H_y^+ + H_y^- = \frac{-j2\omega\epsilon_o k_x \cos(k_x x)\sin(k_y y)\cos(k_z z)}{k^2 - k_z^2} \\
E_{yt} &= E_y^+ + E_y^- = \frac{-2k_z k_y \sin(k_x x)\cos(k_y y)\sin(k_z z)}{k^2 - k_z^2} \\
E_{xt} &= E_x^+ + E_x^- = \frac{-2k_z k_x \cos(k_x x)\sin(k_y y)\sin(k_z z)}{k^2 - k_z^2}.
\end{aligned}$$

TM mode fields:

$$\begin{aligned}
H_x^{\pm} &= \frac{j\omega\epsilon_o \frac{\partial E_z}{\partial y}}{k^2 - k_z^2}, \\
H_y^{\pm} &= \frac{-j\omega\epsilon_o \frac{\partial E_z}{\partial x}}{k^2 - k_z^2}, \\
E_y^{\pm} &= \frac{\mp jk_z \frac{\partial E_z}{\partial y}}{k^2 - k_z^2}, \\
E_x^{\pm} &= \frac{\mp jk_z \frac{\partial E_z}{\partial x}}{k^2 - k_z^2}.
\end{aligned}$$



- Notice that $E_{xt} = 0$ and $E_{yt} = 0$ at both $z = 0$ and $z = d$ provided that $k_z d = l\pi$, as claimed earlier on.
- Furthermore $l = 0$ does not lead to a trivial field since in that case E_{zt}, H_{xt}, and H_{yt} are non vanishing!

- Summarizing the results from above, in a rectangular cavity of dimensions $a > b$ and d and conducting walls, resonant field oscillations at distinct set of frequencies

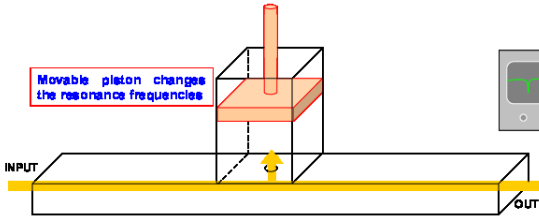
$$f_{mnl} = \frac{c}{2}\sqrt{(\frac{m}{a})^2 + (\frac{n}{b})^2 + (\frac{l}{d})^2}$$

are possible, so long as at least two of the indices m, n, and l are non zero.

- For TE_{mnl} resonances $m = 0$ or $n = 0$ are permitted,
- For TM_{mnl} resonances only $l = 0$ is permitted,
- A resonance of frequency f_{mnl} is is said to be *non-degenerate* if it is allowed for a single mode and it is *degenerate* otherwise.

- **Practical uses of cavities:**

  1. Cavities with small apertures on their walls will interact strongly with external signals (suck them in) having oscillation frequencies matching one of the resonant frequencies, and, conversely, weakly at off-resonant external frequencies. This leads to the usage of cavities as "frequency meters".



  2. Dielectric filled cavities will have resonant frequencies

$$f_{mnl} = \frac{v_p}{2}\sqrt{(\frac{m}{a})^2 + (\frac{n}{b})^2 + (\frac{l}{d})^2} \text{ where } v_p = \frac{1}{\sqrt{\mu\epsilon}}.$$

     Measuring the resonant frequencies of a dielectric-filled cavity is a very accurate means of determining $\sqrt{\mu\epsilon}$.

  3. Cavities filled with active media or devices are a common way of configuring practical signal sources — e.g., lasers.

  4. Microwave ovens are essentially resonant cavities excited by (coupled to) a source operating near some of the resonant frequencies of the cavity that establishes a reasonably smooth field structure where the food is to be placed.

- Our analysis of cavities and waveguides have been based on the assumption of perfectly conducting walls, so far. Waveguide and cavity walls will in practice be very good but imperfect conductors. The implications of this are:

1. Propagating waveguide modes will be weakly attenuated as the field energy is lost into the walls to drive ohmic currents within a few skin-depths of the metallic surface.

   In our idealization of the walls as "perfect conductors", we refer to the depth integral of these volumetric current densities as "surface current densities". In general, an equivalent surface current $\tilde{\mathbf{J}}_s$ on a wall will deliver an average power of

$$S_{loss} = \frac{1}{2} R_s |\tilde{\mathbf{J}}_s|^2 = \frac{1}{2}\sqrt{\frac{\pi f \mu}{\sigma}} |\tilde{\mathbf{J}}_s|^2 \, \frac{\mathrm{W}}{\mathrm{m}^2}$$

   to be dissipated per unit area of the wall (see margin and the ECE 329 notes).

2. Cavity mode oscillations at frequencies f_{mnl} will be damped as a function of time if not "replenished". The rate of energy loss P_{loss} can be calculated by integrating

$$S_{loss} = \frac{1}{2}\sqrt{\frac{\pi f_{mnl} \mu}{\sigma}} |\tilde{\mathbf{J}}_s|^2 \, \frac{\mathrm{W}}{\mathrm{m}^2}$$

   over the 6 cavity walls where at each wall we use $|\tilde{\mathbf{J}}_s|^2 = |\tilde{\mathbf{H}}_{tangential}|^2$. Decay time-constant of the stored mode-energy W, the volume integral of $\frac{1}{4}\epsilon_o |\tilde{\mathbf{E}}|^2 + \frac{1}{4}\mu_o |\tilde{\mathbf{H}}|^2$ within the cavity, is then given by the ratio

$$\tau_{mnl} = \frac{W}{P_{loss}}.$$

Recall from ECE 329, Lecture 26:

Power loss per unit area of a conductor with an equivalent surface current $\tilde{\mathbf{J}}_s$ is

$$S_{loss} = \frac{1}{2} R_s |\tilde{\mathbf{J}}_s|^2$$

where

$$R_s = \sqrt{\frac{\omega\mu}{2\sigma}} = \mathbf{Re}\{\eta_{cond}\}$$

is the surface resistor of the conductor in terms of conductivity σ, permeability μ, and frequency ω.

3. Given the energy dissipation rate τ_{mnl} and the resonant frequency ω_{mnl}, the product

$$Q \equiv \omega_{mnl}\tau_{mnl}$$

is known as **quality factor**. Highly damped modes have small Q, while a high-Q is an indicator of a low-loss cavity.

Review the concept of Q from your ECE 210 notes (see Chpt 12).

4. In cavities with lossy walls in **thermal equilibrium** (i.e., at a steady temperature) it is observed that the average stored energy does not change with time despite the losses in the walls. What that means is that the lossy walls must be radiating as much as they absorb on the average.

The phenomenon of cavity radiation from lossy walls in thermal equilibrium — related to blackbody radiation as well as thermal resistor noise — will explored in the next lecture.

Example 1: Determine f_{nml} and $Q = \tau_{mnl}\omega_{mnl}$ for an air-filled rectangular cavity with $a = b = d = 2$ cm in TE_{101} mode.

Solution: For TE_{m0l} modes $k_y = 0$ and the field expressions derived earlier simplify as

$$\begin{aligned}
H_{zt} &= -j2\cos(k_x x)\cos(k_y y)\sin(k_z z) \to -j2\cos(k_x x)\sin(k_z z) \\
H_{xt} &= \frac{j2k_z k_x \sin(k_x x)\cos(k_y y)\cos(k_z z)}{k^2 - k_z^2} \to \frac{j2k_z \sin(k_x x)\cos(k_z z)}{k_x} \\
H_{yt} &= \frac{j2k_z k_y \cos(k_x x)\sin(k_y y)\cos(k_z z)}{k^2 - k_z^2} \to 0
\end{aligned}$$

$$E_{yt} = \frac{-2\omega\mu_o k_x \sin(k_x x)\cos(k_y y)\sin(k_z z)}{k^2 - k_z^2} \to \frac{-2\omega\mu_o \sin(k_x x)\sin(k_z z)}{k_x}$$
$$E_{xt} = \frac{2\omega\mu_o k_y \cos(k_x x)\sin(k_y y)\sin(k_z z)}{k^2 - k_z^2} \to 0.$$

Therefore, we have

$$\langle \frac{1}{2}\epsilon_o \mathbf{E}\cdot\mathbf{E}\rangle = \epsilon_o \frac{|\tilde{\mathbf{E}}|^2}{4} = \frac{\epsilon_o \omega^2 \mu_o^2 \sin^2(k_x x)\sin^2(k_z z)}{k_x^2}$$

and

$$\langle \frac{1}{2}\mu_o \mathbf{H}\cdot\mathbf{H}\rangle = \mu_o \frac{|\tilde{\mathbf{H}}|^2}{4} = \mu_o[\cos^2(k_x x)\sin^2(k_z z) + \frac{k_z^2 \sin^2(k_x x)\cos^2(k_z z)}{k_x^2}].$$

Volume integrals of these in a cavity with $a = b = c$ replace each trigonometric product in above expressions with $a^3/4$ and thus we obtain

$$W = \frac{a^3}{4}[\frac{\epsilon_o \omega^2 \mu_o^2}{k_x^2} + \mu_o[1 + \frac{k_z^2}{k_x^2}]] = \frac{a^3}{2}\frac{k^2\mu_o}{k_x^2}$$

after using $k^2 = \omega^2\mu_o\epsilon_o$. Also with $a = b = c$ we have $k_x^2 = k_z^2 = k^2/2$ for TE_{101} mode, and hence

$$W = a^3\mu_o.$$

For surface currents on cavity walls we have, on top and bottom walls ($x = 0$ and $x = a$),

$$|\tilde{\mathbf{J}}_s|^2 = |\tilde{H}_z|^2 + |\tilde{H}_y|^2 = 4\sin^2(k_z z) + 0$$

on left and right walls ($y = 0$ and $y = b = a$),

$$|\tilde{\mathbf{J}}_s|^2 = |\tilde{H}_z|^2 + |\tilde{H}_x|^2 = 4\cos^2(k_x x)\sin^2(k_z z) + \frac{4k_z^2 \sin^2(k_x x)\cos^2(k_z z)}{k_x^2}$$

and on front and back walls ($z = 0$ and $z = d = a$),

$$|\tilde{\mathbf{J}}_s|^2 = |\tilde{H}_x|^2 + |\tilde{H}_y|^2 = \frac{4k_z^2 \sin^2(k_x x)}{k_x^2} + 0.$$

Integrating these three expressions over their surfaces, multiplying by 2 (two walls per each expression), and finally scaling by $R_s/2$, we obtain power loss in the walls as

$$P_{loss} = R_s a^2[2 + 1 + \frac{k_z^2}{k_x^2} + \frac{2k_z^2}{k_x^2}] = R_s a^2[2 + 1 + 1 + 2] = 6R_s a^2.$$

Finally

$$\tau = \frac{W}{P_{loss}} = \frac{a^3\mu_o}{6R_s a^2} = \frac{a\mu_o}{6R_s} = \frac{a\mu_o}{6\sqrt{\frac{\omega\mu_o}{2\sigma}}} = \frac{a}{6}\sqrt{\frac{2\sigma\mu_o}{\omega}} \;\Rightarrow\; \omega\tau = \frac{a}{6}\sqrt{2\omega\sigma\mu_o}$$

where ω is the resonance frequency for TE_{101} mode satisfying

$$k^2 = \frac{\omega^2}{c^2} = k_x^2 + k_y^2 + k_z^2 = 2(\frac{\pi}{a})^2 \;\Rightarrow\; \omega = \sqrt{2}\frac{c\pi}{a} \equiv \omega_{101}.$$

Substituting for ω above, we find that

$$\omega\tau = \frac{a}{6}\sqrt{2\omega\sigma\mu_o} = \frac{a}{6}\sqrt{2\sqrt{2}\frac{c\pi}{a}\sigma\mu_o} = \frac{1}{6}\sqrt{2\sqrt{2}c\pi a\sigma\mu_o}$$

which yields for a cavity with copper walls ($\sigma = 6\times 10^7$ S/m)

$$\begin{aligned} Q = \omega\tau &= \frac{\sqrt{2\sqrt{2}}}{6}\sqrt{3\times 10^8\times\pi\times 2\times 10^{-2}\times 6\times 10^7\times 4\pi\times 10^{-7}} \\ &= \frac{\sqrt{2\sqrt{2}}}{6}\sqrt{144\pi^2\times 10^6} = \frac{\sqrt{2\sqrt{2}}}{6}12\pi\times 10^3 \approx 10.56\times 10^3 \sim 10^4. \end{aligned}$$

Also, the resonant frequency of the mode is

$$f_{101} = \frac{\omega_{101}}{2\pi} = \frac{\sqrt{2}\frac{c\pi}{a}}{2\pi} = \frac{c}{\sqrt{2}a} = \frac{3\times 10^8}{\sqrt{2}2\times 10^{-2}} = \frac{30}{2\sqrt{2}}\,\mathrm{GHz} \approx 10\,\mathrm{GHz}.$$

- The result $Q = \tau\omega \sim 10^4$ from Example 1 indicates that the mode oscillates through $10^4/2\pi > 10^3$ cycles over a time period in which the mode energy decays by one e-fold.

37 Resonant modes and field fluctuations

- Since in a rectangular cavity the resonant frequencies

$$f_{mnl} = \frac{c}{2}\sqrt{(\frac{m}{a})^2 + (\frac{n}{b})^2 + (\frac{l}{d})^2} \quad \Rightarrow \quad \frac{2f_{mnl}}{c} = \sqrt{(\frac{m}{a})^2 + (\frac{n}{b})^2 + (\frac{l}{d})^2},$$

  we can consider $2f_{mnl}/c$ to be the "length" of a "vector" $(\frac{m}{a}, \frac{n}{b}, \frac{l}{d})$ pointing away from the origin of a "3D Cartesian space" where each *lattice point*, e.g., $(\frac{m}{a}, \frac{n}{b}, \frac{l}{d}) = (\frac{1}{a}, \frac{2}{b}, \frac{1}{d})$, is associated with two resonant modes (TE and TM) of the cavity.

  - In this space, "volume" per lattice point is $\frac{1}{abd}$, and thus *volume per resonant mode* is $\frac{1/2}{abd}$.
  - Also, all the resonant modes with resonance frequencies f_{mnl} less than a given frequency f can be associated with lattice points residing within one eight (an octant) of a sphere of "radius" $2f/c$ centered about the origin of the same space — only an octant is involved since the indices m, n, l employed are all non-negative.

  Thus, the number of resonant modes with frequencies less than f, to be denoted as the cumulative distribution $C(f)$, is found to be

$$C(f) = \frac{\frac{1}{8} \times (\text{sphere of radius } 2f/c)}{\frac{1/2}{abd}} = \frac{\frac{1}{8} \times \frac{4\pi}{3}(\frac{2f}{c})^3}{\frac{1/2}{abd}} = \frac{8\pi f^3}{3c^3}V$$

where $V = abd$ is the physical volume of the cavity. Consequently, the number density $N(f)$ of the available resonant modes in a cavity of volume V is obtained as

$$N(f) = \frac{dC}{df} = \frac{8\pi f^2}{c^3} V \quad \frac{\text{modes}}{\text{Hz}}$$

which grows quadratically with frequency f. As illustrated later in this lecture, the distribution $N(f)$ has deep theoretical implications.

Example 1: Consider a rectangular cavity with dimensions $a = b = d = 0.3$ m. Determine $N(f)$ for $f = 50$ GHz and the number of resonant modes to be found within a bandwidth of $\Delta f = 1$ GHz centered about $f = 50$ GHz.

Solution: Using the density function derived above, we find that

$$N(50 \times 10^9) = \frac{8\pi(50 \times 10^9)^2}{(3 \times 10^8)^3}(3 \times 10^{-1})^3 = 8\pi \times 25 \times 10^{-7} = 2\pi \times 10^{-5} \frac{\text{modes}}{\text{Hz}}.$$

Thus, the number of resonant modes in a bandwidth of $\Delta f = 1$ GHz centered about $f = 50$ GHz is

$$\text{Number of modes within band} = (2\pi \times 10^{-5} \frac{\text{modes}}{\text{Hz}}) \times 10^9 \text{ Hz} = 20000\pi \approx 60000.$$

Energy spectrum of radiation in enclosed cavities:

- Consider an air-filled rectangular cavity with slightly lossy walls sitting on a table top in some lab where the room temperature is 300 K. Assume that the cavity has been in the room for a long time and has reached thermal equilibrium with the rest of the room — i.e., the walls of the cavity also have $T = 300$ K.

  It turns out that such a cavity will be filled with electromagnetic fields consisting of (i.e., a superposition of) the TE_{mnl} and TM_{mnl} modes distributed across the frequency space with a density function

  $$N(f) = \frac{8\pi f^2}{c^3} V$$

  derived above.

  - The resonant modes with the distribution function just quoted are the result of radiation by random currents flowing on the cavity walls caused by random thermal agitations of the charge carriers located within the walls.
  - As soon as it is (randomly) established, a resonant mode will start decaying because of ohmic losses in cavity walls (see earlier discussions), returning back the radiated energy of the wall back to the wall.
  - In thermal equilibrium the temperature of the wall as well as the *expected* total energy of cavity radiation summed over all of its modes will remain constant.

- The energy density spectrum of the radiation within the cavity, $E(f)$, measured in units of J/m^3/Hz, should be product of $N(f)/V$ with $\langle W(f)\rangle$ representing the expected value (statistical average) of the energy $W(f)$ of each mode at resonant frequency f. Hence,

$$E(f) = \frac{8\pi f^2}{c^3}\langle W(f)\rangle.$$

- What might be the expected mode energy $\langle W(f)\rangle$?

- Each resonant mode such as TE_{101} or TM_{112} can be interpreted as two *degrees of freedom* (one degree for **E** and one for **H**) of the electromagnetic field variations in a closed cavity, just as velocity components v_x, v_y, v_z of any one of N molecules contained within a volume of gas are each considered a "degree of freedom" for the N molecule system.

  - In physical models each degree of freedom in a gas in thermal equilibrium is assigned[1] an expected energy of

$$\langle \frac{1}{2} m v_x^2 \rangle = \frac{1}{2} KT$$

    where $K \equiv 1.38 \times 10^{-23}$ J/K is *Boltzmann's constant* and T is the equilibrium temperature in K.

[1] In thermal equilibrium all particles have by definition equal average energies. Denoting this energy as $\frac{1}{2}KT$ is just a matter of *defining* the **equilibrium temperature** of the gas in terms the **average kinetic energy** of its individual molecules — at a fundamental level that *is* what temperature is! Including the Boltzmann constant K in this assignment is just a matter of setting the scale used for temperature (Kelvin scale by convention). At room temperature (298 K), KT works out to be 0.0256 eV.

- If we naively make a similar assignment (see margin note) to $\langle W(f)\rangle$, e.g., take

$$\langle W(f)\rangle = KT$$

(on account of the fact that TE_{mnl} and TM_{mnl} modes have energies which are the sum of two quadratic terms proportional to $|\tilde{\mathbf{E}}|^2$ and $|\tilde{\mathbf{H}}|^2$), we then immediately run into a difficulty in that $E(f)$ blows up to infinity in the high frequency end because $\langle W(f)\rangle$ has no high-frequency cutoff.

- The difficulty just mentioned — known as "ultraviolet catastrophe" — was well recognized at the beginning of the 20th century, and was resolved by Max Planck's recognition that electromagnetic mode energies $W(f_{mnl})$ have to be *quantized* in chunks of size hf_{mnl}, and

$$\langle W(f_{mnl})\rangle = KT$$

is acceptable only if an "energy quantum" $hf_{nml} \ll KT$.

If $hf_{nml} \gg KT$ for a given mode, then the mode is very seldom excited (to an energy level of one hf_{nml}), and thus the expected value of energy $W(f_{mnl})$ in the mode is an exponentially reduced fraction of an energy quantum hf_{nml} given by

$$\langle W(f_{mnl})\rangle = hf_{mnl}\, e^{-hf_{mnl}/KT}.$$

This effective "cutoff" in $\langle W(f)\rangle$ function eliminates the ultraviolet catastrophe.

$\frac{1}{2}KT$ per quadratic term: Assigning an expected energy of $\frac{1}{2}KT$ per quadratic term in a total energy expression of a large system of elements in thermal equilibrium is a standard procedure used in classical statistical mechanics. This is a consequence of well known experimental results such as: in a gas consisting of a mixture of light and heavy atoms, $\langle \frac{1}{2}mV_x^2\rangle$ of the light atoms match $\langle \frac{1}{2}Mv_x^2\rangle$ of the heavy atoms in thermal equilibrium — all quadratic energy terms get the same $\frac{1}{2}KT$ (classically)!

– Using the **1st** and **2nd laws of thermodynamics** together with the **quantization rule** that he introduced, Planck derived[2] the relation

$$\langle W(f_{mnl})\rangle = \frac{hf_{mnl}}{e^{hf_{mnl}/KT}-1}$$

for the expected mode energies having the limiting cases for

$$hf_{nml} \ll KT \text{ and } hf_{nml} \gg KT$$

just discussed.

– With this result, the energy spectrum within a cavity in thermal equilibrium takes the form

$$E(f) = \frac{N(f)}{V}\langle W(f)\rangle = \frac{8\pi f^2}{c^3}\frac{hf}{e^{hf/KT}-1}\,\frac{\text{J/m}^3}{\text{Hz}}.$$

This derived spectral shape was successfully adjusted to fit the observed energy spectra of cavity radiation by varying the parameter h, which is now known as *Planck's constant*[3] and has the fixed value of 6.626×10^{-34} Js.

Shape independence:
$E(f)$ obtained for the rectangular cavity is actually *independent* of cavity shape. This can be justified by considering two cavities, one rectangular, one not, joined by a small aperture. If the two cavities have the same temperature T, then *by definition* (of T) there cannot be any net energy exchange between the cavities at any f (a *detailed balance* per frequency is required because the aperture may have an f dependent transmittivity) — hence a common $E(f)$ for the two cavities with a common T even if the shapes are different!
$E(f)$ is also independent of the lossiness of the walls (even though Q of the cavity depends on it) and therefore applicable to *all* lossy cavities at thermal equilibrium including those whose walls are *perfect absorbers*, i.e., **blackbodies**.

[2]See *Oliver, B. M.*, "Thermal and Quantum Noise", Proc IEEE, **53**, 436 (1965) for a simplified version of Planck's derivation.

[3]Planck's constant h is one of the three fundamental constants of physics, along with c and G, the gravitational constant, from which *absolute units* for all physical variables can be derived in suitable combinations: e.g., length unit=$\sqrt{hG/c^3}$, time unit=$\sqrt{hG/c^5}$, etc.

38 Cavity radiation and thermal noise

- In a 1D cavity of some length L — e.g. a TL with *shorts* at both ends as discussed in ECE 329 notes — the resonant frequencies are

$$f_m = \frac{c}{\lambda_m} = \frac{c}{2L/m} = \frac{c}{2L}m, \text{ where } m = 1, 2, 3, \cdots$$

which indicates that the mode density in f is

$$N(f) = \frac{2L}{c} \ \frac{\text{modes}}{\text{Hz}}.$$

Therefore the energy density in a 1D cavity in thermal equilibrium (assume vanishingly lossy wires with temperature T) will be

$$E(f) = \frac{N(f)}{L}\langle W(f)\rangle = \frac{2}{c}\frac{hf}{e^{hf/KT}-1}\frac{\text{J/m}}{\text{Hz}}$$

in analogy with the energy density of 3D cavities. This energy density will reside by equal amounts in the traveling wave components of the 1D resonant modes arriving with speed c from the opposite ends of the 1D resonator. Power spectral content $P(f)$ of each of these traveling wave components can thus be calculated as $c/2$ times[1] $E(f)$, i.e.,

$$P(f) = \frac{hf}{e^{hf/KT}-1}\frac{\text{W}}{\text{Hz}}.$$

Cavity radiance: Energy density

$$\begin{aligned} E(f) &= \frac{N(f)}{V}\langle W(f)\rangle \\ &= \frac{8\pi f^2}{c^3}\frac{hf}{e^{hf/KT}-1}\frac{\text{J/m}^3}{\text{Hz}}. \end{aligned}$$

in a 3D cavity in thermal equilibrium resides by equal amounts in the traveling wave components of the cavity modes arriving with speed c from the boundaries of the cavity subtending 4π sterads. Multiplying $E(f)$ by $c/4\pi$ we obtain

$$L(f) = \frac{2f^2}{c^2}\frac{hf}{e^{hf/KT}-1}\frac{\text{W/m}^2/\text{ster}}{\text{Hz}},$$

which is called *radiance* and represents the *power density per unit solid angle* of the waves traveling within the cavity.

Radiance $L(f)$ also represents the spectrum of power *radiated* per unit solid angle by a unit area of a **blackbody surface** at temperature T (since non-reflective walls of a cavity will produce the same $E(f)$ as partial-reflecting walls as mentioned earlier).

[1] Note that per TEM plane wave,

$$c(\frac{1}{4}\epsilon_o|\tilde{\mathbf{E}}|^2 + \frac{1}{4}\mu_o|\tilde{\mathbf{H}}|^2) = \frac{|\tilde{\mathbf{E}}|^2}{4\eta_o} + \eta_o\frac{|\tilde{\mathbf{H}}|^2}{4} = \frac{|\tilde{\mathbf{E}}|^2}{2\eta_o},$$

which confirms that the time-averaged stored energy density times c is indeed the time-averaged power transported per unit area.

- Now replace the shorts at the ends of the resonator with resistors R at temperature T *matching* the characteristic impedance Z_o of the line.

  Since there cannot be any net power exchange between elements in thermal equilibrium (over *any* frequency band — otherwise a net broadband exchange can be arranged for by using filters with suitable frequency responses in violation of the 2nd law of thermodynamics), it follows that matched resistors R will be both absorbing (a full absorption because of impedance matching) and injecting (to a matched load again because of the same fact) the same power density $P(f)$ identified above.

- The upshot is, we need to conclude that any resistor R at a temperature T must have an *available* power density of

$$P(f) = \frac{hf}{e^{hf/KT} - 1} \frac{\text{W}}{\text{Hz}}$$

  fueled by thermal agitation of its internal charge carriers. The resistor is then capable of outputting an average power of

$$P(f)\Delta f = \frac{hf}{e^{hf/KT} - 1}\Delta f \equiv \frac{\langle v^2 \rangle}{4R}$$

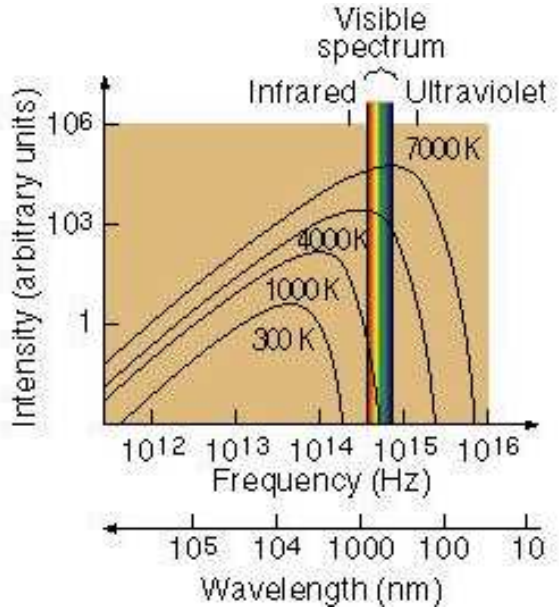
  over a bandwidth Δf, where $\langle v^2 \rangle$ is the mean squared *open-circuit voltage* at the resistor terminals over the same bandwidth.

Unmatched termination case: If $R \neq Z_o$, then a portion $P(f)|\Gamma|^2$ of the incident power $P(f)$ will be reflected from R (rather than being fully absorbed). In that case R emits only a reduced level of power $P(f)(1 - |\Gamma|^2)$. This is a simple example of how $P(f)$ can be emitted in its entirety only by **perfect absorbers** defined to be blackbodies. In 1D, the *blackbody* radiance at temperature T is

$$P(f) = \frac{hf}{e^{hf/KT} - 1} \frac{\text{W}}{\text{Hz}}$$

while in 3D it is

$$L(f) = \frac{2f^2}{c^2} \frac{hf}{e^{hf/KT} - 1} \frac{\text{W/m}^2\text{/ster}}{\text{Hz}}.$$



- It follows that

$$\frac{\langle v^2 \rangle}{4R} = \frac{hf\Delta f}{e^{hf/KT} - 1} \text{ reducing to } \frac{\langle v^2 \rangle}{4R} = KT\Delta f \text{ for } hf \ll KT,$$

  a result known as **Nyquist noise theorem**. The theorem can also be cast as

$$\frac{\langle i^2 \rangle}{4G} = \frac{hf\Delta f}{e^{hf/KT} - 1} \text{ reducing to } \frac{\langle i^2 \rangle}{4G} = KT\Delta f \text{ for } hf \ll KT$$

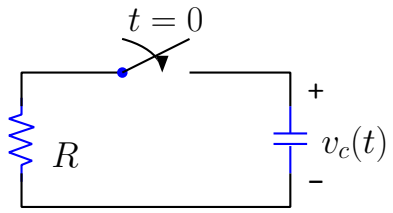
  in terms of mean squared *short-circuit current* $\langle i^2 \rangle$ of the same resistor over the same bandwidth and conductance $G = 1/R$. Note that if the element has an impedance $Z = R + jX = 1/Y$ only the real part of Z should be utilized in connection with power transferred to a matched load Z^*.

- Nyquist noise theorem outlined above has a very powerful generalization known as the **fluctuation-dissipation theorem**:

  - according to this theorem, any linear and dissipative system in thermodynamic equilibrium will exhibit thermally driven fluctuations of its dynamic parameters (e.g., electron density in plasma at finite temperature), and
  - the frequency spectrum of the fluctuations can be obtained by applying the Nyquist noise theorem to an appropriately constructed equivalent circuit model of the system.

- The unavoidable fact of fluctuations and noise encountered in dissipative systems and circuits constitutes both a challenge and an opportunity for the engineer. Consider taking ECE 453 to develop a better understanding of noise issues in communication circuits.

(a) Initial value problem:



Example: Consider the RC circuit shown in the margin. Assuming that the capacitor holds 10 V prior the switch is closed at $t = 0$, the capacitor voltage for $t > 0$ can be expressed as

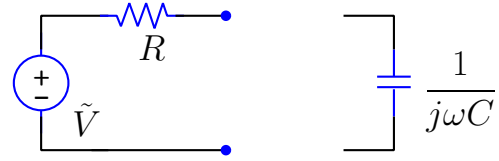
$$v_c(t) = 10e^{-t/RC}$$

using ECE 210 knowledge. This solution implies the dissipation of the initial stored energy within the resistor. But as we have seen in this lecture, dissipative elements such as resistors also produce random thermal voltages and currents. We therefore expect a *non-zero* $v_c(t)$ in the circuit shown in the margin as $t \to \infty$, assuming that the resistor has some non-zero steady-state temperature T measured in Kelvins. Given that the resistor produces an open circuit voltage $v(t)$ with a mean-squared value of

$$\langle v^2 \rangle = 4RKT\Delta f \quad \text{(Nyquist noise formula)}$$

over any bandwidth Δf, let us calculate the mean-squared capacitor voltage $\langle v_c^2 \rangle$ in the circuit over all frequencies f.

(b) Frequency domain Thevenin model of a noisy resistor:



Calculation: The Thevenin equivalent circuit modeling the noisy resistor in the phasor domain is shown in the margin. The model includes a source voltage phasor $\tilde{V}$. A capacitor C with impedance $Z_c = \frac{1}{j\omega C}$ connected across the terminals of the equivalent model circuit will develop a phasor voltage

$$\tilde{V}_c = \tilde{V}\frac{Z_c}{R + Z_c} = \tilde{V}\frac{1}{1 + j\omega RC}$$

as dictated by voltage division. The mean-squared value of a co-sinusoidal oscillation $v_c(t) = \mathrm{Re}\{V_c e^{j\omega t}\}$ with the phasor V_c would then be

$$\frac{1}{2}|\tilde{V}_c|^2 = \frac{1}{2}|\tilde{V}|^2 \frac{1}{|1 + j\omega RC|^2} \equiv \frac{1}{2}|\tilde{V}|^2 ||H(f)|^2$$

where $\frac{1}{2}|\tilde{V}|^2$ is the mean-squared value of open circuit voltage $v(t)$ of the resistor and

$$|H(f)|^2 = \frac{1}{|1 + j2\pi fRC|^2} = \frac{1}{1 + (2\pi fRC)^2}$$

is a frequency dependent scaling factor — the magnitude square of the frequency response function of the circuit — between the two mean-square quantities.

Now, the mean-squared voltage output

$$\langle v^2 \rangle = 4RKT\Delta f \quad \text{(Nyquist noise formula)}$$

of the noisy resistor over a small but finite bandwidth Δf can be scaled likewise to obtain

$$\langle v_c^2 \rangle = \langle v^2 \rangle |H(f)|^2 = \frac{4RKT\Delta f}{1 + (2\pi fRC)^2},$$

the mean-squared voltage output across the capacitor over the same bandwidth provided that $|H(f)|^2$ is fairly constant over the band. For wider bands where the constancy condition is violated, use

$$\langle v_c^2 \rangle = \int_{f_1}^{f_2} \frac{4RKT}{1 + (2\pi fRC)^2} df,$$

whereas the broadband value over *all* frequencies (with $f_1 \to 0$ and $f_2 \to \infty$) is

$$\langle v_c^2 \rangle = \int_0^{\infty} \frac{4RKT}{1 + (2\pi fRC)^2} df = \frac{KT}{C},$$

a value independent of R.

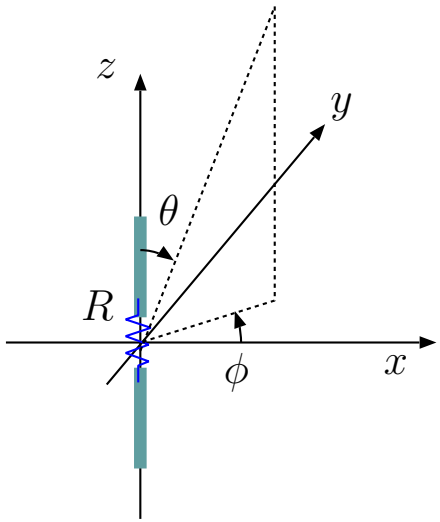
For a 1 pF capacitor $K = 1.38 \times 10^{-23}$ J/K and $T = 300$ K gives an rms (root mean squared) voltage of

$$\langle v_c^2 \rangle^{1/2} = \frac{KT}{C} \approx 0.65\,\text{mV},$$

easily detected in the lab.

39 Antenna noise

Thermodynamic derivation of $A(\theta,\phi) = \frac{\lambda^2}{4\pi}G(\theta,\phi)$:



- Consider a z-polarized half-wave dipole antenna with a transmission impedance of $Z_{ant} = 73\,\Omega$ embedded within a large cavity in thermal equilibrium at some thermodynamic temperature T. A $73\,\Omega$ resistor, also at the same equilibrium temperature T as the rest of the cavity, is connected between the terminals of the dipole as shown in the margin.

  - As we will learn in Lecture 38, a hot resistor will deliver an average power of $KTdf$ over a bandwidth of df to any matched load, which in this case is the dipole antenna.

- The dipole will radiate its average power input $P_t = KTdf$ with a gain $G(\theta,\phi)$ to maintain a radiated power density (Ponting flux) of

$$\frac{KTdf}{4\pi r^2}G(\theta,\phi)\,\frac{\mathrm{W}}{\mathrm{m}^2}$$

  at a distance r within the cavity, amounting to a radiated power of

$$\frac{KT\,df\,d\Omega}{4\pi}G(\theta,\phi)\,\mathrm{W}$$

  over an infinitesimal area of $dS = r^2 d\Omega$ spanning an infinitesimal solid angle increment $d\Omega$. The corresponding outgoing radiated power den-

sity of the dipole

$$\frac{KT}{4\pi}G(\theta,\phi)\,\frac{\text{W}}{\text{ster/Hz}}$$

is carried by $\hat{\theta}$-polarized fields as we know from our studies of dipole radiation.

- The same dipole with its matched termination is also exposed to incoming cavity radiation with a radiance of

$$L(\theta,\phi,f)=\frac{2KT}{\lambda^2}\,\frac{\text{W/m}^2}{\text{ster/Hz}},$$

coming from direction θ,ϕ as we will learn in Lecture 38, of which half the amount

$$\frac{1}{2}L(\theta,\phi,f)=\frac{KT}{\lambda^2}\,\frac{\text{W/m}^2}{\text{ster/Hz}}$$

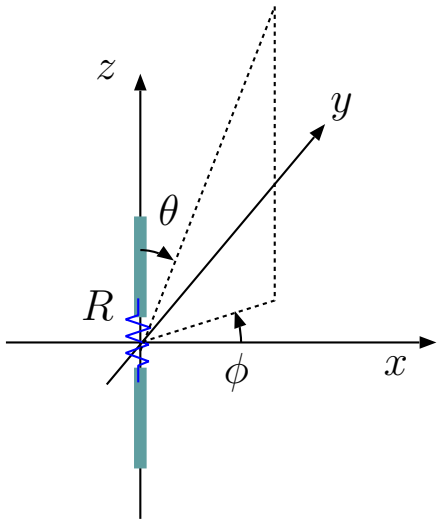
is contained in $\hat{\theta}$-polarized (i.e., co-polarized) incoming waves incident on the dipole.

  - Let the product

$$\frac{1}{2}L(\theta,\phi,f)A(\theta,\phi)=\frac{KT}{\lambda^2}A(\theta,\phi)\,\frac{\text{W}}{\text{ster/Hz}}$$

    of the *effective area* $A(\theta,\phi)$ of the dipole with the co-polarized radiance $\frac{1}{2}L(\theta,\phi,f)$ incident on the dipole represent the *average power per steread per Hz* delivered by the dipole to its impedance

matched $73\,\Omega$ load, balancing the outgoing power loss of the resistor so that it remains at a constant temperature as demanded by the 2nd Law of Thermodynamics for a system in thermal equilibrium.



- In that case

$$\frac{\cancel{KT}}{\lambda^2}A(\theta,\phi)\frac{\mathrm{W}}{\cancel{\mathrm{ster}/\mathrm{Hz}}} = \frac{\cancel{KT}}{4\pi}G(\theta,\phi)\frac{\mathrm{W}}{\cancel{\mathrm{ster}/\mathrm{Hz}}}$$

leading to

$$A(\theta,\phi) = \frac{\lambda^2}{4\pi}G(\theta,\phi).$$

- While this result for the antenna effective area was obtained by considering a half-wave dipole antenna, the same conclusion could have been reached using any type of antenna with a conjugate matched complex load impedance $Z_L = Z_{ant}^*$ connected between its terminals, providing, in thermal equilibrium, both the source of the radiated power of the antenna into the cavity and the sink of the received power extracted from cavity radiation.
  - As such, we conclude that $A = \frac{\lambda^2}{4\pi}G$ relation holds true for all types of antennas and can be used in $P_r = S_{inc}A$ equation to calculate the power delivered by any antenna into its conjugate matched load in terms of the co-polarized incident flux $S_{inc.}$
  - Furthermore, from the universality of $A = \frac{\lambda^2}{4\pi}G$ we directly infer the independence of P_r/P_t ratio from the choice of transmission

and reception roles assigned to a pair of coupled antennas (see Lecture 20) and thus infer the reciprocal two-port network model leading to $P_r = \frac{|V_T|^2}{8Re\{Z_T\}}$ formula (see Lecture 20 again), to provide us, in turn, with a means to calculate $|V_o| = |V_T|$, the magnitude of the open circuit voltage of a receiving antenna.

- Also the phase of the antenna open circuit voltage V_o can be deduced to match the phase of $-d\mathbf{l} \cdot \mathbf{E}$ in the gap region of the receiving antenna for any incoming field $\mathbf{E}$.
- Even the most general

$$V_o = -\int_L d\mathbf{l} \cdot \mathbf{E}(\mathbf{r}) \frac{I(\mathbf{r})}{I_o}$$

formula for an antenna open circuit voltage can be "reverse engineered" to match the $P_r = \frac{|V_T|^2}{8Re\{Z_T\}} = S_{inc} A(\theta, \phi)$ requirement (as we have done in Lecture 20) with $V_T = V_o$.

- All these lead to a simplified description of antenna reciprocity and reception much used by RF engineers — mainly the methods we learned in Lecture 20 — without going through the details of the Reciprocity Theorem based derivations covered in Lecture 21.

- We have seen that the *antenna effective area* $A(\theta, \phi)$ can be used to covert
  - the power density S_{inc} of only the *co-polarized* component of the incident field into [1]
  - the power delivered P_r into a resistive termination conjugate *matching* the antenna input impedance Z_{ant}.
- Consider an arbitrary antenna located at the origin with
  - some termination R,
  - a gain function $G(\theta, \phi)$

  which is exposed to a *spectrum* of incident plane TEM waves with *co*-polarized field components $\tilde{\mathbf{E}}_i$ at a frequency $f = \frac{\omega}{2\pi}$ arriving from directions (θ_i, ϕ_i) with power densities S_i.
- Each co-polarized field $\tilde{\mathbf{E}}_i$ could deliver an average power of up to

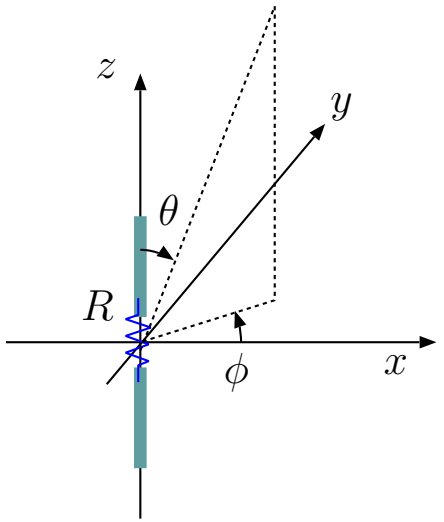
$$P_r = S_i A(\theta_i, \phi_i)$$

  to R one-at-a-time — where $A(\theta, \phi)$ is the appropriate effective area of the antenna for reception — and, furthermore, the expected value of

The relation

$$A(\theta, \phi) = \frac{\lambda^2}{4\pi} G(\theta, \phi)$$

applies to all antennas universally.



[1] The component of the incident field that is not *co*-polarized - called *cross*-polarized - is not detected at all by the receiving antenna.

P_r extracted from the full spectrum of $\tilde{\mathbf{E}}_i$ would be

$$\langle P_r \rangle = \sum_i S_i A(\theta_i, \phi_i)$$

when phasors $\tilde{\mathbf{E}}_i$ have *random phase offsets* distributed independently between 0 and 2π radians (see the margin for an explanation).

– This sum generalizes to an integral

$$\langle P_r \rangle = \int \int dS(\theta, \phi, f) A(\theta, \phi)$$

in case of a continuum of random incident waves over a band of frequencies f, which in turn takes a specific form

$$\langle P_r \rangle = \int \int \frac{1}{2} L(\theta, \phi, f) A(\theta, \phi) d\Omega df,$$

with

$$dS(\theta, \phi, f) = \frac{1}{2} L(\theta, \phi, f) d\Omega df = \frac{KT}{\lambda^2} d\Omega df.$$

in case of co-polarized radiance of cavity radiation at an equilibrium temperature T.

- In a practical application an antenna will be used to detect a specific plane-wave signal of interest, of, say, power density S_s, in the presence of a random continuum $dS(\theta, \phi, f)$ corresponding an some *effective*

***Incoherent power addition*: Consider a voltage**

$$v(t) = V_1 \cos(\omega t) + V_2 \cos(\omega t + \phi)$$

applied across a $1\,\Omega$ resistor where ϕ is a *random* phase shift parameter. Squaring $v(t)$ and taking its time-average it can be shown that the time average-power

$$\begin{aligned} P &= \frac{1}{T} \int_T v^2(t) dt \\ &= \frac{1}{2} V_1^2 + \frac{1}{2} V_2^2 + V_1 V_2 \cos(\phi). \end{aligned}$$

Notice the third term in P. With random ϕ we would be unable to know P, but still its *expected value* is

$$\langle P \rangle = \frac{1}{2} V_1^2 + \frac{1}{2} V_2^2,$$

which is the *incoherent sum* of the time-average power due to signals 1 and 2 *one-at-a-time*.

temperature T_a. In that case, we obtain over a finite frequency band[2] Δf containing S_s, an average received power of

$$\begin{aligned}\langle P_r \rangle &= S_s A(\theta_s, \phi_s) + \int_{\Delta f} \int dS(\theta, \phi, f) A(\theta, \phi) \\ &= S_s A(\theta_s, \phi_s) + KT_a \Delta f,\end{aligned}$$

since

$$\begin{aligned}\int_{\Delta f} \int dS(\theta, \phi, f) A(\theta, \phi) &= \int_{\Delta f} \int \frac{1}{2} L(\theta, \phi, f) A(\theta, \phi) d\Omega df \\ &= \Delta f \int \frac{KT_a}{\lambda^2} A(\theta, \phi) d\Omega = KT_a \Delta f.\end{aligned}$$

- The second term of $\langle P_r \rangle$ above is called noise power, or simply "noise", while the first term represents the signal power, or simply the "signal".

- Since we want the "signal-to-noise ratio" SNR as large as possible, we should select Δf as small as allowed, but regrettably $\Delta f = 0$ is not an option in practical communication applications because of the finite bandwidth of the signal of interest (with $\Delta f = 0$ we would lose the signal together with the noise!).

The effective temperature T_a introduced above deserves some further explanation:

[2] A suitable Δf is imposed by using an appropriate band-pass filter in cascade with the matched antenna termination.

We will routinely use the approximation

$$\begin{aligned}L(f) &= \frac{2f^2}{c^2} \frac{hf}{e^{hf/KT} - 1} \\ &\approx \frac{2KT}{\lambda^2}\end{aligned}$$

in this lecture since we are concerned with radio frequencies f within narrow bands Δf over which

$$hf \ll KT.$$

Recall that radiance $L(f)$ is measured in units of W/m^2/Hz/ster.

- Since in normal usage an antenna will not be placed in a cavity in thermal equilibrium, our usage of

$$L(\theta, \phi, f) = L(f) = \frac{2KT_a}{\lambda^2}$$

above is just a convention that helps us to parametrize the noise power collected by a receiving antenna in terms of an equivalent temperature T_a referred to as "antenna temperature".

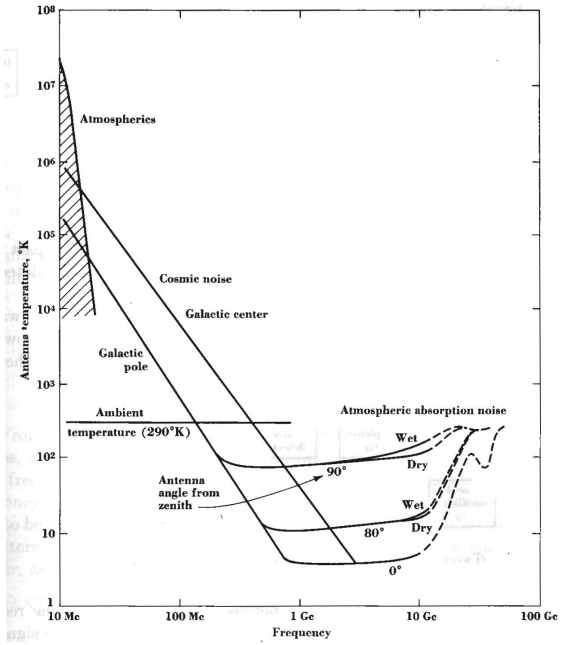
Antenna temperature T_a may or may not be the physical (thermodynamic) temperature of a noise source in the sky:

- When the antenna beam is fully intercepted (at some large distance) by a perfect absorber (a blackbody) of an equilibrium temperature T_b, then then the antenna temperature will be

$$T_a = T_b$$

and measure the thermodynamic temperature of the absorbing body. In radio astronomy this fact is used to measure the surface temperatures of planets such as Venus (using a geometrical correction factor to account for the fact that the planet will not fully intercept the antenna beam).

- The universe can be regarded as a *cavity* sparsely filled with objects (galaxies, stars, planets, living things) *not* in thermal equilibrium with the cavity walls.



Antenna temperature curve from Kraus, "Radio Astronomy" (1966): Note that T_a is dominated by atmospheric noise (lightning, man-made noise) at low frequencies and by thermal emission from absorbing gases in the atmosphere past about a GHz. So-called cosmic noise from radio galaxies (including our own) dominates in VHF and UHF band.

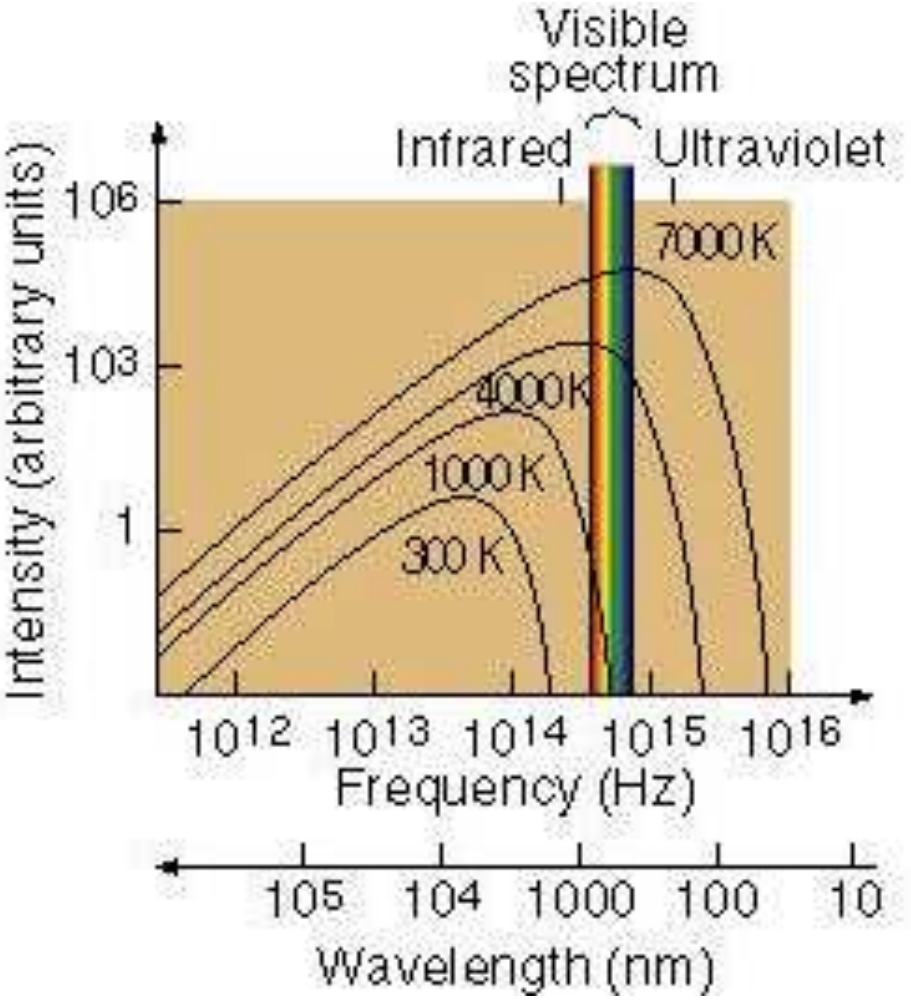
The "walls" of the cavity — at about 13.7 billion light-years away — are known to have an equilibrium temperature of about 3 K (discovered in 1964 and taken to be the strongest evidence for Big Bang) that produces a blackbody radiance curve

$$L_b(f) = \frac{2f^2}{c^2}\frac{hf}{e^{hf/KT_b} - 1}$$

that peaks within the microwave frequency band.

However, galaxies and intergalactic gases radiate by a variety of non-thermal processes (e.g., synchrotron radiation) so that the overall radiance spectrum $L(\theta, \phi, f)$ (the sum of all contributions) is non-thermal, giving rise to a frequency dependent $T_a(\theta, \phi, f)$ curve shown in the margin.

The noise power collected by an antenna can be conveniently calculated by multiplying this "equivalent" antenna temperature $T_a(\theta, \phi, f)$ with $K\Delta f$.



Example 1: A typical GSM cell tower will radiate 10 W of average power at a carrier frequency of 750 MHz. Assuming a 1 km direct path between the cell tower and a GSM hand held receiver we computed the received power of

$$P_r = P_t \frac{\lambda^2 G_1 G_2}{(4\pi r)^2} = 10\,(\frac{0.4 \times 1.5}{4\pi 10^3})^2 \approx 2.28 \times 10^{-8}\,\mathrm{W} = 22.8\,\mathrm{nW}$$

in Example 3 in Lecture 38 assuming that both the tower and the receiver make use of small antennas of directivities $D = 3/2$.

To calculate SNR, signal-to-noise ratio for the hand-held unit, we look up the antenna temperature T_a for $f = 750$ MHz from the chart on page 4 — we see that $T_a \sim 100$ K. With a GSM channel bandwidth of $\Delta f = 200$ kHz, and Boltzmann constant $K = 1.38 \times 10^{-23}$J/K, we deduce a noise power of

$$KT_a\Delta f = 1.38 \times 10^{-23} \times 100 \times 2 \times 10^5 = 2.76 \times 10^{-16}\,\mathrm{W}.$$

If we were to use $T_{room} = 300$ K instead of $T_a = 100$ K, we would have

$$KT_a\Delta f = 1.38 \times 10^{-23} \times 300 \times 2 \times 10^5 \sim 10^{-15}\,\mathrm{W} = 10^{-6}\,\mathrm{nW}.$$

Clearly, the SNR of the hand held unit $\gg 1$ and the system should work successfully.