Electromagnetism

Università degli studi di Roma "La Sapienza"
Physics and Astrophysics BSc

Matteo Cheri

Notes on Electromagnetism

May 24, 2023

Version 0.1

Electromagnetism

Notes on Electromagnetism

Written by

Matteo Cheri

Università degli Studi di Roma "La Sapienza"
Physics and Astrophysics BSc

LaTeX 2_ε inside, VIM powered.

May 24, 2023

Version 0.1

Contents

Part I

Electrostatics

1 The Electric Field

§ 1.1 Electric Forces and the Electric Field

It has been found from the forefathers of Electrodynamics that, empirically, the force exerted between two charged objects has the following characteristics

1. It's directed from one object to the other
2. It depends on the *product* of the charges
3. It's proportional to the inverse squared of the distance between the objects d^{-2}

The experimental results brought with great clarity then, that for two point charges q_1, q_2, said $r = \left\| r_1^i - r_2^i \right\|$ and $\hat{r}^i$ the associated versor, the electrostatic force is

$$F^i = k_e \frac{q_1 q_2}{r^2} \hat{r}^i \tag{1.1}$$

Here, k_e is a coupling constant which takes different values for different choices of units. In the SI system we have

$$k_e = \frac{1}{4\pi\epsilon_0} \tag{1.2}$$

With ϵ_0 being the *permittivity of free space*, which has value

$$\epsilon_0 = 8.85 \cdot 10^{-12} \ \frac{\mathrm{C}}{\mathrm{Nm}^2} \tag{1.3}$$

These forces are obviously additive.
Suppose now that you have a set of n charges q_i and you add an imaginary test charge Q in order to *theoretically* test the force field generated by these charges. We have then

$$F^i = \sum_{j=1}^{n} f^i_{(j)} = \sum_{j=1}^{n} \frac{Q q_{(j)}}{4\pi\epsilon_0 r^2_{(j)}} \hat{r}^i_{(j)} = Q \sum_{j=1}^{n} \frac{q_{(j)}}{4\pi\epsilon_0 r^2_{(j)}} \hat{r}^i_{(j)} \tag{1.4}$$

The element inside the sum can be seen as the *field* generated by the single particle $q_{(j)}$, denoted as $E^i_{(j)}$. This field is the Electrostatic field. It's clear that then we can define a total field E^i by superposition of the single charge fields, and we can write, for a system of charges

$$F^i = Q \sum_{j=1}^{n} E^i_{(j)} = Q E^i \tag{1.5}$$

Then, in general, we can say

$$E^i = \frac{F^i}{Q} \tag{1.6}$$

For our system of n charges the previous calculation is pretty straightforward and we directly get

$$E^i = \sum_{i=1}^{n} \frac{q_{(i)}}{4\pi\epsilon_0} \frac{\hat{r}^i}{r^2} \tag{1.7}$$

The passage to continuous distributions of charge is straightforward. We define the following "transformations"

$$\begin{cases} q_i \longrightarrow \mathrm{d}q \\ \sum_i \longrightarrow \int \end{cases}$$

The electric field of such distribution is then

$$E^i = \frac{1}{4\pi\epsilon_0} \int \frac{\hat{r}^i}{r^2} \mathrm{d}q \tag{1.8}$$

In general, $\mathrm{d}q$ can be expressed mathematically with a charge density which can be linear, superficial or volumetric. I.e.

$$\mathrm{d}q \to \begin{cases} \lambda(\tilde{r}^i)\mathrm{d}l & \text{linear distribution} \\ \sigma(\tilde{r}^i)\mathrm{d}s & \text{superficial distribution} \\ \rho(\tilde{r}^i)\mathrm{d}^3\tilde{x} & \text{volumetric distribution} \end{cases} \tag{1.9}$$

The electric field will then be calculated with the integral (1.8) extended to the appropriate set (a curve, a surface or a volume)

§§ 1.1.1 Divergence of the Electrostatic Field

As we have defined previously the electric field it's clear that if the distribution is complicated enough the integrals might be hard to solve or straight up nonsolvable. We then want to find different ways for calculating the field.
In general a vector field is determined by both its divergence and its curl. We firstly remember the definition of the 3D Dirac delta function $\delta^3(r^i)$, which is simply

$$\delta^3(r^i) = \frac{1}{4\pi} \frac{\partial}{\partial x^i} \left(\frac{r^i}{r^2} \right) \tag{1.10}$$

We then take the definition of E^i for a continuous volumetric distribution and simply apply the divergence operator.

$$\partial_i E^i = \frac{1}{4\pi\epsilon_0} \frac{\partial}{\partial x^i} \iiint_V \rho(\tilde{r}^i) \frac{\hat{r}^i}{r^2} \mathrm{d}^3\tilde{x} \tag{1.11}$$

Noting that the integral is with respect to the primed coordinates (the ones with respect to the distribution) we can bring inside the divergence operator, and remembering that in this case $r = \left\| r^i - \tilde{r}^i \right\|$, with the definition of the 3D delta we get

$$\partial_i E^i = \frac{1}{4\pi\epsilon_0} \iiint_V \rho\left(\tilde{r}^i\right) \delta^3\left(r^i - \tilde{r}^i\right) \mathrm{d}^3\tilde{x} = \frac{1}{\epsilon_0} \rho\left(r^i\right)$$

Therefore, due to the generality of ρ we have that for *every* electrostatic field, the following equation holds

$$\partial_i E^i = \frac{\rho}{\epsilon_0} \tag{1.12}$$

This is Maxwell's first equation for the electrostatic field.
A really important property comes from this equation, *Gauss' law*. This law states that the flux of E^i is proportional to the total charge enclosed by the chosen volume V.
This is a direct consequence of Stokes' theorem for differential forms.
We choose a bounded volume $V \subset \mathbb{R}^3$ and integrate both sides of (1.12)

$$\iiint_V \partial_i E^i \mathrm{d}^3x = \oiint_{\partial V} E^i \hat{n}_i \mathrm{d}s = \frac{1}{\epsilon_0} \iiint_V \rho(r^i) \mathrm{d}^3x$$

Defining the flux of E^i as $\Phi_{\partial V}(E^i)$ we have, then

$$\oiint_{\partial V} E^i \hat{n}_i \mathrm{d}s = \Phi_{\partial V}(E^i) = \frac{Q_V}{\epsilon_0} \tag{1.13}$$

This is the mathematical expression of Gauss' law, where we have written

$$Q_V = \iiint_V \rho(r^i) \mathrm{d}^3x$$

Which is the total charge contained inside the volume V.
This theorem is *fundamental* for the solution of a myriad of electrostatic problems which would take a lot of calculations using (1.8). The main idea is that this can be used in conditions where there are particular symmetries of the system.

Example 1.1.1 (A charged sphere). Suppose that you have a charged sphere with radius R and total charge q and I want to know the electric field inside and outside the sphere. We begin by calculating the field outside using Gauss' law. Due to the radial symmetry of the problem we have that $\hat{n}^i = \hat{r}^i$ and therefore $E^i = E\hat{n}^i$ when we choose a spherical volume.
Let $\partial V = S_r^2$ be our "*gaussian surface*", a sphere of radius r, where the previous relation for E^i holds. We have that for any r

$$\Phi_{S_r^2}\left(E^i\right) = \oiint_{S_r^2} E^i \hat{n}_i \mathrm{d}s = E \oiint_{S_r^2} \mathrm{d}s = 4\pi r^2 E \tag{1.14}$$

The first part on the left of (1.13) is already evaluated. Then we need to calculate only the right side. Noting that there is no charge outside the sphere we have an internal volumetric density of charge $\rho = q/V$. Since V is a sphere we already know its volume, and the calculation it's quite easy

$$\iiint_{V_r} \rho(r^i) \mathrm{d}^3x = \frac{q}{V} \iiint_{V_r} \mathrm{d}^3x = \begin{cases} q\dfrac{V_r}{V} & r < R \\ q & r > R \end{cases} \tag{1.15}$$

Where V_r is the volume contained inside the gaussian sphere S_r^2. Remembering that $V_r = \frac{4}{3}\pi r^3$ and $V = \frac{4}{3}\pi R^3$ we have that

$$4\pi r^2 E = \begin{cases} \dfrac{q}{\epsilon_0}\left(\dfrac{r}{R}\right)^3 & r < R \\ \dfrac{q}{\epsilon_0} & r > R \end{cases} \tag{1.16}$$

Dividing by $4\pi r^2$ and remembering that $E^i = E\hat{n}^i = E\hat{r}^i$ we get the final solution for E^i, both inside and outside the charged sphere

$$E^i = \begin{cases} \dfrac{q}{4\pi\epsilon_0 R^3} r\hat{r}^i & r < R \\ \dfrac{q}{4\pi\epsilon_0 r^2}\hat{r}^i & r > R \end{cases} \tag{1.17}$$

Note how for $r < R$ the field grows linearly (we're adding charge increasing r), and it begins again falling like r^{-2} after we surpass the surface of the sphere at $r = R$. Curiously (but not at random) the field for a charged sphere with constant charge q is identical to the field produced by a point charge at the origin, it's like after we surpassed the surface of the sphere it collapsed all on the origin of the coordinates and became a point charge q at the origin.

The previous statements can be reformulated as a formal method

Method 1 (Gaussian Surfaces)**.** Given an electrostatic system with either spherical, cylindrical or planar symmetries. In order to solve (1.12) we need to choose an appropriate Gaussian surface G which encloses a bounded volume V for which $E^i \propto \hat{n}^i_G$. In this special case, integrating the equation (1.12) and applying Stokes' theorem we have

$$\oiint_G E^i \hat{n}^G_i \mathrm{d}s = E \oiint_G \mathrm{d}s = ES_G \tag{1.18}$$

Where S_G is the surface area of the gaussian surface. With this trick, if we call V the bounded volume such that $\partial V = G$ we have

$$E = \frac{1}{\epsilon_0 S_G} \iiint_V \rho(r^i)\mathrm{d}^3x$$

A rule of thumb for choosing G is the following:

- For spherical symmetry of E (like a point charge or a spherical distribution) G is the sphere of radius r
- For cylindrical symmetry (like a charged cable or a charged cylinder) G is the cylindrical surface of radius r
- for planar symmetry (like a charged plane) G is a "pillbox", i.e. simply a 3D rectangle

§§ 1.1.2 The Scalar Potential

A neat definition we can use is defining the *scalar potential* $V(r^i)$ of the electrostatic field. As usual a potential for a vector field is defined if and only if all closed path integrals of the field are 0 in a simply connected domain, i.e. that the curl of the field is zero in the selected domain.
We of course can choose this proof but it's much easier using this trick.
Take V as a bounded domain of $\mathbb{R}^3$ where there is some charge distribution $\rho(r^i)$ inside. The general formula for the electric field then is the following

$$E^i = \frac{1}{4\pi\epsilon_0} \iiint_V \rho(\tilde{r}^i)\frac{\hat{r}^i}{r^2}\mathrm{d}^3\tilde{x} \tag{1.19}$$

We immediately see that

$$\frac{\partial}{\partial x^i}\left(\frac{1}{r}\right) = -\frac{\hat{r}^i}{r^2}$$

Therefore, noting that the derivation acts only on the unprimed coordinates (i.e. it can go outside the integration without problems) we have

$$E^i = -\frac{\partial}{\partial x_i}\left(\frac{1}{4\pi\epsilon_0}\iiint_V \frac{\rho(\tilde{r}^i)}{r}\mathrm{d}^3\tilde{x}\right) \tag{1.20}$$

By definition of potential then, we can say that $E^i = -\partial^i V(r^i)$, where

$$V(r^i) = \frac{1}{4\pi\epsilon_0}\iiint_V \frac{\rho(\tilde{r}^i)}{r}\mathrm{d}^3\tilde{x} \tag{1.21}$$

This is known as the *scalar potential* of the electrostatic field.
Since by definition the curl of the gradient is always zero, we can immediately write a second constitutive equation for E^i

$$\epsilon^i_{jk}\partial^j E^k = 0 \tag{1.22}$$

This equation is Maxwell's third equation for static fields.
Defining $\mathbb{R}^\star = \mathbb{R}\cup\{\pm\infty\}$ and chosen two points $a, b \in \mathbb{R}^\star$, we have in the language of differential forms

$$\mathrm{d}V = E_i \mathrm{d}x^i \tag{1.23}$$

Therefore, with this definition, we can evaluate the work needed to move a charged particle through some path $\gamma : [a,b] \subset \mathbb{R}^\star \to \mathbb{R}^3$. We have

$$W = \int_\gamma F^i \hat{t}_i \mathrm{d}l = q\int_\gamma E^i \hat{t}_i \mathrm{d}l = -q\int_\gamma \partial^i V \hat{t}_i \mathrm{d}l$$

Writing $\hat{t}^i \mathrm{d}l = \mathrm{d}x^i$ we have

$$W = -q\int_\gamma \partial_i V \mathrm{d}x^i = -q\int_{V(a)}^{V(b)} \mathrm{d}V = q\Delta V \tag{1.24}$$

Therefore, $qV(r^i)$ can be imagined as a "potential energy" of the system. Via this definition, we have that the scalar potential has the following units in the SI system

$$[V] = \frac{[W]}{[q]} = \frac{\mathrm{J}}{\mathrm{C}} = \mathrm{V} \tag{1.25}$$

Where V are *Volts*. With this definition

$$1\ \mathrm{V} = 1\ \frac{\mathrm{J}}{\mathrm{C}} \tag{1.26}$$

From the definition of work we can immediately find a nice trick for evaluating the scalar potential of a distribution. Isolating the last two equalities in the first definition of work for the electric field we have

$$\int_\gamma E_i \mathrm{d}x^i = -\int_{V(a)}^{V(b)} \mathrm{d}V \tag{1.27}$$

Using the path independence of V we have by direct integration

$$V(b) - V(a) = -\int_\gamma E_i \mathrm{d}x_i \tag{1.28}$$

Due to the definition of V we know that it's defined up to a constant, and such constant can be chosen in order to have $V(a) = 0$. The point $a \in \mathbb{R}^\star$ is known as the *reference point* for the potential, and the appropriate choice depends from the charge distribution. The best choice is taking the point where the potential is 0
Suppose now we want to calculate the potential of a point charge in the origin. Since $E^i \to 0$ for $r \to \infty$ we take $a = \infty$, and therefore, since $V(r^i) \to 0$ for $r \to \infty$ we have at some distance $b = r$

$$V(r) = -\frac{q}{4\pi\epsilon_0}\int_\infty^r \frac{1}{r^2}\hat{r}^i \mathrm{d}x^i = -\frac{q}{4\pi\epsilon_0}\int_\infty^r \frac{1}{r^2}\mathrm{d}r = \frac{q}{4\pi\epsilon_0}\frac{1}{r} \tag{1.29}$$

Note that by linearity of the integral, for a system of point charges we have

$$V(r) = \frac{1}{4\pi\epsilon_0}\sum_{i=1}^n \frac{q_i}{r_i} \tag{1.30}$$

Note that this trick doesn't work if the charge extends to infinity since the integral would diverge, in that case the reference point will be some other $a = r_0$

§§ 1.1.3 Maxwell Equations for Electrostatics and Boundary Conditions

So far we found two main equations for the E^i field, these are two coupled partial differential equations known as the *Maxwell equations for Electrostatics*. These equations are

$$\begin{cases} \partial_i E^i = \dfrac{\rho}{\epsilon_0} \\ \epsilon^i_{jk}\partial^j E^k = 0 \end{cases} \tag{1.31}$$

Or, in integral form for a bounded volume V and a regular surface Σ

$$\begin{cases} \oiint_{\partial V} E^i \hat{n}_i \mathrm{d}s = \dfrac{1}{\epsilon_0}\iiint_V \rho \mathrm{d}^3x \\ \oint_{\partial\Sigma} E^i \hat{t}_i \mathrm{d}l = 0 \end{cases} \tag{1.32}$$

Inserting the definition of the potential these two equations collapse in a single equation, which is the *Poisson equation* for the potential

$$\partial_i\partial^i V = -\frac{\rho}{\epsilon_0} \tag{1.33}$$

But, as for every partial differential equation, these make sense if and only if a boundary condition has been provided.
Without loss of generality we can consider an uniformly charged plane with surface density σ. We have using the Gaussian surface trick, choosing a pillbox with surface area A, that

$$E\oiint_G \mathrm{d}s = \frac{\sigma A}{\epsilon_0}$$

Noting that the contribute between the 4 sides is zero, only the two faces remain and $S_G = 2A$, and therefore

$$E = \frac{\sigma}{2\epsilon_0} \tag{1.34}$$

Since $E^i \propto \hat{n}^i$ we have

$$E^i = \frac{\sigma}{2\epsilon_0}\hat{n}^i \tag{1.35}$$

But the normal to the plane changes sign passing through its surface, therefore the field is discontinuous passing through its surface!
For the potential this is not true. By definition of potential we're checking the line integral along the tangent to the border of this Gaussian surface, which doesn't change sign when we pass through the surface. Therefore we have that $V \in C^2(V) \cup C^0(\partial V)$ while the field is discontinuous on the border. These conditions are valid for every regular surface.
Consider that, locally, every regular surface can be considered as "flat" or euclidean, therefore the Gaussian pillbox trick works well.
Noting that the outward normal of the pillbox above and below the "plane" is equal to $\pm\hat{n}^i$ where $\hat{n}^i$ is the normal to this plane. Therefore, by the previous calculations we must have that passing through the surface (locally)

$$\left(E^i_{in} + E^i_{out}\right)\hat{n}_i = 0 \tag{1.36}$$

I.e., the field outside this "plane" is opposite in sign to the field inside the "plane". Going back to the main general surface, via integration, we have that this result must hold generally, which emphasizes the discontinuity of the electric field.

§§ 1.1.4 Energy of the Electrostatic Field

Considering again the definition of work for a particle as $W = q\Delta V$ we can calculate it for a set of particles. Considering the interaction between particles we have that $W \propto q_i q_j$ where $i, j = 1, \cdots, n$, and noting that a charge doesn't self interact, i.e. $q_i q_j = 0$ for $i = j$ and that the usual multiplication between scalar is commutative, i.e. $q_i q_j = q_j q_i$ we have

$$W = \frac{1}{8\pi\epsilon_0}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{q_i q_j}{r_{ij}^2} = \frac{1}{2}\sum_{i=1}^{n} q_i V(r_i) \tag{1.37}$$

Passing to continuous distributions we get

$$W = \frac{1}{2}\iiint_V \rho(r^i)V(r^i)\mathrm{d}^3x$$

Using the first Maxwell equation we have $\rho = \epsilon_0\partial_i E^i$, therefore

$$W = \frac{\epsilon_0}{2}\iiint_V V(r^i)\partial_i E^i \mathrm{d}^3x$$

Integrating by parts and applying Stokes' theorem we get

$$W = \frac{\epsilon_0}{2}\left(\oiint_{\partial V} VE^i \mathrm{d}s - \iiint_V E^i\partial_i V \mathrm{d}^3x\right) \tag{1.38}$$

Noting that E^i extends to infinity where it becomes zero, we have on the limit $V \to \mathbb{R}^3$, where we use $E^i = -\partial^i V$, that the total energy stored in a charge distribution is

$$W = \frac{\epsilon_0}{2}\iiint_{\mathbb{R}^3} E^2 \mathrm{d}^3x \tag{1.39}$$

§ 1.2 Conductors

The main real problem that somebody will encounter solving problems in electrostatics is problems with *conductors*. A conductor is a rigid body for which there are free charges which can move after the application of an electric field. An example of conductor is a metallic body in the rigid body approximation.
One main property of conductors is that inside of it the electric field is zero.
Imagine taking a neutral box conductor, and then apply an electric field parallel to the sides of the box. The free charges will then move due to the action of the electrostatic force towards the field (if $q > 0$) or against the field (if $q < 0$). Since the conductor was neutral and charges must be conserved since they cannot pop into existence randomly, we have that the field generated by the single negative and positive charges on the surface of the conductor will be equal in magnitude and opposite in sign, therefore the total field inside is 0, even though the field outside is nonzero.
A second property of conductors is that the charge density inside the conductor is 0 inside. Using Gauss law and the first property of conductors we have

$$\rho = \epsilon_0 \partial_i E^i = 0 \tag{1.40}$$

This is always true for conductors, since as we said before $E^i = 0$ inside.
One main explanation of this is that inside there is as much positive charge density ρ_+ and ρ_-. In fact, from Gauss' law we have

$$\rho = \rho_+ + \rho_- = 0 \implies \rho_+ = \rho_-$$

This indicates that the charges of the conductor will then be all on the surface, and therefore the conductor is an equipotential surface. In fact

$$\partial^i V_{in} = -E^i_{in} = 0 \implies V_{in} = k, \qquad k \in \mathbb{R} \tag{1.41}$$

In order to bring out other properties of the electric field in presence of conductors, we can consider the surface of separation between two materials. Consider a rectangular loop going through both materials. We have from the third Maxwell equation for electrostatics

$$\oint_A E^i \hat{t}_i \mathrm{d}l = 0 \tag{1.42}$$

Since the conductor is rectangular, separating the line integral into 4 integrals, where 2 go parallel to the surface and 2 are normal to it, we have that the two normal integrals taking a clockwise path must cancel each other and therefore we have

$$\oint_1 E^i_1 \hat{t}^1_i \mathrm{d}l + \oint_2 E^i_2 \hat{t}^2_i \mathrm{d}l = 0 \tag{1.43}$$

Since $\hat{t}^i_1 = -\hat{t}^i_2$ we have that, locally

$$\left(E^i_1 - E^i_2\right) \hat{t}^1_i = 0 \tag{1.44}$$

Therefore, the electric field tangent to the surface is continuous and therefore conserved. Since a charged conductor has a zero electric field inside and there is no external field, then

$$E^i_{in} = 0, \quad E^i_{in} \hat{t}_i = 0, \quad E^i_{ext} \hat{t}_i = 0$$

But since in general a vector can always be decomposed in a tangent component to the surface and a normal component to the surface we have

$$E^i_{in}\hat{n}_i = 0, \quad E^i_{ext}\hat{n}_i \neq 0$$

This because the conductor is charged. This means that there is a discontinuity in the field and the field itself must be normal to the surface of the conductor due to the continuity of the tangential component of the field. If E^i_c is the electric field generated by a conductor we have then

$$E^i_c = E\hat{n}^i \tag{1.45}$$

Consider now the potential inside and outside the conductor, V_{ext}, V_{in}. Considering that the charges we are moving are electrons with $q = -e$, where e is the fundamental charge

$$e = 1.6021766208(98) \cdot 10^{-19}\ \mathrm{C} \tag{1.46}$$

We have that the work needed to bring outside the conductor our electron will be

$$\Delta U = -e\Delta V = -e\left(V_{ext} - V_{in}\right) \tag{1.47}$$

We define the work function as $L = \Delta U/e$ and it must obviously be positive since we're applying energy to the system in order to bring out an electron. We have

$$L = V_{in} - V_{ext} > 0 \implies V_{in} > V_{ext} \tag{1.48}$$

Due to all of these consideration, and noting that $\mathrm{d}V = -E_i \mathrm{d}x_i$ we have that the potential of a conductor will be defined as

$$V_0(r) = -\int_r^\infty E^i\hat{t}_i \mathrm{d}l = -\int_r^{r_0} E^i\hat{t}_i \mathrm{d}l \tag{1.49}$$

Where r_0 is the "first" radius immediately outside the conductor.

§§ 1.2.1 Coulomb Theorem

Consider now a conductor V and take a small cylinder orthogonal to its surface. Considering that the charge on a conductor is only on the surface we have using Gauss' law on the differential flux of E^i that

$$\mathrm{d}\Phi(E^i) = E^i\hat{n}_i \mathrm{d}S = \frac{\sigma}{\epsilon_0}\mathrm{d}S \tag{1.50}$$

Considering the equality in terms of norms of the E^i field and remembering that $E^{\|}\hat{n}^i$ we have that

$$E^i = \frac{\sigma}{\epsilon_0}\hat{n} \tag{1.51}$$

You can immediately see that this field is twice the field generated by a charged infinite plane. Let's consider what's happening with some more precision.
In that small cylinder $\mathrm{d}S$ we will have that the total external field will be composed from the contribution

of the charge inside the cylinder and the one outside. The same should be for the inside, but the inside field *must* be zero

$$E^i_{ext} = \left(E^i_{ext}\right)^{\mathrm{d}S} + \left(E^i_{ext}\right)^{S-\mathrm{d}S} \neq 0$$
$$E^i_{in} = \left(E^i_{in}\right)^{\mathrm{d}S} + \left(E^i_{in}\right)^{S-\mathrm{d}S} = 0$$

The field $\left(E^i\right)^{S-\mathrm{d}S}$ doesn't change and it must be the same as the field generated by $\mathrm{d}S$. Applying Gauss' theorem to the small surface element $\mathrm{d}S$ and noting that it must be the same of a plane with surface area $\mathrm{d}S$ we have

$$\left(E^i_{in}\right)^{S-\mathrm{d}S} = \left(E^i_{ext}\right)^{S-\mathrm{d}S} = -\left(E^i_{in}\right)^{\mathrm{d}S} = \frac{\sigma}{2\epsilon_0}\hat{n}^i$$

Therefore, finally

$$E^i_{ext} = \frac{\sigma}{2\epsilon_0}\hat{n} + \frac{\sigma}{2\epsilon_0}\hat{n} = \frac{\sigma}{\epsilon_0}\hat{n}$$

Where we used again that $\left(E^i_{ext}\right)^{\mathrm{d}S} = \frac{\sigma}{2\epsilon_0}\hat{n}$

§§ 1.2.2 Induced Charges

Consider now some conductor which is empty inside. Inside the first conductor we insert another conductor charged with charge Q. At $t = 0$ the external conductor is neutral, and therefore $Q_{ext} = 0$. Since charge must be conserved, we have that at $t > 0$ when we insert the new conductor inside the total charge must remain neutral, therefore

$$Q_{int} + Q_{ext} = 0$$

From Gauss' theorem, taking a surface inside the conductor that includes inside itself the internal surface of the conductor but not the external one. For Gauss we have

$$\Phi(E^i) = 0 = \frac{Q_V}{\epsilon_0} \implies Q + Q_{in} = 0$$

Therefore, there must be an *induced charge* Q_{in} on the internal surface of the conductor, such that

$$Q_{in} = -Q \tag{1.52}$$

From this, substituting before, we have that on the external surface we measure the charge we added inside the conductor, via the process of charge induction

$$Q_{ext} = -Q_{in} = Q \tag{1.53}$$

Note that this comes directly for having charge conservation.
Consider now the same empty conductor but don't add any charge inside of it, but rather charge the whole conductor with some positive charge Q. What happens inside the hole? Is there any charge?
By Gauss' theorem we have, since $E^i = 0$ inside the conductor, that the total charge inside the conductor is zero $Q_{in} = 0$.
There could still be a charge balance inside, where $Q^+_{in} - Q^-_{in} = Q_{in} = 0$. Supposing this true we can take a closed path that goes inside the hole. By definition of E^i the line integral on this path γ must be

zero. Divide the path into 1, that goes inside the hole, where there should be a field E^i between the two charges Q^+_{in} and Q^-_{in}, and path 2 which is inside the conductor. Then we would have

$$\oint_\gamma E^i \hat{t}_i \mathrm{d}l = \int_1 E^i \hat{t}^1_i \mathrm{d}l + \int_2 E^i \hat{t}^2_i \mathrm{d}l = \int_1 E^i \hat{t}^1_i \mathrm{d}l$$

Since the path is closed, call D the surface enclosed by the path, we have

$$\oint_\gamma E^i \hat{t}_i \mathrm{d}l = \oiint_D \epsilon^i_{jk} \partial^j E^k \hat{n}_i \mathrm{d}s = \int_1 E^i \hat{t}^1_i \mathrm{d}l \implies \epsilon^i_{jk} \partial^j E^k \neq 0 \tag{1.54}$$

This is in clear contradiction with Maxwell's equation for electrostatics (which we have already demonstrated that they generally hold), therefore all the charge is safely distributed on the *external* surface of the conductor, as we expected.

Exercise 1.2.1 (Two Charged Spheres)**.** Suppose that you have two metal spheres connected by a wire. One has radius R_1 and the other has radius R_2. At $t > 0$ we deposit some charge Q on the system. What will be the total charge distributed on the two spheres? (Q_1, Q_2)

S o l u t i o n

The potentials on the two spheres must be equal, and we know already from previous calculations that

$$\begin{aligned} V_1 &= \frac{1}{4\pi\epsilon_0}\frac{Q_1}{R_1} \\ V_2 &= \frac{1}{4\pi\epsilon_0}\frac{Q_2}{R_2} \\ V_1 &= V_2 \end{aligned} \tag{1.55}$$

From the previous equation we have that

$$Q_2 = \frac{R_2}{R_1} Q_1 \tag{1.56}$$

The total charge, on the other hand, will be $Q = Q_1 + Q_2$, therefore

$$Q = Q_1 + \frac{R_2}{R_1} Q_1 = \frac{R_1 + R_2}{R_1} Q_1 \implies Q_1 = \frac{R_1}{R_1 + R_2} Q \tag{1.57}$$

And, analogously

$$Q_2 = \frac{R_2}{R_1 + R_2} Q \tag{1.58}$$

From Gauss' theorem, if the spheres have surface charges σ_i, $i = 1, 2$ we must also have

$$\frac{Q_1}{R_1} = \frac{Q_2}{R_2} \implies \frac{4\pi R_1^2 \sigma_1}{R_1} = \frac{4\pi R_2^2 \sigma_2}{R_2} \tag{1.59}$$

I.e.

$$\sigma_1 R_1 = \sigma_2 R_2 \implies \sigma_1 = \frac{R_2}{R_1}\sigma_2$$

Since $\frac{R_1\sigma_1}{\epsilon_0} = \frac{R_2\sigma_2}{\epsilon_0}$, we must also have that the fluxes of the fields multiplied by R_i are equal, i.e. the electric fields are scaled as follows

$$E_2 = \frac{R_1}{R_2}E_1 \tag{1.60}$$

□

§§ 1.2.3 Capacity

Consider an isolated conductor on which there is some charge Q, distributed with density σ on its surface, such that the conductor is equipotential. We have that for every point in the conductor, by definition

$$\begin{aligned} V(r) &= \frac{1}{4\pi\epsilon_0}\iint_S \frac{\sigma}{r}\mathrm{d}s \\ Q &= \iint_S \sigma \mathrm{d}s \end{aligned} \tag{1.61}$$

It's clear that by this definition that if we vary σ to a new $\sigma' = \alpha\sigma$ with $\alpha \in \mathbb{R}$, we also have that $V' = \alpha V,\ Q' = \alpha Q$.
The following rate is then called the *capacity* of the conductor

$$C = \frac{Q}{V} \tag{1.62}$$

This is clearly only dependent on the geometry of the system. The capacity is measured in *Farads*, where

$$1\ \mathrm{F} = 1\ \frac{\mathrm{C}}{\mathrm{V}}$$

Example 1.2.1 (Capacity of a Spherical Conductor). Take now a spherical conductor with charge Q. We have

$$V = \frac{Q}{4\pi\epsilon_0 R}$$

Therefore

$$C = 4\pi\epsilon_0 R \tag{1.63}$$

This lets us redefine ϵ_0 in terms of Farads. In fact

$$[\epsilon_0] = \frac{[C]}{[R]} = \frac{\mathrm{F}}{\mathrm{m}}$$

Therefore

$$\epsilon_0 = 8.854\ \frac{\mathrm{F}}{\mathrm{m}} \tag{1.64}$$

In the case that we have multiple conductors one close to the other the problem gets slightly more complex.
Add a charge Q_1 to the first conductor, which will have potential V_1, which will induce a charge Q_2 and therefore a potential V_2 on the second. If I change the charge to $Q_1' = \alpha Q_1$ we will have a basically

identical result to the previous problem. Inverting the system and setting the charge on the second conductor Q_2 we will have a symmetrical system, for which we can write

$$V_i = \sum_{j=1}^{n} p_{ij} Q_j \tag{1.65}$$

The p_{ij} are the potential coefficients, for which holds $p_{ij} = p_{ji} > 0$, $p_{ii} \geq p_{ij}\ i \neq j$.
Due to the fact that the potential is unequivocally determined we must be able to solve the inverse problem, therefore we also know that $\det p_{ij} \neq 0$, and therefore

$$Q_i = \sum_{j=1}^{n} c_{ij} V_j \tag{1.66}$$

The matrix c_{ij} is known as the capacitance matrix, and we have $p_{ij} = c_{ij}^{-1}$. The diagonal elements c_{ii} are known as the *capacity coefficients*, while the off diagonal $c_{ij},\ i \neq j$ are known as the induction coefficients.
For this matrix hold the following properties, known as *Maxwell inequalities*

$$\begin{cases} c_{ij} = c_{ji} \\ c_{ii} > 0 \\ c_{ij} < 0 \quad i \neq j \\ \sum_{j=1}^{n} c_{ij} \geq 0 \end{cases} \tag{1.67}$$

§§ 1.2.4 Capacitors

Let's take again two conductors in total induction as for our previous system of two concentric conductors where one inside is set at a charge Q. Grounding the external surface we get that the external shell will be at a fixed $V = 0$, while the internal surface will have an induced charge $-Q$. Between these two surfaces there will be a potential difference ΔV, for which it's possible to evaluate the capacitance as

$$C = \frac{Q}{\Delta V}$$

Writing this in terms of the potential matrix $V_i = \sum_j p_{ij} Q_j$ we have the following system of equations, where $Q_1 = Q,\ Q_2 = -Q$

$$\begin{cases} V_1 = p_{11} Q - p_{12} Q \\ V_2 = p_{21} Q - p_{22} Q \end{cases}$$

Subtracting the second from the first we have

$$\Delta V = (p_{11} + p_{12} - 2p_{12})\, Q$$

Therefore

$$C = \frac{Q}{\Delta V} = \frac{1}{p_{11} + p_{12} - 2p_{21}} \tag{1.68}$$

Or in terms of the capacitance matrix c_{ij}

$$C = \frac{\det(c_{ij})}{c_{11} + c_{12} - 2c_{12}} \tag{1.69}$$

Finding the capacitance using these matrices tho is a quite long calculations, therefore we directly use the line integral of the E^i field for determining it, therefore, since

$$\Delta V_{12} = \int_2^1 E_i \mathrm{d}x^i$$

Example 1.2.2 (Spherical Capacitor). Consider now a spherical capacitor for which the outer shell is grounded, we have

$$E^i = \frac{Q}{4\pi\epsilon_0}\frac{\hat{r}^i}{r^2}$$

Therefore

$$\Delta V = \frac{Q}{4\pi\epsilon_0}\int_{r_2}^{r_1}\frac{1}{r^2}\mathrm{d}r = \frac{Q}{4\pi\epsilon_0}\left(\frac{1}{r_1} - \frac{1}{r_2}\right)$$

Therefore

$$C_s = \frac{4\pi\epsilon_0 r_1 r_2}{r_2 - r_1} \tag{1.70}$$

Example 1.2.3 (Cylindrical Capacitor). For a cylindrical capacitor made of two conducting cylinders of radius R_1 and R_2 and length $l >> R_2$ and total charge λl we have that the electric field is

$$2\pi l r E = \frac{\lambda l}{\epsilon_0}$$
$$E^i = \frac{\lambda}{2\pi\epsilon_0}\frac{\hat{r}}{r}$$

Therefore

$$\Delta V = \int_1^2 E_i \mathrm{d}x^i = \frac{\lambda}{2\pi\epsilon_0}\log\left(\frac{R_2}{R_1}\right)$$

Therefore, since $Q = \lambda l$, we have

$$C_c = \frac{2\pi\epsilon_0}{\log\left(\frac{R_2}{R_1}\right)} \tag{1.71}$$

Example 1.2.4 (Parallel Plane Capacitor). For two parallel plane conductors for which $d << \sqrt{S}$ where S is the surface area of the plane we have that

$$E = \frac{\sigma}{\epsilon_0},\ Q = S\sigma$$

Therefore

$$\Delta V = \frac{\sigma}{\epsilon_0}d$$

Where d is the distance between the plates, and therefore

$$C = \frac{\epsilon_0 S}{d} \tag{1.72}$$

§§ 1.2.5 Forces on a Conductor

Consider a charged conductor with surface area S. Considering a small element $\mathrm{d}S$ we have that the external field generated by the remaining surface is

$$\left(E^i_{ext}\right)^{S-\mathrm{d}S} = \frac{\sigma}{2\epsilon_0}\hat{n}$$

The total charge in $\mathrm{d}S$ is $\sigma \mathrm{d}S$, and therefore the (infinitesimal) force on the area element $\mathrm{d}S$ is, by definition of electrostatic force

$$\mathrm{d}F^i = \sigma \left(E^i\right)^{S-\mathrm{d}S} \mathrm{d}S = \frac{\sigma^2}{2\epsilon_0}\hat{n}^i \mathrm{d}S = \frac{1}{2}\epsilon_0 E^2 \hat{n}^i \mathrm{d}S = u\hat{n}\mathrm{d}S \tag{1.73}$$

Where we identified the energy density of the field u as

$$u = \frac{1}{2}\epsilon_0 E^2$$

Deriving everything by $\mathrm{d}S$, we have that the *electrostatic pressure* p^i on the infinitesimal element of the surface of the conductor is

$$p^i = \frac{\mathrm{d}F^i}{\mathrm{d}S} = u\hat{n}^i \tag{1.74}$$

Consider now a virtual displacement of the external surface of the conductor, where we move it by δr orthogonally to the previous surface, then, the (virtual) work necessary for such displacement is

$$\delta L = \delta F^i_{ext} \delta r_i = \delta U$$

Where we used that $\delta L = \delta U$, and F^i_{ext} as the "extraction force". Since $F^i_{ext} = -F^i$ we have that

$$\delta F_r = -\frac{\delta U}{\delta r}$$

But, by definition

$$\delta U = -\frac{1}{2}\epsilon_0 E^2 \delta r \mathrm{d}S$$

Therefore, as before

$$\delta F_r = u\mathrm{d}S$$

For constant charge, we might think to apply this to a charged parallel plate capacitor, for which we know that the infinitesimal work needed to charge it, i.e. to move the charges from infinity towards our capacitor, is

$$\mathrm{d}W = V\mathrm{d}q = \frac{q}{C}\mathrm{d}q \implies W = \frac{1}{2}\frac{Q^2}{C}$$

For a parallel plate capacitor therefore

$$U(x) = \frac{1}{2}\frac{Q^2 x}{\epsilon_0 S}$$

Therefore

$$F = -\frac{\partial U}{\partial x} = -\frac{1}{2}\frac{Q^2}{\epsilon_0 S}$$

This force is attractive (obvious from the system).
What if $V = cost.$ but the charge isn't constant? We know that

$$V = \frac{Q}{C}$$

And since both C, V are constants (one depends only on the geometry and the other is set constant by the system) Q can be the only one to have changed.
This means that there is some generator that charges up the capacitor, with work

$$\delta W_g = V\delta Q = V^2\delta C$$

Where we have $Q = VC$. From our previous relations we have

$$\delta W_{ext} = F_{ext}\delta x, \qquad \delta U = \delta W_g + \delta W_{ext}$$

Therefore, since $F_{ext} = -F$ we have that

$$\delta U = \delta W_g - \delta W$$

And, for the generator

$$\delta(CV^2) = \delta W + \delta\left(\frac{1}{2}CV^2\right)$$

Finally

$$\delta W = \delta U = F\delta x$$

Remembering that $U = CV^2/2$ and $C = S\epsilon_0/x$ we have through derivation that

$$F = -\frac{1}{2}\frac{S\epsilon_0 V^2}{x^2} = -\frac{1}{2}\frac{C^2V^2}{S\epsilon_0} = -\frac{1}{2}\frac{Q^2}{S\epsilon_0}$$

Which is the same result as before.
It's clear that for a charged conductor then the force is the mechanical moment of the system. It can be derived using the virtual work theorem, noting that $\delta L_{ext} = \delta U = -\delta L$, therefore

$$\delta L = F^i\delta x_i + L_i\delta\theta^i = -\delta U$$

Where, in the limit $\delta x, \delta\theta \to \mathrm{d}x, \mathrm{d}\theta$

$$F_x = -\frac{\partial U}{\partial x}$$
$$L_\theta = -\frac{\partial U}{\partial \theta}$$

2 The Electrostatic Potential

§ 2.1 Poisson and Laplace Equations

§§ 2.1.1 Green Identities

From the equations of Maxwell for electrostatics, we have seen that inserting the relation between the electrostatic field and the potential we get a second order partial differential equation known as the *Poisson equation*

$$\nabla^2 V = \partial^i \partial_i V(r^i) - \frac{\rho}{\epsilon_0} \tag{2.1}$$

And its homogeneous counterpart where $\rho = 0$, the *Laplace equation*

$$\partial^i \partial_i V = 0 \tag{2.2}$$

There are two fundamental theorems that we're gonna use for solving PDEs (Partial differential equations).

Theorem 2.1 (First Green Identity). *Given two functions $\varphi, \psi \in C^2(V)$ with V being a bounded set, we have*

$$\iiint_V \left(\varphi \partial^i \partial_i \psi + \partial^i \varphi \partial_i \psi\right) \mathrm{d}^3 x = \oiint_{\partial V} \varphi \frac{\partial \psi}{\partial x^i} \hat{n}^i \mathrm{d}s = \oiint_{\partial V} \varphi \frac{\partial \psi}{\partial n} \mathrm{d}s \tag{2.3}$$

Proof. Taken $A_i = \varphi \partial_i \psi$ we have that

$$\partial^i A_i = \partial^i \left(\varphi \partial_i \psi\right) = \varphi \partial^i \partial_i \psi + \partial^i \varphi \partial_i \psi$$

Therefore

$$\iiint_V \partial^i A_i \mathrm{d}^3 x = \iiint_V \left(\varphi \partial^i \partial_i \psi + \partial^i \varphi \partial_i \psi\right) \mathrm{d}^3 x = \oiint_{\partial V} \varphi \frac{\partial \psi}{\partial n} \mathrm{d}s = \oiint_{\partial V} A_i \hat{n}^i \mathrm{d}s$$

□

Theorem 2.2 (Second Green Identity). *Given two functions $\varphi, \psi \in C^2\left(\mathbb{R}^3\right)$, again from stokes theorem one has*

$$\iiint_V \left(\varphi \partial^i \partial_i \psi - \psi \partial^i \partial_i \varphi\right) \mathrm{d}^3 x = \oiint_{\partial V} \left(\varphi \frac{\partial \psi}{\partial n} - \psi \frac{\partial \varphi}{\partial n}\right) \mathrm{d}s \tag{2.4}$$

Proof. Taken two vector fields $A_i = \varphi\partial_i\psi$, $B_i = \psi\partial_i\varphi$, analogously as before we have

$$\partial^i\left(A_i - B_i\right) = \varphi\partial^i\partial_i\psi - \psi\partial^i\partial_i\varphi$$

Applying Stokes' theorem to the previous definition we have the proof, since

$$A_i - B_i = \varphi\partial_i\psi - \psi\partial_i\varphi$$

□

With these two theorems, we can easily modify Poisson's equation into an integral equation which can help us find useful informations on the shape of V.
Taking (2.4) where we set

$$V(r^i) = \varphi$$
$$\frac{1}{r} = \psi$$

We get, remembering that $\partial^i\partial_i(r^{-1}) = -4\pi\delta^3(r^i - \tilde{r}^i)$ and $\partial^i\partial_i V = -\rho/\epsilon_0$

$$\iiint_V \left(V(\tilde{r}^i)\partial^i\partial_i\left(\frac{1}{r}\right) - \frac{1}{r}\partial^i\partial_i V\right)\mathrm{d}^3x = \iiint_V \left(-4\pi V\delta^3(r^i) + \frac{\rho}{r\epsilon_0}\right)\mathrm{d}^3x$$

Therefore

$$\iiint_V \left(-4\pi V(\tilde{r}^i)\delta^3(r^i) + \frac{\rho}{r\epsilon_0}\right)\mathrm{d}^3x = \oiint_{\partial V}\left(V(r^i)\frac{\partial}{\partial n}\left(\frac{1}{r}\right) - \frac{1}{r}\frac{\partial V}{\partial n}\right)\mathrm{d}s$$

Bringing to the left the surface integral and solving for $V(r^i)$ after having applied the Dirac delta we have

$$V(r^i) = \frac{1}{4\pi\epsilon_0}\iiint_V \frac{\rho}{r}\mathrm{d}^3x - \frac{1}{4\pi}\oiint_{\partial V}\left(V(r^i)\frac{\partial}{\partial n}\left(\frac{1}{r}\right) - \frac{1}{r}\frac{\partial V}{\partial n}\right)\mathrm{d}s \tag{2.5}$$

This result, is the general solution for Poisson's equation, consistent with a known volume charge ρ and a surface charge $\sigma = \epsilon_0\partial_n V$. Note how the solution depends on the boundary values of V.

§§ 2.1.2 Boundary Conditions and Uniqueness of the Solution

Given the general solution (2.5) how can we choose for appropriate boundary values such that the solution exists and is unique $\forall r^i \in V$ where V is a bounded and closed set?
One way is to specify V at the boundary, i.e. using Dirichlet boundary conditions, or to specify $E_n = E^i\hat{n}_i = -\partial_n V$ in the boundary ∂V, i.e. using Neumann boundary conditions.
Supposing Dirichlet boundary conditions for V we have that the solution is unique. Why?
Let V be the usual bounded set of $\mathbb{R}^3$ in which we have

$$\partial^i\partial_i V = -\frac{\rho}{\epsilon_0} \qquad \forall x^i \in V$$

Then, let $U = V_1 - V_2$ where V_1, V_2 are two solutions to Poisson's equation. By definition, then, U solves Laplace's equation

$$\partial^i\partial_i U = \partial^i\partial_i V_1 - \partial^i\partial_i V_2 = 0$$

At the boundary therefore we must have

$$U, \qquad \frac{\partial U}{\partial n} = 0 \qquad \forall x^i \in \partial V$$

From Green's 1st identity we also have that

$$\iiint_V U\partial^i\partial_i U + \partial^i U\partial_i U \mathrm{d}^3x = \oiint_{\partial V} U\frac{\partial U}{\partial n}\mathrm{d}s$$

Using that U must solve Laplace's equation and it must also be zero at the boundary, we have, writing $\partial^i U\partial_i U = |\partial U|^2$

$$\iiint_V |\partial U|^2 \mathrm{d}^3x = 0$$

This last integral implies that $|\partial U|^2 = 0$ and therefore $\partial_i U = 0\ \forall x^i \in V$, i.e. U is constant. Since $U \in C^2$ and it must be 0 in ∂V the constant must be 0 and therefore

$$V_1 = V_2$$

Which implies $\exists! V : V \to \mathbb{R}$ which solves Poisson's equation where V is defined on the boundary. With Neumann conditions this implies that the two solutions are linearly dependent, still implying the uniqueness of the solution.
It's also clear that using mixed Dirichlet/Neumann boundary conditions will give rise to a well behaved and unique solution.

§§ 2.1.3 Method of Images

A cool method for finding a solution of the Poisson and Laplace equations is the *method of images*, where we choose some imaginary charges put in some special positions such that the potential found solves the PDE and therefore is unique.

Example 2.1.1 (A Toy Problem). Suppose that some point charge q is held at some distance d from a grounded infinite conducting plane put at $z = 0$. What is $V(r^i)$ above the plane where there is q?
Note that it cannot be $q/4\pi\epsilon_0 r$ since there is an induced charge on the surface of the plane where $Q_i = -q$.
We imagine removing the plane and setting a charge $-q$ on the opposite side of the first charge. In this case the potential will be simply the sum of the two potentials of the single charge, where

$$V(x,y,z) = \frac{q}{4\pi\epsilon_0}\left(\frac{1}{\sqrt{x^2+y^2+(z-d)^2}} - \frac{1}{\sqrt{x^2+y^2+(z+d)^2}}\right)$$

Note that this potential goes to 0 at infinity and it's 0 at $z = 0$ where there should be our plane. Due to the uniqueness of the solution we have that this is the solution to the first problem. We can also calculate the induced surface charge. We know that the surface charge will be proportional to the normal derivative of the potential at $z = 0$ (see the general solution of Poisson's equation), therefore, since the normal to the plane is the $\hat{z}^i$ versor, we have that

$$\frac{\partial V}{\partial n} = \frac{\partial V}{\partial z} = \frac{q}{4\pi\epsilon_0}\left(\frac{z+d}{(x^2+y^2+(z+d)^2)^{3/2}} - \frac{z-d}{(x^2+y^2+(z-d)^2)^{3/2}}\right)$$

Therefore, taking $z = 0$ and multiplying by $-\epsilon_0$ we have that the induced surface charge on the plane is:

$$\sigma(x, y) = -\frac{qd}{2\pi \left(x^2 + y^2 + d^2\right)^{3/2}}$$

Note that integrating σ over all the plane we get back that the total induced charge is $-q$ as expected.

The method of images is a particular method that uses the symmetries of the problem in order to carve out a solution to Poisson's equation and it can't be used in most situations. In those other situations we need to actually solve the partial differential equation and find the potential through integration, using a cool method that will be explained in the next section

§ 2.2 Separation of Variables

The main line of attack for Laplace's equation is the *separation of variables*, i.e. taking the following Ansatz for the potential $V(x, y, z)$

$$V(x, y, z) = f(x)g(y)h(z)$$

This Ansatz tho it's only usable when either the surface charge distribution σ or V are defined on the boundary of the set V, i.e. when our PDE has a defined boundary value problem with either Dirichlet or Neumann conditions.
Take as an example the following 2D problem.

Example 2.2.1 (Two Infinite Planes). Suppose that there are two infinite plates (grounded) parallel to each other and to the xz plane. One is at $y = 0$ and the other is at $y = a$. At $x = 0$ the left end of this strip is closed by an infinitely vertical strip at some fixed potential $V_0(y)$. Find $V(x, y, z)$ of the system. Since the system is independent from z we gotta solve the following differential equation

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0$$

Where we have the following boundary conditions

$$\begin{cases} V(x, 0) = V(x, a) = 0 \\ V(0, y) = V_0(y) \\ \lim\limits_{x \to \infty} V(x, y) = 0 \end{cases}$$

We begin by separating the variables and writing $V(x, y) = f(x)g(y)$. We substitute into the differential equation and then divide by $f(x)g(y)$ and we get

$$\frac{1}{f(x)} \frac{\mathrm{d}^2 f}{\mathrm{d}x^2} + \frac{1}{g(y)} \frac{\mathrm{d}^2 g}{\mathrm{d}y^2} = 0$$

Note that now we have a sum of two functions depending on only one variable, i.e. $X(x) + Y(y) = 0$ This means that these functions must be equal, opposite in sign and constant, therefore the differential equation decouples into two ordinary differential equations

$$\begin{cases} \dfrac{\mathrm{d}^2 f}{\mathrm{d}x^2} = k f(x) \\ \dfrac{\mathrm{d}^2 g}{\mathrm{d}y^2} = -k g(y) \end{cases}$$

These two equations are of easy solution, and therefore we get

$$\begin{cases} f(x) = Ae^{kx} + Be^{-kx} \\ g(y) = C\cos(ky) + D\sin(ky) \end{cases}$$

Imposing the boundary conditions we get

$$\lim_{x\to\infty} f(x) = 0 \implies A = 0$$
$$g(0) = 0 \implies C = 0$$

The searched potential therefore has the following shape

$$V(x,y) = De^{-kx}\sin(ky)$$

Imposing $V(x,a) = 0$ we have the following constraint on the coupling constant k

$$V(x,a) = De^{-kx}\sin ka = 0 \implies k_n = \frac{n\pi}{a}$$

Therefore, we finally have

$$V_n(x,y) = D_n e^{-\frac{n\pi x}{a}}\sin\left(\frac{n\pi y}{a}\right)$$

The general solution of our problem then will be a linear superposition of *all* solutions, therefore

$$V(x,y) = \sum_{n=0}^{\infty} C_n e^{-\frac{n\pi x}{a}}\sin\left(\frac{n\pi y}{a}\right)$$

This is clearly the Fourier series solution of V, therefore the constants C_n will be found using Fourier's trick and multiplying on the left by $\sin(k_{n'}y)$ and integrating on the expansion interval, which for us is $[0,a]$. We have then, for $V(0,a) = V_0(y)$

$$\sum_{n=0}^{\infty} C_n \int_0^a \sin\left(\frac{k\pi y}{a}\right)\sin\left(\frac{n\pi y}{a}\right)\mathrm{d}y = \int_0^a V_0(y)\sin\left(\frac{k\pi y}{a}\right)\mathrm{d}y$$

Remembering that

$$\int_0^a \sin\left(\frac{k\pi y}{a}\right)\sin\left(\frac{n\pi y}{a}\right)\mathrm{d}y = \frac{a}{2}\delta_{kn}$$

We have

$$C_n = \frac{2}{a}\int_0^a V_0(y)\sin\left(\frac{n\pi y}{a}\right)\mathrm{d}y$$

I.e. C_n are the Fourier coefficients of the function $V_0(y)$. If $V_0(y) = V_0$ is constant the integral can be solved quickly, and we get

$$C_n = \frac{2V_0}{a}\left(1 - \cos(n\pi)\right) = \begin{cases} \frac{4V_0}{n\pi} & n \mod 2k = 0 \\ 0 & n \mod 2k+1 = 0 \end{cases}$$

And the complete solution is then

$$V(x,y) = \frac{4V_0}{\pi}\sum_{n=0}^{\infty}\frac{e^{-\frac{(2n+1)\pi x}{a}}}{2n+1}\sin\left(\frac{(2n+1)\pi y}{a}\right)$$

§§ 2.2.1 Laplace Equation in Spherical Coordinates

What happens when the boundaries exhibit spherical symmetry? We change to spherical coordinates!. The Laplacian in spherical coordinates is

$$\partial_i \partial^i = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial}{\partial r}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial}{\partial \theta}\right) + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2}{\partial \varphi^2}$$

The Laplace equation therefore becomes

$$\partial_i \partial^i V = \frac{1}{r^2}\left(r^2 \frac{\partial V}{\partial r}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial V}{\partial \theta}\right) + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2 V}{\partial \varphi^2} = 0 \tag{2.6}$$

We suppose that the system has azimuthal symmetry, i.e. $\partial_\varphi V = 0$ and we solve the equation using the separation of variables, supposing $V(r,\theta) = f(r)g(\theta)$, then after dividing by V and multiplying by r^2 we get the following equation

$$\frac{1}{f(r)}\frac{\partial}{\partial r}\left(r^2 \frac{\partial f}{\partial r}\right) + \frac{1}{g(\theta)\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial g}{\partial \theta}\right) = 0$$

The equation can be then separated. Taking $c_1 = -c_2 = l(l+1)$ We get two ordinary differential equations

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}r}\left(r^2 \dfrac{\mathrm{d}f}{\mathrm{d}r}\right) = l(l+1)f(r) \\ \dfrac{\mathrm{d}}{\mathrm{d}\theta}\left(\sin\theta \dfrac{\mathrm{d}g}{\mathrm{d}\theta}\right) = -l(l+1)\sin\theta g(\theta) \end{cases} \tag{2.7}$$

The first equation has a power series solution, while the second is a special differential equation solved by the Legendre polynomials $P_l(\cos\theta)$, a complete and orthogonal set of polynomials defined by the recursive relation using the Rodrigues' formula

$$P_l(x) = \frac{1}{2^l l!}\frac{\mathrm{d}^l}{\mathrm{d}x^l}\left[(x^2-1)^l\right] \tag{2.8}$$

The solutions for the two differential equations are then

$$\begin{cases} f(r) = Ar^l + \dfrac{B}{r^{l+1}} \\ g_l(\theta) = P_l(\cos\theta) \end{cases} \tag{2.9}$$

The potential will then be, after superposition of all solutions in l, the following

$$V(r,\theta) = \sum_{l=0}^{\infty}\left(A_l r^l + \frac{B_l}{r^{l+1}}\right)P_l(\cos\theta) \tag{2.10}$$

An example of using this solution is the following

Example 2.2.2 (A Hollow Sphere). Consider a hollow sphere with radius R, find V inside the sphere considering that the surface of the sphere is at some fixed potential $V_0(\theta)$.

The differential equation that must be solved here is the following

$$\begin{cases} \partial_i \partial^i V(r,\theta) = 0 \\ V(\theta, R) = V_0(\theta) \\ \lim_{r\to 0} V(r,\theta) = 0 \end{cases}$$

From the third condition we need that $B_l = 0$, if not the potential would blow up at the center, therefore the first sketch of the solution will be from the general solution (2.10)

$$V(r,\theta) = \sum_{l=0}^{\infty} A_l r^l P_l(\cos\theta)$$

From the second equation we have that at R it must be equal to $V_0(\theta)$. From (2.8) we can also get, using induction, the completeness relation for P_l.

$$\int_{-1}^{1} P_l(x) P_k(x) \mathrm{d}x = \int_0^{\pi} P_l(\cos\theta) P_k(\cos\theta) \sin\theta \mathrm{d}\theta = \frac{2}{2l+1}\delta_{lk} \tag{2.11}$$

Therefore, using Fourier's trick to the potential we found, we get that

$$A_l R^l \frac{2}{2l+1}\delta_{lk} = \int_0^{\pi} V_0(\theta) P_k(\cos\theta)\sin\theta \mathrm{d}\theta$$

This implies that the coefficients A_l we're searching are

$$A_l = \frac{2l+1}{2R^l}\int_0^{\pi} V_0(\theta) P_l(\cos\theta)\sin\theta \mathrm{d}\theta$$

The complete potential inside the sphere is then

$$V(r,\theta) = \sum_{l=0}^{\infty} \frac{2l+1}{2}\left(\frac{r}{R}\right)^l P_l(\cos\theta)\int_0^{\pi} V_0(\theta) P_l(\cos\theta)\sin\theta \mathrm{d}\theta$$

§ 2.3 Multipole Expansion of the Potential

§§ 2.3.1 Electric Dipoles

It's clear that from our calculations, at large distances from the distribution the electrostatic potential behaves approximatively like the potential of a single point charge

$$V(r) \approx \frac{q}{4\pi\epsilon_0}\frac{1}{r}$$

Note that if $Q_{tot} = 0$ we don't have necessarily that $V \approx 0$ at large distances! Take as an example the *electric dipole*. Take two point charges with charge $\pm q$ and position them at some distance d

between them. Writing r_+ and r_- as the distances of each charge from the point considered we can immediately write the potential of such system by superimposing the potentials of each single charge

$$V(r) = \frac{q}{4\pi\epsilon_0}\left(\frac{1}{r_+} - \frac{1}{r_-}\right)$$

Noting that the distance d between the two charges and the distance from the origin of each writes a triangle, we can write

$$r_\pm^2 = r^2 + \frac{d^2}{4} \mp rd\cos\theta = r^2\left(1 + \frac{d^2}{4r^2} \mp \frac{d}{r}\cos\theta\right)$$

In our case $r_\pm >> d$ since we're far from the system, and therefore, approximating to the first order in $\frac{d}{r}$

$$\frac{1}{r_\pm} \approx \frac{1}{r}\left(1 \pm \frac{d}{2r}\cos\theta\right)$$

Therefore

$$\frac{1}{r_+} - \frac{1}{r_-} \approx \frac{d}{r^2}\cos\theta$$

Which, by substitution into our previous definition of the potential, gives

$$V(r) = \frac{1}{4\pi\epsilon_0}\frac{qd\cos\theta}{r^2} \tag{2.12}$$

The term on the right, qd, is known as the *electric dipole moment* of the distribution p.
In general, a potential can be approximated in a *multipole series*. The first term (the dominant one) is known as the *monopole term* of the potential, and it's equal to the potential of a single point charge

$$V_{mon}(r) = \frac{Q}{4\pi\epsilon_0 r} = \frac{1}{4\pi\epsilon_0}\frac{1}{r}\iiint_V \rho(\tilde{r}^i)\mathrm{d}^3\tilde{x}^i \tag{2.13}$$

If the total charge $Q = 0$, as for the previous case, the dominant term will be the *dipole term* of the potential

$$V_{dip}(r) = \frac{1}{4\pi\epsilon_0}\frac{p^i\hat{r}_i}{r^2} = \frac{1}{4\pi\epsilon_0}\frac{\hat{r}_i}{r^2}\iiint_V \tilde{r}^i\cos\theta\rho(\tilde{r}^i)\mathrm{d}^3\tilde{x} \tag{2.14}$$

The vector p^i is what we have defined as the dipole moment of the system, which is equal to

$$p^i = \iiint_V \tilde{r}^i\rho\left(\tilde{r}^i\right)\mathrm{d}^3\tilde{x} \tag{2.15}$$

For the previous case of the two charges, we easily have

$$p^i = qr_+^i - qr_-^i = qd^i \tag{2.16}$$

Where d^i is the vector connecting the two charges.
Note that in the case that the dipole moment of the potential is zero, there will be other terms that will

dominate the expansion, such as *quadrupole terms*, *octupole terms* and so on. The general formula for finding these coefficients can be extracted from the general shape of the potential in integral form

$$\frac{1}{4\pi\epsilon_0}\iiint_V \frac{\rho(\tilde{r}^i)}{\|r^i-\tilde{r}^i\|}\mathrm{d}^3\tilde{x}$$

Using

$$\|r^i-\tilde{r}^i\| = r^2\left(1+\left(\frac{\tilde{r}}{r}\right)^2-2\left(\frac{\tilde{r}}{r}\right)\cos\theta\right)$$

And supposing $\|r^i-\tilde{r}^i\| = r\sqrt{1+\varepsilon}$, where we choose ε as follows

$$\varepsilon = \left(\frac{\tilde{r}}{r}\right)\left(\frac{\tilde{r}}{r}-2\cos\theta\right)$$

We have, for $1+\varepsilon\to 0$, which is the case for long distances from the potential

$$\frac{1}{\|r^i-\tilde{r}^i\|} \approx \frac{1}{r}\left(1-\frac{1}{2}\varepsilon+\frac{3}{8}\varepsilon^2-\frac{5}{16}\varepsilon^3+\cdots\right) \tag{2.17}$$

Rewriting in terms of $\tilde{r}/r,\ \cos\theta$, we have on the right a series of cosines, which is known as the *Legendre Polynomials* in $\cos\theta$ $P_l(\cos\theta)$, which are the solutions to the angular part of the Laplace equation in polar coordinates. The function on the left of the series approximation is known as the *generating function* of the polynomials.

$$\frac{1}{\|r^i-\tilde{r}^i\|} = \frac{1}{r}\sum_{l=0}^{\infty}\left(\frac{\tilde{r}}{r}\right)^l P_l\left(\cos\theta\right) \tag{2.18}$$

In general, we have then that the complete multipole expansion of the electrostatic potential is

$$V_{mult}(r) = \frac{1}{4\pi\epsilon_0}\sum_{l=0}^{\infty}\frac{1}{r^{l+1}}\iiint_V \tilde{r}^l P_l\left(\cos\theta\right)\rho\left(\tilde{r}^i\right)\mathrm{d}^3\tilde{x} \tag{2.19}$$

Note that this gives consistently that the potential goes as $1/r$ for monopoles, $1/r^2$ for dipoles, $1/r^3$ for quadrupoles and so on, and approximate charge distributions at great distances as a sum of simpler problems, a single point charge for the monopole, two point charges for the dipole, four point charges in a square for a quadrupole and so on.

3 Electricity in Matter

§ 3.1 Polarization

So far we dealt with electrostatics with conductors only. We begin to consider non-conducting materials, which are known as *dielectrics*.
What changes from electrostatics with conductors? Experimentally it can be seen using capacitors.
Consider a parallel plane capacitor on which we put a charge Q and then fill the space between the places with some isotropic and homogeneous dielectric.
It can be seen that if V_0 is the potential without the dielectric, then in this case $\Delta V < \Delta V_0$. From the definition of capacitance then

$$C > C_0$$

Experimentally it's seen that, independently from the shape of the capacitor

$$\frac{C}{C_0} = \epsilon_r \tag{3.1}$$

This is known as the *relative dielectric constant*, which is from what we have seen greater than 1 and non-dimensional.
We can then write

$$C = \epsilon_r C_0 = \epsilon_r \frac{\epsilon_0 S}{d} = \frac{\epsilon S}{d} \tag{3.2}$$

Where we defined $\epsilon = \epsilon_r \epsilon_0$, which is the *dielectric constant of the medium*.
Using the known formulas for the capacitance we get that $\Delta V = \Delta V_0 / \epsilon_r$ and therefore $E = E_0 / \epsilon_r$, and this phenomenon can be explained as if we added a surface charge distribution on the two plates, and therefore

$$E = \frac{\sigma + \sigma'}{\epsilon_0}, \qquad E_0 = \frac{\sigma}{\epsilon_0}$$

I.e.

$$\sigma + \sigma' = \frac{\sigma}{\epsilon_r} \implies \sigma' = \frac{1 - \epsilon_r}{\epsilon_r}\sigma$$

We decide to cleverly distribute this charge on the positively charged plate as a negative charge distribution and vice versa on the other plate.
These charges are due to the *polarization* of the medium.

We also have that if we put a point charge inside a dielectric we get a new "scaled" Coulomb law

$$E = \frac{E_0}{\epsilon_r} = \frac{q}{4\pi\epsilon_0\epsilon_r}\frac{1}{r^2} = \frac{q}{4\pi\epsilon}\frac{1}{r^2} \tag{3.3}$$

§§ 3.1.1 The Polarization Field

Consider now an atom A. An atom in general it's a neutral object composed of a positively charged nucleus with charge $q = Ze$ and Z electrons with charge $q = -Ze$ Inserting it into a constant electric field we have that if it's not big enough to ionize the atom (making a conductor) it will move the nucleus and electrons till they get to a stable point, generating a dipole moment p^i. This process is known as *polarization* of the medium.
It's clear that this dipole moment is linearly coupled to the electric field with the following relationship

$$p^i = \alpha E^i \tag{3.4}$$

The coupling constant α is known as the *atomic polarizability* and depends on the chosen atom A. For anisotropic media, like molecules, this coupling constant becomes the *atomic polarizability tensor*, with the following relation

$$p^i = \alpha^i_k E^k \tag{3.5}$$

Consider now a molecule with a "built in" polarization, (i.e. *polar molecules*) like water. What happens when we apply an E field?
If E is uniform then the force on the positive charge cancels the one on the negative, $F_+ = -F_-$, however there is still a torque to consider

$$\tau^i = \epsilon^i{}_{jk} r^j F^k_+ + \epsilon^i{}_{jk} r^j F^k_- \tag{3.6}$$

Since $r^i = \pm d^i/2$ we have, substituting $F_\pm = \pm qE$

$$\tau^i = \frac{q}{2}\epsilon^i{}_{jk} d^j E^k + \frac{q}{2}\epsilon^i{}_{jk} d^j E^k$$

This is nonzero, in fact we have

$$\tau^i = q\epsilon^i{}_{jk} d^j E^k = \epsilon^i{}_{jk} p^j E^k \tag{3.7}$$

I.e., since $p^i = qd^i$ is the dipole moment of the molecule (which is nonzero), there is an induced torque when applying the field, which rotates the molecules until $p^i \parallel E^i$, and therefore $\tau^i = 0$.
Note that if the field is nonuniform we won't have anymore $F_+ = -F_-$, and we will have a net force applied to our dipole (the molecule)

$$F^i = F^i_+ + F^i_- = q\Delta E^i$$

For small dipoles, i.e. for small ΔE^i, we can approximate it to

$$\Delta E^i \approx d^i \partial_i E^j$$

And therefore the net force applied on the dipole is

$$F^i = qd^j \partial_j E^i = p^j \partial_j E^i \tag{3.8}$$

Now consider an element with an amount of molecules of the order of 10^{23}. All these tiny dipoles induced from the electric field or from the single molecule itself will sum up to a general dipole field, called the *polarization field* of the medium. By definition we have, that if V is some volume then

$$P^i = \lim_{V\to 0} \frac{1}{V} \sum_{\alpha=1}^{N} p^i_{(\alpha)} = \langle p^i \rangle \frac{\mathrm{d}N}{\mathrm{d}V} \tag{3.9}$$

Here we have indicated with $\langle p^i \rangle$ the average dipole moment of the system.
Now let's write the potential for a single molecule. Since the molecule can be approximated as a dipole, we know already then that

$$V(r) = \frac{p^i \hat{r}_i}{4\pi\epsilon_0 r^2}$$

From our previous definition of polarization field, then, integrating over all the dielectric and using $\mathrm{d}V \to \mathrm{d}^3x$ we have

$$V_{pol}(r) = \frac{1}{4\pi\epsilon_0} \iiint_V \frac{P^i \hat{r}_i}{r^2} \mathrm{d}^3\tilde{x} \tag{3.10}$$

Looking closely inside the integral, we can rewrite an identity inside that will ease our calculations, in fact

$$\frac{\hat{r}_i}{r^2} = \frac{\partial}{\partial x^i}\left(\frac{1}{r}\right)$$

With a clever trick then we can write, using product rules

$$P^i \frac{\partial}{\partial x^i}\left(\frac{1}{r}\right) = \frac{\partial}{\partial x^i}\left(\frac{P^i}{r}\right) - \frac{1}{r}\frac{\partial P^i}{\partial x^i}$$

Therefore, reinserting into our definition of V and applying Stokes when possible, we have

$$V_{pol}(r) = \frac{1}{4\pi\epsilon_0}\left[\oiint_{\partial V} \frac{P^i \hat{n}_i}{r}\mathrm{d}s - \iiint_V \frac{1}{r}\partial_i P^i \mathrm{d}^3x\right] \tag{3.11}$$

This potential resembles *a lot* the potential given from a volumetric charge plus some surface charge in some closed bound set, like

$$V_v(r) = \frac{1}{4\pi\epsilon_0}\left[\oiint_{\partial V} \frac{\sigma(r^i)}{r}\mathrm{d}s + \iiint_V \frac{\rho(r^i)}{r}\mathrm{d}^3x\right]$$

And, reinterpreting the polarization field as a field generated by a *bound charge*, we can define two simple equations that will make our V similar to V_v. Then, if

$$\begin{cases} P^i \hat{n}_i = \sigma_b \\ \partial_i P^i = -\rho_b \end{cases} \tag{3.12}$$

And defined as V_σ, V_ρ the two potentials generated by this "bound charge", we have that the total potential generated by a polarized medium is

$$V_{pol}(r^i) = V_\sigma(r^i) + V_\rho(r^i) \tag{3.13}$$

A nice observation from the second equation of (3.12) is that if the dielectric is homogeneous, the dipole moments inside the object will average to 0, and therefore P^i will be independent from the position inside the dielectric, i.e.

$$\partial_i P^i = 0 = -\rho_b$$

And all bound charges will be on the surface with distribution σ_b

§ 3.2 Perfect Dielectrics

§§ 3.2.1 Local Electric Field

So far we defined a dielectric as a cluster of molecules and atoms. It's clear so far that each atom and molecule has its little microscopic e^i field, therefore the electric field inside a dipole can change greatly between points, depending on where we measure the field, if near or far away from an electron (considering that the distances are $d \approx 10^{-10}\ \mathrm{m}$ "far" can be a negligible quantity in relation to the dimension of the dielectric).
Take now a really small part of the dielectric, in this small element of dielectric we will have inside some sphere S molecules which are polarized when an external field E^i gets applied.
We consider 2 major cases:

1. There are no molecules inside S and therefore there will be only the bound surface charge $\sigma_b = P^i \hat{n}_i$ with $\hat{n}^i$ being the outward normal of the conductor
2. There are molecules inside S and therefore, there will also be a field generated by the polarization of the molecules

The field at the center of S, E^i_S will then be the sum of these three fields we considered, the external polarizing field E^i, the field $\tilde{E}^i$ generated by the bound surface charge, and the field E^i_{dip} generated by the molecular dipoles. Therefore

$$E^i_S = E^i + \tilde{E}^i + E^i_{dip} \tag{3.14}$$

Due to the homogeneity of the dielectric we must have that $\partial_i P^i = 0$, and therefore the field generated by the dipoles and the bound surface charge must balance themselves, $\tilde{E}^i + E^i_{dip} = 0$.
As we said before the molecule itself generates a small microscopic field e^i, therefore we define a *local field* or *Lorentz field* inside the dielectric by subtracting this e^i. We have that this field E^i_{loc} is

$$E^i_{loc} = E^i + \tilde{E}^i + E^i_{dip} - e^i = E^i + \tilde{E}^i + \underline{E}^i \tag{3.15}$$

Where we defined $\underline{E}^i = E^i_{dip} - e^i$. What's this field then?
We begin by evaluating $\tilde{E}^i$, which is the field generated by the surface charge. Then by definition of the E^i field itself we can immediately say

$$\mathrm{d}\tilde{E}^i = \frac{1}{4\pi\epsilon_0}\frac{\sigma_b \hat{r}^i}{r^2}\mathrm{d}s$$

Due to the symmetries imposed on the system (homogeneity of the dielectric,...) we have that $\mathrm{d}\tilde{E}^z = -\left\|\mathrm{d}\tilde{E}^i\right\|\cos\theta$, and therefore, remembering that $\sigma_b = P^i\hat{n}_i = -P\cos\theta$ ($\hat{n}^i$ is the *outward* normal)

$$\mathrm{d}\tilde{E}^z = -\frac{\sigma_b \cos\theta}{4\pi\epsilon_0 r^2}\mathrm{d}s$$

Since $\mathrm{d}s = r^2 \mathrm{d}\Omega$ we then have

$$\mathrm{d}\tilde{E}^z = -\frac{P\cos^2\theta}{4\pi\epsilon_0}\mathrm{d}\Omega \tag{3.16}$$

Integrating, we have

$$\tilde{E}^z = -\frac{P}{4\pi\epsilon_0}\int_0^{2\pi}\mathrm{d}\phi\int_{-\pi}^{\pi}\cos^2\theta\sin\theta\mathrm{d}\theta = \frac{P}{2\epsilon_0}\int_{-1}^{1}\cos^2\theta\mathrm{d}\left(\cos\theta\right) = \frac{P}{3\epsilon_0} \tag{3.17}$$

Therefore, we firstly found that

$$\tilde{E} = \frac{P}{3\epsilon_0} \tag{3.18}$$

We only miss evaluating the field generated by the dipoles minus the microscopic molecular electric field. We only need to know what's the field generated by an isotropic dipole.
We know already that an electric dipole has the following scalar potential

$$V(r) = \frac{p^i\hat{r}_i}{4\pi\epsilon_0 r^2} = \frac{p^i r_i}{4\pi\epsilon_0 r^3}$$

Taking the gradient we have

$$\frac{\partial V}{\partial x_i} = \frac{1}{4\pi\epsilon_0}\left(\frac{1}{r^3}\frac{\partial}{\partial x_i}\left(p^j r_j\right) + p^j r_j\frac{\partial}{\partial x_i}\left(\frac{1}{r^3}\right)\right)$$

Expanding and writing explicitly the gradient of a radial function with the usual formula, we have

$$\frac{\partial V}{\partial x_i} = \frac{1}{4\pi\epsilon_0}\left[\frac{1}{r^3}\left(\frac{\partial p^j}{\partial x_i}r_j + p^j\frac{\partial r_j}{\partial x_i}\right) - \frac{3(p^j r_j)r^i}{r^5}\right]$$

Using $\partial^i p^j = 0$ and $\partial^i r_j = \delta^i_j$ we have that

$$\frac{\partial}{\partial x_i}\left(p^j r_j\right) = p^j\delta^i_j = p^i$$

And therefore

$$\frac{\partial V}{\partial x_i} = \frac{1}{4\pi\epsilon_0}\left(\frac{p^i}{r^3} - \frac{3(p^j r_j)r^i}{r^5}\right)$$

Writing $\hat{r}^i = r^i/r$ we have finally, multiplying by -1

$$E^i = -\frac{\partial V}{\partial x_i} = \frac{1}{4\pi\epsilon_0 r^3}\left(3\left(p^j\hat{r}_j\right)\hat{r}^i - p^i\right) \tag{3.19}$$

In our special isotropic case inside a little sphere S, inside a dipole itself we have that this field will be oriented on the z axis, with a constant dipole moment of the $\alpha-$th molecule $p_{(\alpha)}$, then

$$\underline{E} = \underline{E}^z = \sum_{\alpha=1}^{N}\frac{p_{(\alpha)}\left(3z^2_{(\alpha)} - r^2_{(\alpha)}\right)}{4\pi\epsilon_0 r_{(\alpha)^5}} \tag{3.20}$$

(Note that here we took the opposite process and rewrote the non normalized vector r^i for ease of calculation).
Since the $p_{(\alpha)}$ are uniformly distributed around the dielectric we must have that

$$\sum_{\alpha=1}^{N} \frac{x_{(\alpha)}^2}{r_{(\alpha)}^5} = \sum_{\alpha=1}^{N} \frac{y_{(\alpha)}^2}{r_{(\alpha)}^5} = \sum_{\alpha=1}^{N} \frac{z_{(\alpha)}^2}{r_{(\alpha)}^5} = \frac{1}{3} \sum_{\alpha=1}^{N} \frac{r_\alpha^2}{r_{(\alpha)}^5} \tag{3.21}$$

Simply inserting it back into the definition of $\underline{E}$ we get then $\underline{E} = 0$.
The final result for the Lorentz field (the local field inside a dielectric), considering all the microscopic variables, is

$$E_{loc}^i = E^i + \frac{P^i}{3\epsilon_0} \tag{3.22}$$

I.e. it only depends on the external applied field E^i and the polarization of the dielectric P^i (divided by $3\epsilon_0$)

§§ 3.2.2 Susceptibility and the Clausius-Mossotti relation

So far we can finally conclude that with a good approximation the polarization of the dielectric P^i must depend on this local field E_{loc}^i, which basically decides how a certain material gets polarized. Therefore, using the definition of P^i and defining the numerical volumetric density of molecules $\frac{\mathrm{d}N}{\mathrm{d}V} = n$

$$P^i = n \left\langle p^i \right\rangle = n\alpha E_{loc}^i \tag{3.23}$$

§§§ 3.2.2.1 Gases and Vapors

Let's now consider different relations between the Lorentz field and the polarization field. The easiest case to consider is a gas. In this case, if we take the perfect gas approximation, i.e. the density is low enough, we can say that the molecules are too far apart in order for their fields to interact between each other, therefore $E_{loc}^i \approx E^i$.
We also have to consider thermal excitations of the molecules of the gas, and therefore the coupling constant α must be split in two parts. One, α_d, dependent on the molecule itself, and one α_t depending on the temperature of the gas and the specific polarization of the molecule

$$\alpha = \alpha_d + \alpha_t = \alpha_d + \frac{p_0^2}{3kT}$$
$$P^i = n\alpha E^i = n\left(\alpha_d + \frac{p_0}{3kT}\right) E^i$$

We then define the *electric susceptibility* of the medium χ via the following relation

$$P^i = \epsilon_0 \chi E^i \tag{3.24}$$

Therefore, for a gas

$$\chi(T) = \frac{n}{\epsilon_0}\left(\alpha_d + \frac{p_0}{3kT}\right) = \epsilon_r - 1 \tag{3.25}$$

Where ϵ_r is the relative permittivity of the substance, as we will see later

§§§ 3.2.2.2 Liquids and Amorphous Substances

For liquids everything changes a little bit. Since the density isn't low enough, the molecules will be packed and their local field will comprise of the external field applied plus the field generated by the polarization. We have

$$E^i_{loc} = E^i + \frac{P^i}{3\epsilon_0}$$
$$P^i = n\alpha E^i_{loc}$$

Then, by mere substitution

$$P^i = n\alpha\left(E^i + \frac{P^i}{3\epsilon_0}\right)$$

Solving for P^i (bringing it on the left and taking it outside the product with the constants) we have then

$$P^i = \frac{n\alpha}{1 - \frac{n\alpha}{3\epsilon_0}} E^i = \epsilon_0 \chi E^i \tag{3.26}$$

Now, solving for α, we have after some algebra, the *Clausius-Mossotti relation*, which ties α, a microscopic quantity, to ϵ_r via χ, a macroscopic quantity

$$\alpha = \frac{\epsilon_0}{n}\frac{3(\epsilon_r - 1)}{\epsilon_r + 2} \tag{3.27}$$

§§§ 3.2.2.3 Anisotropic Solids, Electrets and Piezoelectricity

In general when the solid is anisotropic, as we defined before the polarizability is not a simple constant but a tensor, where

$$P^i = \alpha^i_j E^j \tag{3.28}$$

For other materials, α can also be nonlinear. Take for example *electrets*. An *electret* or a *ferroelectric material* is a material which keeps a permanent polarization inside after turning off the external field, showing magnet-like behavior, like *hysteresis*. In this case α is non-unique.
Another example of a non-linear relation comes from *piezoelectric materials*. *Piezoelectricity* is a phenomenon given by substances that polarize under mechanical pressure, like quartz. In these materials α must depend on the mechanical pressure itself.

§§ 3.2.3 The Electric Displacement Field

So far, adding the theory on dielectrics, we can build multiple equations describing the polarization P^i, bound charges ρ_b, σ_b and the relation between P^i and E^i.
From Gauss' law we know that the divergence of the E^i field is equal to the (total) volumetric charge divided by ϵ_0. With dielectrics we then gotta consider also bound charges, therefore

$$\partial_i E^i = \frac{\rho + \rho_b}{\epsilon_0}$$

Remembering that the bound volumetric charge is tied to the polarization with the differential equation

$$\partial_i P^i = -\rho_b$$

We then have

$$\partial_i E^i = \frac{\rho}{\epsilon_0} - \frac{1}{\epsilon_0}\partial_i P^i$$

Multiplying by ϵ_0 and bringing $\partial_i P^i$ on the left and using the linearity of ∂_i we have

$$\partial_i \left(\epsilon_0 E^i + P^i\right) = \rho$$

We define the vector on the left as the *Electric Displacement field* D^i

$$D^i = \epsilon_0 E^i + P^i \tag{3.29}$$

And we immediately get from the previous equation, the equivalent Gauss law for this field

$$\partial_i D^i = \rho \tag{3.30}$$

With this field, the first and third Maxwell equations in dielectrics become two coupled partial differential equations for two different fields, E^i and D^i

$$\begin{cases} \partial_i D^i = \rho \\ \epsilon^i{}_{jk}\partial^j E^k = 0 \end{cases} \tag{3.31}$$

This is solvable only if we know the relations between D^i and E^i, or in general how P^i is related to E^i. In a perfect dielectric we have that the polarizability tensor α^i_j is independent of the position, time and electric field (note that a gas cannot be a perfect dielectric since α depends on the temperature).
We will study only isotropic perfect dielectrics, also known as *linear dielectrics*, where $\alpha^i_j = \alpha\delta^i_j$, and we can write for these, as we saw before

$$P^i = \alpha E^i = \epsilon_0 \chi E^i \tag{3.32}$$

Therefore, from our previous definition of D^i and noting that $\chi = \epsilon_r - 1$,

$$D^i = \epsilon_0 E^i + P^i = \epsilon_0 E^i + \epsilon_0 \chi E^i = \epsilon_0 \left(1 + \chi\right) E^i = \epsilon_0 \epsilon_r E^i \tag{3.33}$$

Using $\epsilon = \epsilon_0 \epsilon_r$ we have then, that in linear dielectrics the D^i field is linearly dependent on the E^i field, where

$$D^i = \epsilon E^i \tag{3.34}$$

Note that outside a dielectric (i.e. in free space) we must have $P^i = 0$, and therefore

$$D^i_f = \epsilon_0 E^i_f \tag{3.35}$$

Maxwell's equations for a linear dielectric then modify to a much simpler variant which differs from the usual electrostatic maxwell equations by simply setting $\epsilon_0 \to \epsilon$

$$\begin{cases} \partial_i E^i = \dfrac{\rho}{\epsilon} \\ \epsilon^i{}_{jk}\partial^j E^k = 0 \end{cases} \tag{3.36}$$

Note that in free space

$$\partial_i \left(\epsilon_0 E^i_f\right) = \rho$$

And in a dielectric

$$\partial_i \left(\epsilon E^i\right) = \rho$$

Then, we must have

$$\partial_i \left(\epsilon E^i\right) = \partial_i \left(\epsilon_0 E_f^i\right)$$

Integrating and using the first principle of the calculus of variation then it's obvious that

$$\epsilon_0 E_f^i = \epsilon E^i \implies E^i = \frac{1}{\epsilon_r} E_f^i \tag{3.37}$$

Where we used $\epsilon = \epsilon_r \epsilon_0$. This is the exact same experimental result that we found empirically before with the parallel plate capacitor

§ 3.3 Maxwell Equations for Electrostatics in Linear Dielectrics

We can now begin defining all the various laws we derived for electrostatic fields in free space in presence of dielectrics, using the linear relations that we found before.
From Gauss' law for the D^i field integrating we immediately have

$$\iiint_V \partial_i D^i \mathrm{d}^3x = \oiint_{\partial V} D^i \hat{n}_i \mathrm{d}s = Q_{loc} = \iiint_V \rho \mathrm{d}^3x \tag{3.38}$$

And, analogously, the Coulomb theorem for surface charges

$$D^i = \sigma \hat{n}^i \tag{3.39}$$

Note that we didn't indicate the *total* charge inside V, Q_V, since we're not considering the polarization bound charge Q_b! We're only considering the "free" charge, which is not due to polarization effects of the dielectric.
We have a bit of luck tho when dielectrics are linear, then with a simple multiplication of the third Maxwell equation by ϵ we also get a coupled set of equations for the D^i field

$$\begin{cases} \partial_i D^i = \rho \\ \epsilon^i{}_{jk} \partial^j D^k = 0 \end{cases} \tag{3.40}$$

Due to the clear linear relations between E^i and P^i it's also possible to know the polarization inside the medium, which is not always obvious and measurable (it's clear only for perfect dielectrics). Since $E^i = D^i/\epsilon_0\epsilon_r$ and $\chi = \epsilon_r - 1$ we have

$$P^i = \epsilon_0 \chi E^i = \epsilon_0(\epsilon_r - 1)\frac{D^i}{\epsilon_0 \epsilon_r} = \frac{\epsilon_r - 1}{\epsilon_r} D^i \tag{3.41}$$

Example 3.3.1 (A Charged Dielectric Sphere). Take as an example a sphere composed of dielectric material of radius R with charge Q.
From Gauss' theorem for D^i we have, for a spherical Gaussian surface with $r > R$

$$\Phi\left(D^i\right) = 4\pi R D = Q \implies D = \frac{Q}{4\pi r^2}$$

Since $E^i = \epsilon^{-1} D^i$ and $P^i = \epsilon_0 \chi E^i$ we have

$$P^i = \epsilon_0 \chi \frac{D^i}{\epsilon} = \frac{\epsilon_0(\epsilon_r - 1)}{\epsilon_0 \epsilon_r} D^i = \left(\frac{\epsilon_r - 1}{\epsilon_3} \right) \frac{Q}{4\pi r^2} \hat{r}^i$$

And

$$E^i = \frac{Q}{4\pi\epsilon r^2} \hat{r}^i$$

The bound polarization charge distributions are then found using the known formulas, and therefore for the surface polarization charge

$$\sigma_b = P^i \hat{n}_i = -P^i \hat{r}_i = -\frac{\epsilon_r - 1}{\epsilon_r} \frac{Q}{4\pi R^2} = -\frac{\epsilon_r - 1}{\epsilon_r} \sigma$$

The total polarization charge is

$$Q_b = 4\pi R^2 \sigma_b = -\frac{\epsilon_r - 1}{\epsilon_r} Q$$

And therefore the total charge is

$$Q_t = Q + Q_b = Q \left(1 - \frac{\epsilon_r - 1}{\epsilon_r} \right) = \frac{Q}{\epsilon_r}$$

While, for the volumetric polarization charge we have

$$\rho_b = -\partial_i P^i = \frac{1}{r^2} \frac{\mathrm{d}}{\mathrm{d}r} \left(r^2 P^r \right) = -\frac{1}{r^2} \frac{\mathrm{d}}{\mathrm{d}r} \left(\frac{\epsilon_r - 1}{\epsilon_r} \frac{Q}{4\pi} \right) = 0$$

I.e. $\rho_b = 0$ as we expected. Since the dielectric is neutral there also must be a charge $-Q_b > 0$ at $r \to \infty$.

Example 3.3.2 (A Parallel Plate Capacitor). This example is quite simple. We know from Gauss' theorem for the surface charge and D^i that

$$D = \sigma$$

Therefore

$$E = \frac{D}{\epsilon} = \frac{\sigma}{\epsilon}$$

The polarization field instead is

$$P = \epsilon_0 \chi E = \chi \epsilon_0 \frac{\sigma}{\epsilon} = \frac{\epsilon_r - 1}{\epsilon_r} \sigma$$

And, the polarization surface charge (remembering that we take the outer normal) is

$$\sigma_b = P^i \hat{n}_i = -P = -\frac{\epsilon_r - 1}{\epsilon_r} \sigma$$

The potential difference between the plates is simply

$$\Delta V = Ed = \frac{D}{\epsilon} d = \frac{\sigma}{\epsilon} d = \sigma S \frac{d}{\epsilon S} = Q \frac{d}{\epsilon S} = \frac{Q}{C}$$

Note that

$$C = \frac{S\epsilon}{d} = \epsilon_r C_0$$

As we expected.

§§ 3.3.1 Boundary Conditions

Suppose now that we have multiple dielectric regions. On the boundaries of these regions, passing from one dielectric to another, it's clear that the fields D^i, E^i have discontinuities and therefore we cannot use the differential equations anymore.
We might try by either solving the equations for every dielectric region, or instead by directly solving Poisson's equation with appropriate boundary conditions for each dielectric.
While we cannot use the differential equations (local) in the boundaries of the dielectrics we instead have that the integral relations still hold, where

$$\begin{cases} \oiint D^i \hat{n}_i \mathrm{d}s = 0 \\ \oiint E^i \hat{n}_i \mathrm{d}s \neq 0 \end{cases} \tag{3.42}$$

These imply that

1. The dielectric is neutral, $\sigma = 0$
2. The dielectric is polarized, $\sigma_b \neq 0$

Considering an infinitesimal cylinder centered on the boundary of two different dielectric regions, we have that inside the cylinder the flux of D must be 0, i.e.

$$D_1^i \hat{n}_i^1 \mathrm{d}s + D_2^i \hat{n}_2 \mathrm{d}s = 0$$

Therefore, noting that $\hat{n}_1^i = -\hat{n}_2^i$ we get

$$\left(D_1^i - D_2^i\right) \hat{n}_i^1 \mathrm{d}s = 0 \implies D_{n1} = D_{n2} \tag{3.43}$$

While, for E, using $D_{n1} = \epsilon_1 E_{n1}$

$$\frac{E_{n1}}{E_{n2}} = \frac{\epsilon_2}{\epsilon_1} \neq 1 \tag{3.44}$$

Considering that $\epsilon^i{}_{jk} \partial^j E^k = 0$ we can write instead, for the line integral on the closed curve describing the cylinder instead that

$$\begin{aligned} E_{t1} &= E_{t2} \\ \frac{D_{t1}}{D_{t2}} &= \frac{\epsilon_1}{\epsilon_2} \end{aligned} \tag{3.45}$$

The boundary conditions between two dielectrics then become the following connection relations

$$E_{t1} = E_{t2}, \qquad D_{n1} = D_{n2} \tag{3.46}$$

Example 3.3.3 (Parallel Plate Capacitor with 2 Dielectrics Inside). Consider now a parallel plate capacitor with surface area S, composed inside of two dielectrics, one thick d_1 with permeability ϵ_1 and one thick d_2 with permeability ϵ_2. If we smear on the plates a charge Q we have that by our previous definitions that D only sees the charge Q but not the polarization charges, that E sees. Since the field is normal to the plates we must have that between the two dielectrics

$$D_1 = D_2 = D$$

And, for what we've seen before

$$D = \sigma = \frac{Q}{S}$$

The potential difference is then

$$\Delta V = E_1 d_1 + E_2 d_2 = D\left(\frac{d_1}{\epsilon_1} + \frac{d_2}{\epsilon_2}\right) = \frac{Q}{S}\left(\frac{d_1}{\epsilon_1} + \frac{d_2}{\epsilon_2}\right)$$

Note that

$$\frac{\Delta V}{Q} = \frac{1}{C} = \frac{d_1}{\epsilon_1 S} + \frac{d_2}{\epsilon_2 S} = \frac{1}{C_1} + \frac{1}{C_2}$$

I.e. this parallel plate capacitor works exactly as a series of two capacitors! From what we've seen before we can write then the potential difference of these "2" capacitors

$$\Delta V_1 = E_1 d_1 = \frac{D}{\epsilon_1} d_1$$
$$\Delta V_2 = E_2 d_2 = \frac{D}{\epsilon_2} d_2$$

Or noting that

$$\Delta V = \frac{\epsilon_1 d_2 + \epsilon_2 d_1}{\epsilon_1 \epsilon_2} D$$

We can write

$$\Delta V_1 = \frac{\epsilon_2 d_1}{\epsilon_2 d_1 + \epsilon_1 d_2} \Delta V$$
$$\Delta V_2 = \frac{\epsilon_1 d_2}{\epsilon_2 d_1 + \epsilon_1 d_2} \Delta V$$

§ 3.4 Electrostatic Energy with Dielectrics

We know already that the electrostatic energy in free space is given by the following formula

$$U = \frac{1}{2} \iiint_{\mathbb{R}^3} \rho V \, \mathrm{d}^3 x$$

In presence of dielectrics this still holds if we consider that $\rho = \rho_b + \rho_f$ where ρ_b, ρ_f are the bound polarization charges and the free charges respectively.
Remembering that $\partial_i D^i = \rho$ and integrating by parts, we get for a volume V

$$U = \frac{1}{2} \iiint_V \frac{\partial}{\partial x^i} \left(D^i V\right) \mathrm{d}^3 x - \iiint_V D^i \partial_i V \, \mathrm{d}^3 x \tag{3.47}$$

Sending $V \to \mathbb{R}^3$ we get that the first integral is zero (it becomes a surface integral with Stokes' theorem and goes to 0), therefore, for a dielectric, remembering that $-\partial_i V = E_i$

$$U = \frac{1}{2} \iiint_{\mathbb{R}^3} D^i E_i \mathrm{d}^3 x \tag{3.48}$$

Which implies that the volumetric energy density for a dielectric is

$$u = \frac{1}{2} D^i E_i \tag{3.49}$$

For a perfect isotropic dielectric $D^i = \epsilon E^i$, therefore

$$u = \frac{1}{2} D^i E_i = \frac{\epsilon}{2} E^2 = \frac{1}{2} \frac{D^2}{\epsilon} \tag{3.50}$$

Which, if integrated, give the exact identical result for free space if we substitute $\epsilon_0 \to \epsilon$

Part II

Magnetostatics

4 The Magnetic Field

§ 4.1 Electric Currents

As for all materials (like dielectrics) we can see conductors as crystalline solids with atoms positioned at the nodes of the crystalline reticle, the valence electrons are free to roam on on all the volume of the conductor via thermal agitations.
Classically, this gas of electrons is in thermal equilibrium with kinetic energy

$$\frac{1}{2}m_e \langle v^2 \rangle = \frac{3}{2}kT$$

This implies a classical root mean square velocity of

$$v_t = \sqrt{\langle v^2 \rangle} = \sqrt{\frac{3kT}{m_e}} \approx 120 \ \frac{\mathrm{km}}{\mathrm{s}}$$

If we add over this an electric field we get an ordinate motion of charges, that we call *current*.
Consider now a conductor where charges are moving and consider a generic section of it, if in this section S in some time interval $\mathrm{d}t$ a charge $\mathrm{d}q$ passes through that, we define the current I as

$$I = \frac{\mathrm{d}q}{\mathrm{d}t} \tag{4.1}$$

The units of current in the SI are therefore

$$[I] = \frac{[Q]}{[t]} = 1 \ \frac{\mathrm{C}}{\mathrm{s}} = 1 \ \mathrm{A} \tag{4.2}$$

These units are known as *Ampères*.
In a conductor, if a current is present, we have that if we have two sections with different potentials V_A and V_B the work needed to move charges between the plates is

$$\mathrm{d}W = V_A \mathrm{d}q - V_B \mathrm{d}q = I \Delta V \mathrm{d}t \implies \frac{\mathrm{d}W}{\mathrm{d}t} = I \Delta V \tag{4.3}$$

This represents the work employed by the electric field for moving the charges.
In general the electric field will induce a force F_e^i, which, for Newton's second law will be

$$F_e^i = qE^i = ma^i \implies a^i = \frac{qE^i}{m} \tag{4.4}$$

Remembering that for electrons $q = -e$ we have that the acceleration will be on the opposite direction from the field. The average velocity of electrons in a time Δt will then be, by definition

$$\left\langle v^i \right\rangle = \frac{1}{2}\Delta v^i = -\frac{eE^i}{2m}\Delta t$$

This, between collisions, that happen on average in a time $\Delta t = \lambda / v_t$, where λ is the mean free path of the electron and v_t is the thermal velocity of the electrons. Therefore

$$\left\langle v^i \right\rangle = -\frac{e}{2m}\frac{\lambda}{\sqrt{\frac{3kT}{m}}}E^i = -e\frac{\lambda}{\sqrt{3kTm}}E^i$$

I.e. $\left\langle v^i \right\rangle \propto E^i$.
Formally for a group of charges q inside a conductor where current is flowing, taken a differential cross-sectional surface $\mathrm{d}s$ we have that the charge flowing through that is, if n is the volumetric density of charges

$$\mathrm{d}Q = qn \left\langle v^i \right\rangle \hat{n}_i \mathrm{d}s\mathrm{d}t$$

We define the *current density* as the vector J^i, where

$$J^i = nq \left\langle v^i \right\rangle$$

Then, we have

$$\mathrm{d}Q = J^i \hat{n}_i \mathrm{d}s\mathrm{d}t \tag{4.5}$$

Therefore, the charge through this flux tube is

$$\mathrm{d}I = \frac{\mathrm{d}Q}{\mathrm{d}t} = J^i \hat{n}_i \mathrm{d}s$$

Integrating, we get

$$I = \int \mathrm{d}I = \oiint_S J^i \hat{n}_i \mathrm{d}s \tag{4.6}$$

It's clear that the current density then has the following units

$$[J] = [n][q][\langle v \rangle] = \ \mathrm{m}^{-3} \cdot \mathrm{C} \cdot \mathrm{ms}^{-1} = \ \frac{\mathrm{A}}{\mathrm{m}^2}$$

Considering that the average velocity of the charges is $\langle v \rangle \propto E$, since it's collinear with the E^i field (with direction depending on the sign of the charge) we have that with a special kind of conductors (known as *Ohmic conductors*) we can define what is known as *Ohm's law* which ties directly current flow density to the electric field applied

$$J^i = \sigma E^i \tag{4.7}$$

σ is known as the *electric conductibility*, and indicates how much one is able to move charges applying an electric field

§§ 4.1.1 Charge Conservation

Consider a bounded volume V with inside it some variable charge $Q(t)$. Since charge doesn't pop up randomically into existence (yet) we have that if $Q(t)$ changes there must be some charge going inside or coming outside V depending from the current itself. Then, if the current goes *out*, the charge must *reduce*. Then

$$-\mathrm{d}Q = I\mathrm{d}t = \oiint_{\partial V} J^i \hat{n}_i \mathrm{d}s\mathrm{d}t \tag{4.8}$$

In general we can write Q as a volume integral, and deriving the previous equation with respect to time, we have (bringing inside the integral the time derivative)

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = \iiint_V \frac{\partial \rho}{\partial t}\mathrm{d}^3x = -\oiint_{\partial V} J^i \hat{n}_i \mathrm{d}s$$

Using Stokes' theorem and bringing everything to the left then we have that

$$\iiint_V \left(\frac{\partial \rho}{\partial t} + \frac{\partial J^i}{\partial x^i}\right)\mathrm{d}^3x = 0 \tag{4.9}$$

This implies the *continuity equation*

$$\partial_i J^i + \partial_t \rho = 0 \tag{4.10}$$

We define now *stationary currents* as currents for which $\partial_t \rho = 0$, i.e., where the continuity equation becomes

$$\partial_i J^i = 0 \tag{4.11}$$

§ 4.2 Lorentz Force and Magnetic Fields

§§ 4.2.1 Lorentz Force Law

Consider a small (closed) test circuit where there is a stationary current and a mobile rod long $\mathrm{d}l$ with its tangent in the same direction of I. If we put it in a permanent magnetic field or next to cables where current is passing through we have that it will experience a force $\mathrm{d}F$ such that

$$\mathrm{d}F \propto I, \qquad \mathrm{d}F \propto \mathrm{d}l, \qquad \mathrm{d}F^i \perp \hat{t}^i \mathrm{d}l \tag{4.12}$$

Noting that there is a direction for which $F = 0$ which coincides with the direction of the magnetic needle in that point we can write what is known as *Lorentz's force law*, which introduces a new field, the *magnetic induction field* B^i which is oriented as the magnetic needle in the point considered

$$\mathrm{d}F^i = I\epsilon^i{}_{jk}\hat{t}^j B^k \mathrm{d}l \tag{4.13}$$

Rewriting $I\hat{t}^i\mathrm{d}l$ in terms of the current density J^i we have

$$I\hat{t}^i\mathrm{d}l = J\hat{t}^i\mathrm{d}S\mathrm{d}l = J^i\mathrm{d}^3x$$

Therefore

$$\mathrm{d}F^i = \epsilon^i{}_{jk}J^j B^k \mathrm{d}^3x \tag{4.14}$$

Or, writing $J^i = nqv^i$ and integrating

$$F^i = \iiint_V \epsilon^i{}_{jk} J^i B^k \mathrm{d}^3x = q \iiint_V \epsilon^i{}_{jk} v^i B^k n \mathrm{d}^3x$$

Writing again $n\mathrm{d}^3x = \mathrm{d}N$ we have then integrating the final and better known shape of Lorentz's force law

$$F^i = q\epsilon^i{}_{jk} v^j B^k \tag{4.15}$$

Note that by this definition, we have that *magnetic forces do no work*. In fact we have

$$\mathrm{d}W = F^i \mathrm{d}x_i = q\epsilon^i{}_{jk} v_i v^j B^k \mathrm{d}t = 0 \tag{4.16}$$

Note that this is obvious by the definition of the triple product $\epsilon^i{}_{jk} v_i v^j B^k$. In boldface notation this can be written as a determinant of the matrix which has as colums the vectors v^i, v^j, B^k, therefore since two colums are equal, it must be zero by definition of the determinant.
What's the unit for magnetic fields in the SI system?

$$[B] = \left[\frac{F}{qv}\right] = \frac{\mathrm{kg}\cdot\mathrm{m}\cdot\mathrm{s}^{-2}}{\mathrm{C}\cdot\mathrm{m}\cdot\mathrm{s}^{-1}} = \mathrm{kg}\cdot\mathrm{C}^{-1}\mathrm{s}^{-1} = \frac{\mathrm{Vs}}{\mathrm{m}^2} = \frac{\mathrm{Wb}}{\mathrm{m}^2} = \mathrm{T} \tag{4.17}$$

The unit Wb is known as *Weber* and it's a measure of *magnetic flux*, while the one on the right measures the irradiance of the field and is known as *Tesla*.
The *total* force on an electric charge can then be written as a sum of the Coulomb force and the Lorentz force, where

$$F^i = qE^i + q\epsilon^i{}_{jk} v^j B^k \tag{4.18}$$

It's clear that if $v = 0$ there won't be any Lorentz force, and therefore B^i acts only on moving charges. Do another supposition: Take a set of moving charges with current I. If I do a relativistic frame change and I move to a frame where the charges have $v = 0$ then B disappears magically. It means that there must be an *electromagnetic field* which transforms covariantly with respect to Lorentz transformations, which can be seen as a mix of magnetic and electric fields.
Basically, the magnetic field can be seen as some kind of "relativistic effect" on charges

§§ 4.2.2 Motion of Charges in Magnetic Fields

Consider a charge with some velocity v in a magnetic field, we have then that in absolute value we have a *centripetal* force

$$F = qvB$$

Since the force is centripetal the motion will be circular on a plane, take it as the xy plane. In this plane

$$\frac{mv^2}{R} = qvB \implies \frac{v}{R} = \frac{qB}{m} = \omega$$

ω is the frequency of the circular motion, and the radius is

$$R = \frac{mv}{qB}$$

With period $T = 2\pi m/qB$ independent from the speed of the particle, and momentum $p = mv = qBR$. If the initial velocity goes along the z axis it's clear that the motion then is elicoidal.

Let's write this properly. Writing $v \to v^i$ in vector form, and $B^i = B\delta^i_3$ we have evaluating the cross product (use your preferred method, note that using Levi-Civita symbols it's easy to find when it's positive in 3 dimensions. Knowing that it's positive for even permutations, you can write 12312 as the first row and then writing 3 subsequent numbers removing the first one of the row. You'll get then three rows with values 123,231,312 which are all the positive elements of $\epsilon^i{}_{jk}$. obviously the negative ones will be 132,213,321 and the cross product vector can be then be written easily in terms of the components of the product of the two vectors multiplied, $v^i B^j$ in our case. Another way is to write the two vectors one over another and then work more or less like if it's a determinant starting from above. When we hide the x component of the first vector we will have the x component of the product, when we hide the second we will have *minus* the y component and at last the z coponent. Another way is to write a determinant in vector quaternion notation as follows

$$\epsilon^i{}_{jk} a^j b^k \to \det \left| \begin{pmatrix} \hat{\mathbf{e}}_1 & \hat{\mathbf{e}}_2 & \hat{\mathbf{e}}_3 \\ a^i & a^2 & a^3 \\ b^1 & b^2 & b^3 \end{pmatrix} \right|$$

Remember that each way is valid as long as you get the right answer, especially for cross products which have their particularities) After this long parenthesis, we calculate our cross product with our favorite method, and we get, (Remember that $B^i \parallel z$, i.e. $B^i = B\delta^i_3$ this can ease calculations)

$$m\frac{\mathrm{d}v^i}{\mathrm{d}t} = q\epsilon^i{}_{jk} v^j B^k = qB\epsilon^i{}_{jk} v^j \delta^k_3 \to qB \begin{pmatrix} v_y \\ -v_x \\ 0 \end{pmatrix}$$

Note that $\frac{\mathrm{d}v^3}{\mathrm{d}t} = 0$ since the B field is acting only in the xy plane. Then

$$\begin{pmatrix} \frac{\mathrm{d}v^x}{\mathrm{d}t} \\ \frac{\mathrm{d}v^y}{\mathrm{d}t} \\ \frac{\mathrm{d}v^z}{\mathrm{d}t} \end{pmatrix} = \frac{qB}{m} \begin{pmatrix} v^y \\ -v^x \\ 0 \end{pmatrix} \tag{4.19}$$

Rewriting everything as a 2D problem in the xy plane and remembering that $v_z = v_0$ is constant and equal to the initial value of the velocity, we have

$$\begin{pmatrix} \frac{\mathrm{d}v^x}{\mathrm{d}t} \\ \frac{\mathrm{d}v^y}{\mathrm{d}t} \end{pmatrix} = \frac{qB}{m} \begin{pmatrix} v^y \\ -v^x \end{pmatrix} \tag{4.20}$$

From (4.20), deriving again we get

$$\begin{pmatrix} \frac{\mathrm{d}^2v^x}{\mathrm{d}t^2} \\ \frac{\mathrm{d}^2v^y}{\mathrm{d}t^2} \end{pmatrix} = \frac{qB}{m} \begin{pmatrix} \frac{\mathrm{d}v^y}{\mathrm{d}t} \\ -\frac{\mathrm{d}v^x}{\mathrm{d}t} \end{pmatrix}$$

Connecting this one to (4.20) we have two identical differential equations

$$\begin{pmatrix} \frac{\mathrm{d}^2v^x}{\mathrm{d}t^2} \\ \frac{\mathrm{d}^2v^y}{\mathrm{d}t^2} \end{pmatrix} = \left(\frac{qB}{m}\right)^2 \begin{pmatrix} \frac{\mathrm{d}v^x}{\mathrm{d}t} \\ \frac{\mathrm{d}v^y}{\mathrm{d}t} \end{pmatrix} \tag{4.21}$$

Solving these equations and inserting the initial values we have the final solutions

$$\begin{pmatrix} v^x \\ v^y \\ v^z \end{pmatrix} = \begin{pmatrix} v_{0\perp}\cos(\omega t) \\ -v_{0\perp}\sin(\omega t) \\ v_0 \end{pmatrix} \tag{4.22}$$

And this is the equation of an elicoidal motion, as we expected.

§§ 4.2.3 Mechanical Actions on Circuits

We know already a formula for magnetic forces, given from Lorentz's force law. In its differential form we have *Laplace's second formula*

$$\mathrm{d}F^i = I\epsilon^i{}_{jk}\mathrm{d}l^j B^k \tag{4.23}$$

Where $\mathrm{d}l^i = \hat{t}^i \mathrm{d}l$.
We always consider here *steady* currents, therefore in all our circuits there will be current generators that keep I constant. For a rigid circuit we have that the total force exerted on the charges is

$$F^i = I\oint \epsilon^i{}_{jk}\mathrm{d}l^j B^k = I\oint \epsilon^i{}_{jk}\hat{t}^j B^k \mathrm{d}l \tag{4.24}$$

Note that if B is uniform then, for a closed circuit the total force exerted on the charges is 0

$$F^i = I\oint \epsilon^i{}_{jk}\hat{t}^j B^k \mathrm{d}l = I\epsilon^i{}_{jk}B^k \oint \hat{t}^j \mathrm{d}l = 0$$

Since

$$\oint \hat{t}^i \mathrm{d}l = 0$$

$$M^i = I\oint \epsilon^i{}_{jk} r^j \epsilon^k{}_{lm}\hat{t}^l B^m \mathrm{d}l \tag{4.25}$$

Note that it's the usual classical formula $\mathbf{M} = \mathbf{r}\times\mathbf{F}$ extended to the whole circuit

Example 4.2.1 (A Closed Circuit). Consider a closed rectangular rigid circuit with a current I, submerged in an uniform B field.
If the rectangle has sides a and b where $a \perp b$ and the sides b are the ones "above" and "below". If the circuit isn't normal to the B field, and denominating with F_1, F_4 the forces acting on the sides long a, while F_3, F_1 the ones acting on the sides long b, we have

$$F_1 = F_3 = BIb\cos\theta$$

(Remember that the total current on one cable is Il where the cable is long l)
While

$$F_2 = F_4 = BIa$$

Using the right hand rule for cross products we know then that F_1, F_3 must be coplanar and therefore don't contribute to a force couple, while F_2, F_4 do. The angular momentum M^i depends on the orientation of B, and therefore we car write

$$M = BIab\sin\theta = BIS\sin\theta$$

Where $S = ab$ is the surface area. Therefore, rewriting $B\sin\theta = \|\mathbf{B}\times\hat{\mathbf{n}}\|$ we have

$$M^i = IS\epsilon^i{}_{jk}\hat{n}^j B^k = \epsilon^i{}_{jk}m^j B^k$$

Via analogy we define the *magnetic dipole moment* of the circuit m^i as

$$m^i = IS\hat{n}^i \tag{4.26}$$

§ 4.3 Maxwell's Equations for Magnetostatics

§§ 4.3.1 Biot-Savart's Law

In general, experimentally it has been found that the magnetic field follows the right hand rule and depends in magnitude from the current density and the inverse of the distance from the magnetic object. Experimentally it has been found that for a wire with static current density $\mathrm{d}I^i = I\mathrm{d}l^i = I^i\mathrm{d}l$ the magnetic field is

$$B^i = \frac{\mu_0}{4\pi}\int_\lambda \frac{\epsilon^i{}_{jk} I^j \hat{r}^k}{r^2}\mathrm{d}l \tag{4.27}$$

This equation is known as the law of *Biot and Savart*. The constant μ_0 is a coupling constant known as the *permeability of free space* and has units of Newton/Ampere squared. It has the following (exact) value in the SI system

$$\mu_0 = 4\pi\cdot 10^{-7}\ \frac{\mathrm{N}}{\mathrm{A}^2} \tag{4.28}$$

Note that this holds dimensionally with the B from what we found before from Lorentz's force law.

Example 4.3.1 (A Live Wire). Consider a wire along the x axis where a stationary current goes from right to left. Using Biot-Savart we can find the field B^i at a distance s from the wire. We have that for the symmetries of the system we're working on a 2D plane xy. The radius vector from the current to the point where we're evaluating the field describes a triangle in this plane, we indicate the angle between the height of the triangle (s) and the radius vector r as θ. Using trigonometry we have that the length of the basis of this triangle is $l = s\tan\theta$, therefore

$$\mathrm{d}l = \frac{s}{\cos^2\theta}\mathrm{d}\theta$$

Also, we have that in this triangle considering the angle α between the current and the radius vector we have that the cross product of the two has the following magnitude

$$\|\mathrm{d}\mathbf{l}\times\hat{\mathbf{r}}\| = r\sin\alpha\mathrm{d}l = r\frac{\sin\alpha}{\cos^2\theta}\mathrm{d}\theta$$

Using the fact that we're in an euclidean triangle (duh) we must have that $\alpha+\theta+\pi/2=\pi$, therefore $\alpha = \pi-\theta$. Still using trigonometry we can also say that

$$s = r\cos\theta \implies r = \frac{s}{\cos\theta} \implies \frac{1}{r^2} = \frac{\cos^2\theta}{s^2}$$

From Biot-Savart then

$$\mathrm{d}B = \frac{\mu_0}{4\pi}\frac{\cos\theta}{r^2}\mathrm{d}l = \frac{\mu_0\cos\theta}{4\pi}\left(\frac{s}{\cos^2\theta}\frac{\cos^2\theta}{s^2}\right)\mathrm{d}\theta$$

Therefore

$$\mathrm{d}B = \frac{\mu_0\cos\theta}{4\pi s}\mathrm{d}\theta \tag{4.29}$$

Considering that θ goes from some θ_0 to some θ_1 if the wire is finite, and $\theta_0 = -\pi/2, \theta_1 = \pi/2$ for an *infinite* wire, we have in the second case

$$B(s) = \frac{\mu_0}{4\pi s}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cos\theta\mathrm{d}\theta = \frac{\mu_0}{2\pi s} \tag{4.30}$$

Using the right hand rule we can also find the direction of B, being always normal to both r and I we have that in 3 dimensions, using cylindrical coordinates

$$B^i = \frac{\mu_0}{2\pi s}\hat{\varphi}^i \tag{4.31}$$

§§ 4.3.2 Divergence and Curl of B

Consider now an infinite wire with a static current I^i going through it. If we take a closed loop C around the wire it's clear that if we integrate B in that loop the integral can't be 0. From Biot-Savart we have, using what we found previously for infinite wires

$$B = \frac{\mu_0 I}{2\pi r}\oint_C \mathrm{d}l = \mu_0 I \tag{4.32}$$

In general it's clear that for some loop C around multiple currents $I_{(i)}$, we will have that as for Gauss' law for electric fields with charges, that

$$\oint_C B^i \hat{t}_i \mathrm{d}l = \mu_0 I_{in} \tag{4.33}$$

Where I_{in} is the total current inside the loop. If we have a current density J^i it's clear that then,

$$I_{in} = \iint_S J^i \hat{n}_i \mathrm{d}s$$

Where S is the surface such that $\partial S = C$. Reconnecting and using Stokes' theorem on B we have

$$\iint_S \epsilon^i{}_{jk}\partial^j B^k \hat{n}_i \mathrm{d}s = \mu_0 \iint_S J^i \hat{n}_i \mathrm{d}s \tag{4.34}$$

We get then that *for straight wires*

$$\epsilon^i{}_{jk}\partial^j B^k = \mu_0 J^i \tag{4.35}$$

What about the divergence of B? We begin again with Biot-Savart law in a general setting with a volume current J^i. We have

$$B^i = \frac{\mu_0}{4\pi}\iiint_V \frac{1}{r^2}\epsilon^i{}_{jk}J^j \hat{r}^k \mathrm{d}^3x$$

We then apply the divergence operator to both sides

$$\partial_i B^i = \frac{\mu_0}{4\pi}\iiint_V \frac{\partial}{\partial x^i}\left(\epsilon^i{}_{jk}J^j\frac{\hat{r}^k}{r^2}\right)\mathrm{d}^3x$$

Using the antisymmetry of $\epsilon^i{}_{jk}$ we can rewrite the divergence of the cross product as follows

$$\frac{\partial}{\partial x^i}\epsilon^i{}_{jk}J^j\frac{\hat{r}^k}{r^2} = \frac{\hat{r}_i}{r^2}\epsilon^i{}_{jk}\partial^j J^k - J_i\epsilon^i{}_{jk}\partial^j\left(\frac{\hat{r}^k}{r^2}\right)$$

Remember: J^i doesn't depend from how far we're measuring the field, therefore $\partial_i J^k = 0$! The first part on the right hand side of the previous equation is zero, therefore we have

$$\frac{\partial}{\partial x^i}\epsilon^i{}_{jk}J^j\frac{\hat{r}^k}{r^2} = -J_i\epsilon^i{}_{jk}\partial^j\left(\frac{\hat{r}^k}{r^2}\right) = 0 \tag{4.36}$$

Where we used that the curl of $\hat{r}^i/r^2$ is 0 (it's the gradient of another function). Therefore we have that

$$\partial_i B^i = 0 \tag{4.37}$$

Note that this holds everywhere since we started from a generic current distribution! We can also recompute the curl of B from this generic current distribution. Remembering again that J depends only on the local coordinates of the distribution itself and not from how far we're measuring B we have, firstly, using the identities of $\epsilon^i{}_{jk}$, and juggling a bit

$$\epsilon^i{}_{jk}\partial^j\epsilon^k{}_{lm}J^l\frac{\hat{r}^m}{r^2} = J^i\partial_l\left(\frac{\hat{r}^l}{r^2}\right) - \left(J^l\partial_l\right)\frac{\hat{r}^i}{r^2} \tag{4.38}$$

Using the definition of the 3D Dirac delta for evaluating the first divergence we have

$$\epsilon^i{}_{jk}\epsilon^k{}_{lm}\partial^j J^l\frac{\hat{r}^m}{r^2} = 4\pi\delta^3(r)J^i - \left(J^l\partial_l\right)\frac{\hat{r}^i}{r^2} \tag{4.39}$$

Looking closely at the second side, we have

$$\frac{\partial}{\partial x^i}\left(J^i\frac{\hat{r}^j}{r^2}\right) = \frac{\hat{r}^j}{r^2}\partial_i J^i + \left(J^i\partial_i\right)\frac{\hat{r}^j}{r^2}$$

Noting that for steady currents $\partial_i J^i = 0$ we have then

$$\left(J^l\partial_l\right)\frac{\hat{r}^i}{r^2} = \frac{\partial}{\partial x^j}\left(J^i\frac{\hat{r}^j}{r^2}\right)$$

Reinserting everything into the definition of B^i we have

$$\epsilon^i{}_{jk}\partial^j B^k = \mu_0\iiint_V J^i\delta^3(r)\mathrm{d}^3x + \frac{\mu_0}{4\pi}\oiint_{\partial V}\frac{\hat{r}^i}{r^2}J^j\hat{n}_j\mathrm{d}s \tag{4.40}$$

The surface integral is safely equal to zero since all currents are safely inside the surface (by definition), and collapsing the first integral we have

$$\epsilon^i{}_{jk}\partial^j B^k = \mu_0 J^i \tag{4.41}$$

This relation is *general* and holds for all current distributions J^i. This is known formally as *Ampere's law for Magnetostatics*.

Analogously to Gauss' law for E, we have some tricks for evaluating B using Ampere's law. We draw some convenient closed loops for which B comes out easily from the integral. As we have Gaussian surfaces, we have for B *Amperian loops*.

Integrating both sides of (4.41) we have applying Stokes that if we chose a nice enough loop, i.e. one where B is parallel to the tangent of the loop it will come out from the integral and we will have on the right $\mu_0 I_{in}$, easing all calculations

§§ 4.3.3 The Parallel Between Magnetostatics and Electrostatics

We have now found 4 Maxwell equations, two for static charges and two for static currents. These equations are, in free space

$$\begin{cases} \partial_i E^i = \dfrac{\rho}{\epsilon_0} \\ \epsilon^i{}_{jk}\partial^j E^k = 0 \\ \partial_i B^i = 0 \\ \epsilon^i{}_{jk}\partial^j B^k = \mu_0 J^i \end{cases} \tag{4.42}$$

These equations are asymmetric! If we made a symmetric parallel between the two and supposed $\partial_i B^i = \mu_0 \rho_m$ where ρ_m is a "magnetic charge", the magnetic counterpart to the electric charge, these equation would be perfectly symmetric. The fact that $\partial_i B^i = 0$ always shows that magnetic charges, or magnetic monopoles, cannot exist in nature due to the actual shape of B.
This particular parallel, lets us think that there might also be a potential for B. We are already sure that it can't be a scalar potential in general, since the curl of B is nonzero.

§§§ 4.3.3.1 Magnetic Potentials

Let's look back at Biot-Savart's law, we have

$$B^i = \frac{\mu_0}{4\pi}\iiint_V \frac{1}{r^2}\epsilon^i{}_{jk}J^j\hat{r}^k \mathrm{d}^3x = \frac{\mu_0}{4\pi}\iiint_V \epsilon^i{}_{jk}J^j\partial^k\left(\frac{1}{r}\right)\mathrm{d}^3x$$

Using $\boldsymbol{\nabla}\times\mathbf{J} = 0$ we can use a little property of the epsilon symbol and bring outside the curl. We have

$$B^i = \frac{\mu_0}{4\pi}\epsilon^i{}_{jk}\frac{\partial}{\partial x_j}\iiint_V \frac{J^k}{r}\mathrm{d}^3x \tag{4.43}$$

I.e. we wrote B in terms of the curl of some other vector. This vector is the *magnetic vector potential* A^i and, writing

$$A^i = \frac{\mu_0}{4\pi}\iiint_V \frac{J^i}{r}\mathrm{d}^3x \tag{4.44}$$

It's clear that since the curl of a gradient is zero, this potential is defined up to gradients of continuous functions.
We will usually choose the *Coulomb Gauge*, where $\partial_i A^i = 0$, this means that since

$$A^i = A^i_0 + \partial^i\lambda \implies \partial_i A^i = 0 \iff \partial^i\partial_i\lambda = -\partial_i A^i_0$$

This is a Poisson equation and the solution is easily

$$\lambda(r) = \frac{1}{4\pi}\iiint_V \frac{\partial_i A^i_0}{r}\mathrm{d}^3x$$

In this potential formulation, if we take Ampere's law we have

$$\epsilon^i{}_{jk}\partial^j B^k = \epsilon^i{}_{jk}\partial^j\epsilon^k{}_{lm}\partial^l A^m = \partial_i\partial_l A^l - \partial^l\partial_l A^i = \mu_0 J^i \tag{4.45}$$

Taking the previous gauge choice we have 3 Poisson equations for each component of A

$$\partial_l \partial^l A^i = -\mu_0 J^i \tag{4.46}$$

If $J(r) \to 0$ we have then that the solution is, as before

$$A^i = \frac{\mu_0}{4\pi} \iiint_V \frac{J^i}{r} \mathrm{d}^3 x$$

§§ 4.3.4 Boundary Conditions for the Magnetic Field

Going back to electrostatic boundary conditions, we have that E is discontinuous on the surface *charge*. Instead, B, by analogy of the formulas is discontinuous on surface *currents*.
Checking again Maxwell's equations, we must have that if E is discontinuous on the *normal* to the surface, B must be discontinuous on the *tangent*, in fact, taken a pillbox P that goes through this surface current, then, since $\partial_i B^i = 0$ *always* we have

$$\iint_P B^i \hat{n}_i \mathrm{d}s = 0 \tag{4.47}$$

This immediately implies, noting that the normal vector above is opposite to the normal vector below the surface, that

$$B^i \hat{n}_i^{ab} = B^i \hat{n}_i^{bel} \tag{4.48}$$

I.e., the normal components of the magnetic field are continuous.
Take now an Amperian loop that goes through the surface. Then, by Ampere's law we get

$$\oint_A B^i \hat{t}_i \mathrm{d}l = \mu_0 I_{enc}$$

Where $I_{enc} = KI$ if K is our surface current density. Therefore, expanding the integral

$$B^i \hat{t}_i^{ab} - B^i \hat{t}_i^{bel} = \mu_0 K \tag{4.49}$$

Note that this holds since if we check the parallel components of the tangent of this loop, we must have $I_{enc} = 0$. Both these conditions can be summarized into a single equation.
Let K^i be the surface current density, then if we measure the magnetic field above (B^i_{ab}) and below B^i_{bel}, it must be continuous passing through the current surface and discontinuous on its tangent, i.e.

$$B^i_{ab} - B^i_{bel} = \mu_0 \epsilon^i{}_{jk} K^j \hat{n}^j \tag{4.50}$$

§ 4.4 Magnetic Dipoles

As for electrostatic potentials, it's possible to expand the magnetic potential into multipoles.
Remembering that if $r^i = r^i - \tilde{r}^i$, we can write r^{-1} in terms of Legendre functions

$$\frac{1}{r} = \frac{1}{\sqrt{r^2 + (\tilde{r})^2 - 2r\tilde{r}\cos\theta}} = \sum_{l=0}^{\infty} \left(\frac{\tilde{r}}{r}\right)^2 P_l(\cos\theta)$$

Therefore, for any Amperian loop A with current $I\hat{t}^i$ we have

$$A^i(r) = \frac{\mu_0 I}{4\pi}\sum_{l=0}^{\infty}\frac{1}{r^{l+1}}\oint_A (\tilde{r})^l\, P_l(\cos\theta)\hat{t}^i \mathrm{d}l \tag{4.51}$$

Expanding till quadrupoles ($l=2$) we then can write

$$A^i \approx \frac{\mu_0 I}{4\pi r}\oint_A \hat{t}^i \mathrm{d}l + \frac{\mu_0 I}{4\pi r^2}\oint_A \tilde{r}\cos\theta\hat{t}^i\mathrm{d}l + \frac{\mu_0 I}{4\pi r^3}\oint_A \frac{(\tilde{r})^2}{2}\left(3\cos\theta - 1\right)\hat{t}^i\mathrm{d}l \tag{4.52}$$

Note that the monopole term ($l=0$) is zero, as we expected! Again, there are no magnetic monopoles. We concentrate on the $l=1$ term, the dipole term. We have then that the magnetic dipole is

$$A^i_{dip}(r) = \frac{\mu_0 I}{4\pi r^2}\oint_A \tilde{r}\cos\theta\hat{t}^i\mathrm{d}l \tag{4.53}$$

Remembering that by definition $\hat{r}^i\hat{\tilde{r}}_i = \cos\theta$ we can write everything in terms of scalar products. Since

$$\tilde{r}\cos\theta = \tilde{r}\hat{r}$$

We can write the integral in terms of a surface integral. Call S the surface enclosed by the loop, then

$$A^i_{dip}(r) = \frac{\mu_0 I}{4\pi r^2}\oint_A (\tilde{r}^j\hat{r}_j)\hat{t}^i\mathrm{d}l = \frac{\mu_0 I}{4\pi r^2}\left(-\epsilon^i{}_{jk}\hat{r}^j\iint_S \hat{n}^k\mathrm{d}s\right)$$

Remembering the definition of magnetic dipole, and switching the indexes on the cross product, we have

$$A^i_{dip}(r) = \frac{\mu_0}{4\pi r^2}\epsilon^i{}_{jk}m^j\hat{r}_k \tag{4.54}$$

Look how this is clearly equivalent to the field generated by a small dipole m^i generated by a current loop on a circuit A.

§§ 4.4.1 Interaction Between Currents

Consider two rigid circuits with some current I_i, $i=1,2$. Suppose these two circuits are small enough so that we can consider them as two magnetic dipoles.
In the second circuit, there is a force $\mathrm{d}F_{21}$ generated from the B_1 field of the first circuit. We have

$$\mathrm{d}F^i_{21} = I_2\epsilon^i{}_{jk}\mathrm{d}l^j_2 B^k_1 = I_2\epsilon^i{}_{jk}\mathrm{d}l^j_2\oint_1\frac{\mu_0 I_1}{4\pi r^2_{12}}\epsilon^k{}_{lm}\mathrm{d}l^l_1\hat{r}^k_{12}$$

Where we took r_{12} as the distance between the two circuit elements $\mathrm{d}l_i$. The total force will be then, integrating

$$F^i_{12} = \frac{\mu_0 I_1 I_2}{4\pi}\oint_1\oint_2\frac{1}{r^2_{12}}\epsilon^i{}_{jk}\epsilon^k{}_{lm}\mathrm{d}l^j_2\mathrm{d}l^l_1\hat{r}^m_{12} = \frac{\mu_0 I_1 I_2}{4\pi}\oint_1\oint_2\left(\mathrm{d}l^i_1\left(\frac{\hat{r}^l_{12}}{r^2_{12}}\mathrm{d}l^2_l\right) - \frac{\hat{r}^i_{12}}{r^2_{12}}\mathrm{d}l^l_1\mathrm{d}l^2_l\right) \tag{4.55}$$

Where we used the properties of $\epsilon^i{}_{jk}$ for computing the products. On the right hand side the first term can be seen as the curl of a vector (with respect to the 2nd line integral) integrated on the surface enclosed by the second loop. This curl is zero, and therefore we have finally

$$F_{12}^i = -\frac{\mu_0 I_1 I_2}{4\pi} \oint_1 \oint_2 \frac{\hat{r}_{12}^i}{r_{12}^2} \mathrm{d}l_1^l \mathrm{d}l_l^2 \tag{4.56}$$

Note that $F_{12} = -F_{21}$ simply because $r_{12} = -r_{21}$! Therefore what we found respects Newton's third law.
Consider now two infinite parallel wires with the same currents, we have that if $\mathrm{d}l_l^1 \mathrm{d}l_2^l \geq 0$ or $\mathrm{d}l_l^1 \mathrm{d}l_2^l \leq 0$ (i.e. the two currents have the same direction or opposite direction), we will get wither an attractive force or a repulsive force.
It's clear that this force will be orthogonal to the wires (see Lorentz's force law) and therefore we're interested to check only the orthogonal component.
If the two parallel wires are distant a from each other we can write immediately $r_{12}\cos\theta = a$, and $\mathrm{d}l_1 = \frac{a}{\cos^2\theta}\mathrm{d}\theta$. Noting that the two wires are infinite, using the transformation $\mathrm{d}l_1 \to \mathrm{d}\theta$ we get (dividing by $\mathrm{d}l_2$)

$$\frac{\mathrm{d}F_{21}^\perp}{\mathrm{d}l_2} = -\frac{\mu_0 I_1 I_2}{4\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left(\frac{\cos^2\theta}{a^2}\right) \cos\theta \left(\frac{a}{\cos^2\theta}\mathrm{d}\theta\right) = -\frac{\mu_0 I_1 I_2}{4\pi a} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos\theta \mathrm{d}\theta$$

Integrating, we immediately get

$$\frac{\mathrm{d}F_{12}^\perp}{\mathrm{d}l_2} = -\frac{\mu_0 I_1 I_2}{2\pi a} \tag{4.57}$$

Note this is negative only if $\mathrm{d}l_1^l \mathrm{d}l_l^2 > 0$, i.e. when the currents are parallel.

5 Magnetism in Matter

If we insert some material in a region where there is a B field there are three observed effects

1. Mechanical forces on the body are observed
2. The field around the bodies is modified by their presence
3. The bodies can be *magnetized*, i.e. they behave like a magnet

If we take as our experimental test field the one produced by a solenoid (a conductive spring where charges move in a closed loop) it can be verified immediately that all substances are distinguishable in three categories

1. Ferromagnets, which get attracted by the B field of the solenoid
2. Paramagnets, which get *weakly* attracted by the field
3. Diamagnets, which get *weakly* repulsed by the field

All these different behaviors are directly correlated from macroscopic proprieties.
Atoms themselves can be thought as small loop circuits (imagine electrons "going around" the nucleus), and therefore generate some magnetic dipole m^i. These dipoles interact with the field and tend to orient themselves in the same direction as B, i.e. the bodies get *magnetized*.

§ 5.1 Magnetization

The discussion of magnetism in matter is similar to the one on electricity in matter, and therefore it's good practice to begin with a microscopic approach to the problem.
Consider a small Hydrogen atom, one proton and one electron. Since $m_p \approx 2000 m_e$ we can consider the nucleus as locked in place, while electrons move around in a circular orbit with radius r_0.
The electron experiences the following centripetal coulomb force

$$F_c = m_e a_c = \frac{1}{4\pi\epsilon_0}\frac{e^2}{r_0^2}$$

I.e., using $a_c = v_0^2/r_0 = \omega_0^2 r_0$ we get

$$m_e \omega_0^2 r_0 = \frac{1}{4\pi\epsilon_0}\frac{e^2}{r_0^2}$$

Since $\omega_0 = 2\pi T_0^{-1}$ with T_0 being the period of the orbit we get

$$T_0 = \frac{4\pi}{e}\sqrt{\pi\epsilon_0 m_e r_0^3}$$

All these calculations are needed to find the magnetic dipole momentum of the electron, m^i. Using a bit of quantum mechanics and remembering that the electron is in a bound state ($E < 0$) we can find r_0 using the ionization energy (i.e. the energy needed to bring the electron to $r = \infty$ with $v = 0$) we have that

$$E = -I_e \implies r_0 = \frac{1}{8\pi\epsilon_0}\frac{e^2}{I} \approx 0.5$$

Note that we used *experimental data* (for now). In the setup we made we basically made a toy hydrogen atom, for which $I = 13.6\ \mathrm{eV}$, where $1\ \mathrm{eV} = 1.6 \times 10^{-19}\ \mathrm{C \cdot V} = 1.6 \times 10^{-19}\ \mathrm{J}$.
With this we get $T_0 \approx 1.5 \times 10^{-16}\ \mathrm{s}$ as our orbit period, and the current associated with a single electron moving around a proton (a simple toy atom, hydrogen in this case) is

$$I = \frac{e}{T_0} \approx 1\ \mathrm{mA}$$

Using $m^i = IS\hat{n}^i$ we have that the magnetic momentum of this system is

$$m = I\pi r_0^2 = \frac{e\pi r_0^2}{T_0} = 9.35 \times 10^{24}\ \mathrm{A \cdot m}$$

And the angular momentum is

$$L = m_e v_0 r_0 = m_e \frac{2\pi r_0^2}{T_0} \implies \frac{m}{L} = \frac{e}{2m_e}$$

The last constant is known as the *gyromagnetic factor* g of the electron and is a general result also valid in quantum mechanics.
Writing $L = \hbar l$ in a semiclassical fashion (you'll understand later, probably, or you already know) we get a new fundamental constant tied to the gyromagnetic factor g

$$m = gL = \hbar g l = \frac{\hbar e}{2m_e} l = \mu_B l$$

Where μ_B is known as *Bohr's magneton*, for which $\mu_B \approx 9.27 \cdot 10^{-24}\ \mathrm{A \cdot m^2}$

§§ 5.1.1 The Magnetization Field

After the small "quantum" digression, we can get back to our classical treatment of Electrodynamics. We've seen that all atoms must have a magnetic dipole moment m^i tied to the "orbital" nature of bound electrons in nuclear fields. Analogously to dipole moments in dielectrics this must determine the magnetic properties in matter.
We define the *Magnetization irradiance* M^i as follows

$$M^i = \lim_{\Delta V \to 0} \frac{\Delta N}{\Delta V} \left\langle m^i \right\rangle \tag{5.1}$$

Where ΔN is the numerical density of atoms.
In SI units we have

$$[M] = \frac{\mathrm{A}}{\mathrm{m}}$$

And rearranging a bit the previous terms, and using $\Delta V \to \mathrm{d}V$

$$\mathrm{d}m^i = M^i \mathrm{d}V \tag{5.2}$$

We begin by considering an uniform magnetization M^i inside a magnetized medium. It's clear that indide the body all the currents will cancel out and we'll be left only with surface effects, which will be magnetization-induced currents that will follow the right hand rule since there's no compensation outside the magnet.
Obviously, if M^i is not uniform, we will also have volumetric currents. Surface currents will be indicated with J_{ms} and volumetric currents with J_{mv}.
Using equation (4.54) we can see the relations between M^i and these currents. remembering equation (5.2) we can write for a magnetized body V

$$A^i = \frac{\mu_0}{4\pi} \int_V \frac{1}{r^3} \epsilon^i{}_{jk} M^j r^k \mathrm{d}^3 x' \tag{5.3}$$

Bringing $1/r^3$ inside the cross product and remembering that $r^i/r^3 = -\partial_i(r^{-1})$ and then applying a simple vector analysis identity ($\mathbf{v} \times \nabla f = f\nabla \times \mathbf{v} - \nabla \times (f\mathbf{v})$) we get two integrals

$$A^i = \frac{\mu_0}{4\pi} \left(\int_V \frac{\epsilon^i{}_{jk} \partial^j M^k}{r} \mathrm{d}^3 x' - \int_V \epsilon^i{}_{jk} \partial^j \left(\frac{M^k}{r} \right) \mathrm{d}^3 x' \right) \tag{5.4}$$

Using $\int_V \nabla \times \mathbf{v} \mathrm{d}V = -\int_{\partial V} \mathbf{v} \times \hat{n} \mathrm{d}s$ on the second one we get

$$A^i = \frac{\mu_0}{4\pi} \int_V \frac{\epsilon^i{}_{jk} \partial^j M^k}{r} \mathrm{d}^3 x' + \frac{\mu_0}{4\pi} \int_{\partial V} \frac{\epsilon^i{}_{jk} M^j n^k}{r} \mathrm{d}s' \tag{5.5}$$

Since the vector potential has an unique solution (it's defined from a Poisson equation with well defined conditions) We can interpret the first curl as our volumetric current density and the second cross product as our surface current densities, giving us the relations between the magnetization currents and the magnetization vector M^i

$$\begin{aligned} J^i_{mv} &= \epsilon^i{}_{jk} \partial^j M^k \\ J^i_{ms} &= \epsilon^i{}_{jk} M^j \hat{n}^k \end{aligned} \tag{5.6}$$

§ 5.2 Maxwell Equations for Magnetostatics in Magnetic Media

Taking back what we found for the B^i field we can try to build up again the Maxwell equation for magnetostatics in magnetized media.
As we already have found we have

$$\begin{aligned} \partial_i B^i &= 0 \\ \epsilon^i{}_{jk} \partial^j B^k &= \mu_0 J^i \end{aligned}$$

We now must consider that J^i indicates the total current, so we will consider it as the sum of "free" extra currents J^i_f and the previously found magnetization currents J^i_m.
Inside the magnetized volume V we can replace J^i_m with the curl of M^i and, bringing it to the left we can write guardando il forno e una

$$\epsilon^i{}_{jk}\partial^j\left(\frac{B^k}{\mu_0} - M^k\right) = J^i$$

We can define an auxiliary field inside this curl, which we will call the "magnetic field" H^i

$$H^i = \frac{B^i}{\mu_0} - M^i \tag{5.7}$$

Rewriting everything, we get Maxwell's equation for magnetostatics in media

$$\begin{cases} \partial_i B^i = 0 \\ \epsilon^i{}_{jk}\partial^j H^k = J^i_f \end{cases} \tag{5.8}$$

These equations can be solved only if we know the functional relations between B and H or M and B, or if we manage to find some conditions that can help us

§§ 5.2.1 Boundary Conditions

In order to solve these equations tho we need to consider what happens at the surface ∂V of the body. Suppose that we have two magnetized bodies separated by a surface S_s. Taken a small loop l on this separation surface, which encompasses a surface S, we can use the second equation of (5.8) we get

$$\oint_l H^i t_i \mathrm{d}l = \iint_S \epsilon^i{}_{jk}\partial^j H^k \hat{n}_i \mathrm{d}s = \iint_S J^i \hat{n}_i \mathrm{d}s = \sum_i I_i \tag{5.9}$$

Therefore, the closed line integral of H^i is the sum of the (free) currents enclosed by the loop.
Considering the same loop in the case where there are no free currents, equations (5.8) give the boundary conditions for B and H in matter.

$$\begin{cases} B_{n_1} = B_{n_2} \\ H_{t_1} = H_{t_2} \end{cases} \tag{5.10}$$

Where n_i, t_i are the normal and tangent components of the field between substance 1 and 2.
By definition of H^i we can see already that in vacuum $B^i_0 = \mu_0 H^i_0$ since $M^i = 0$. As for dielectrics in isotropic and homogeneous substances we can write $B^i = \mu H^i$ with $\mu_0 = \mu_0\mu_r$ where μ_r is the relative magnetic permeability.
For anisotropic substances μ can be described as a rank 2 tensor. Contrary to dielectrics, μ in general depends from the B field irradiance, and is constant only for diamagnetic or paramagnetic substances. For ferromagnets $\mu = \mu(B)$.
With this definition, we can calculate the magnetization field of the body. We have $B = \mu H$ therefore

$$H^i = \mu_r H^i - M^i \implies M^i = (\mu_r - 1)H^i = \chi_m H^i \implies \mu M^i = \chi_m B^i \tag{5.11}$$

Where $\mu_r - 1 = \chi_m$ is the magnetic susceptibility.
Inserting that back to the definition of H^i we have

$$B^i = \mu_0 (1 + \chi_m) H^i \tag{5.12}$$

By definition, the value of χ_m defines the alignment of the magnetization with respect to the magnetic field. In general for values of χ_m between 10^{-5} to 10^{-3} we have an orientation of atomic magnetic dipoles and therefore paramagnetism.
For negative values we get diamagnetic effects and for very big positive effects we get ferromagnetic effects

§ 5.3 Ferromagnets and Hysteresis Cycles

In ferromagnets ($\chi_m >> 1$) the dependence $B(H)$ or $M(H)$ is really complex and the relations aren't unique and can change a lot for small changes on composition of the material.
For analyzing it we start with the unmagnetized material ($H = B = M = 0$) and place it inside a solenoid, for which we know already that, thanks to the Maxwell equations that $H = nI$, with n being the number of loops of the solenoid and I being the total current of the solenoid.
Changing I we have that B changes way quicker than H, with a strong contribution from the magnetization of the element through the relation

$$B = \mu_0 H + \mu_0 M$$

The growth is exponential until a saturation H_s value is reached. This growth is known as the "first magnetization curve". After this value the growth of B is linear in H till a maximum H_m due to a saturation in M, which reaches a saturation maximum M_s.
Shutting the current off ($I = 0$) we get to $H = 0$ and a residue magnetic induction field B_r can be measured.
Inverting the current's direction B goes down till 0, for $\mu_0 H = -\mu_0 M$, i.e. the magnetic field H reaches the coercive magnetic field value where $H_c = -M_c$. From here on, the fields quickly reaches a negative minimum at $H = -H_m$.
Making H grow again from the minimum the field B will reach $-B_r$ at $H = 0$ and will reconnect to the first cycle maximum at H_m.

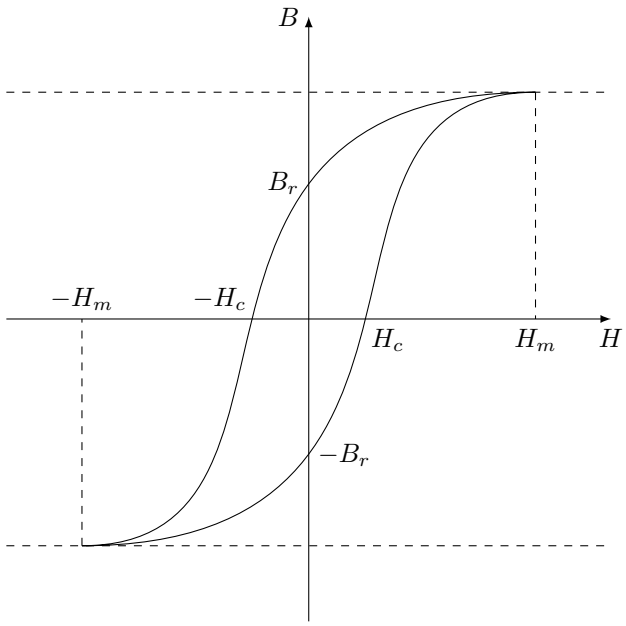


Figure 5.1: Example of an hysteresis cycle, without the first magnetization curve.

This full cycle is known as the magnetic hysteresis cycle and if it's drawn it's clear that $B(H)$ is not a function in the proper sense of it, since the value depends on what happened before to the material, and in general we have that

$$\mu(H) = \frac{B}{H} \tag{5.13}$$

It's also possible to draw a demagnetization cycle on the $B - H$ plane from any point making smaller and smaller hysteresis cycles, and with a simple analogy to $p - V$ planes in thermodynamics, one can calculate the work made per unit volume of the material with the relation $\mathrm{d}W = B\mathrm{d}H$ (remember that in thermodynamics $\mathrm{d}W = p\mathrm{d}V$ when there is no external work acting on the system).
From the relationship we found before for B one can write the differential magnetic permeability of a body as

$$\mu_d = \frac{\mathrm{d}B}{\mathrm{d}H}$$

Or its relative counterpart

$$\mu_{d_r} = \frac{\mu_d}{\mu_0} = \frac{1}{\mu_0}\frac{\mathrm{d}B}{\mathrm{d}H}$$

Another experimental result on ferromagnets is the *law of Curie-Weiss*, which states that for temperatures over a critical value T_c, a ferromagnet becomes a paramagnet, and its susceptibility goes as

$$\chi_m = \frac{k\rho}{T - T_c} \tag{5.14}$$

Where k is a constant and ρ is the material's density

§ 5.4 Local Magnetic Field

For evaluating the local counterpart of the magnetic field, since we can consider ourselves in vacuum, we're free to choose between using B and $H = B/\mu_0$. For notational ease H is "better".
Using the same exact path taken to find the Lorentz local field in dielectrics (3.22) we can say that the magnetic field around some atom, at its center is

$$H^i_{loc} = H^i + \frac{1}{3}M^i \tag{5.15}$$

This local field considers that the contribute of all the small dipoles inside the sphere around the atom sum to zero.
Since for paramagnetic and diamagnetic substances $M << H$ we could even write $H_{loc} \approx H$.
This doesn't hold for ferromagnets, and thanks to Weiss we get a reformulation of the local field

$$H^i_{loc,fm} = H^i + \gamma M^i \tag{5.16}$$

The constant γ is known as Weiss' constant, and $10^3 < \gamma < 10^4$. It has been justified by considering the ferromagnet as divided in multiple sectors where atomic dipoles have zones of common orientations, where the biggest zone is the one oriented with the magnetic field H. The zone engulfs the whole magnet then slowly.

§§ 5.4.1 Larmor Precession

Consider now a single atom, completely unaligned with the field. We have already found that its magnetic moment is

$$m^i_0 = -\frac{e}{2m_e}L^i$$

And its torque is

$$\tau^i = \epsilon^i{}_{jk} m^j_0 B^k_{loc}$$

With B_{loc} being our local B field.
By definition of torque τ we have

$$\frac{\mathrm{d}L^i}{\mathrm{d}t} = \epsilon^i{}_{jk} m^j_0 B^k_{loc} = \frac{e}{2m_e}\epsilon^i{}_{jk} B^j_{loc} L^k = \epsilon^i{}_{jk}\omega^i_L L^k \tag{5.17}$$

The last result gives us the Poisson formula for L, which indicates that it completes a precession motion with angular velocity ω_L, known commonly as Larmor precession. This speed is by definition parallel to the local field, and it's associated to a current given by this precession and the charged nature of the electron

$$I_L = -\frac{e}{T_L} = -\frac{e\omega_L}{2\pi} \tag{5.18}$$

This current is therefore tied to a magnetic moment, for which $m_L = I_L \tilde{S}$ where $\tilde{S}$ is the area of the orbit of the electron projected onto the same direction of the local field. Inserting a bit of numbers in the previous statement we have

$$m^i_L = -\frac{e}{2\pi}\tilde{S}\omega^i_L = -\frac{e^2}{4\pi m_e}\tilde{S}B^i_{loc}$$

Averaging the coordinates of the position of the electron we know already that

$$\langle x\rangle^2 = \langle y\rangle^2 = \langle z\rangle^2$$
$$\langle x\rangle^2 + \langle y\rangle^2 + \langle z\rangle^2 = \langle r\rangle^2$$

Therefore

$$\tilde{S} = \pi\left(\langle x\rangle^2 + \langle y\rangle^2\right) = \frac{2\pi}{3}\langle r\rangle^2$$

Therefore

$$m_L^i = -\frac{e^2}{6m_e}\langle r\rangle^2 B_{loc}^i \tag{5.19}$$

Summing for all Z electrons in an atom, and remembering that the average radius of an electron is a_B, the Bohr radius

$$m_L^i = -\frac{Ze^2a^2}{6m_e}B_{loc}^i \tag{5.20}$$

Note that we used $\omega_0 << \omega_L$ as an approximation, together with $B_l << 4\pi m_e T_0^{-1} e^{-1}$ ($B_{loc} << 5\cdot 10^5$ T), which is almost always verified.
This intrinsic atomic moment is always present by definition, and it always opposes the local field

§§§ 5.4.1.1 Microscopic Interpretation of Diamagnets

Considering atoms where there is no atomic magnetic moment we have only Larmor effects, and by definition therefore the magnetization will be

$$M^i = nm_L^i = -\frac{n\mu_0 Ze^2a^2}{6m_e}H_{loc}^i = \alpha_d H_{loc}^i \tag{5.21}$$

By definition $\alpha_d << 1$ and therefore, using (5.15) we write

$$M^i = \frac{3\alpha_d}{3-\alpha_d}H^i \approx \alpha_d H^i \tag{5.22}$$

By definition $M^i = \chi_m H^i$, i.e. $\chi_m \approx \alpha_d < 0$. This susceptibility doesn't depend on the temperature, is negative and for reasonable values of a, Z, n $\chi_m \approx -10^{-5}$ as we said before for diamagnets

§§ 5.4.2 Langevin Function

Going back to substances where its composing atoms have their own atomic magnetic moment m_0^i, we have that thermal agitation tends to bring them to a disorder in their orientation.
In order to evaluate this Langevin proposed to utilize a function which could be used to evaluate the average magnetic momentum. Called L this Langevin function we have

$$\langle m^i\rangle = \langle m_0^i\rangle L(y) \tag{5.23}$$

Where

$$L(y) = \coth(y) - \frac{1}{y} = \coth\left(\frac{m_0^i B_i^{loc}}{kT}\right) - \frac{kT}{m_0^i B_i^{loc}} \tag{5.24}$$

By definition we have that this function is limited at $\pm\infty$ by ± 1 and it's uneven ($L(y) = -L(-y)$)

§§§ 5.4.2.1 Paramagnets

For paramagnets we have atoms (or molecules, as always) with proper magnetic moment m_0, but in general $m_0 B_{loc} << kT$, i.e. $y << 1$ and we can use a power series approximation on Langevin's function at the first order, which implies the following statements

$$\langle m \rangle \approx m_0 \frac{y}{3} = \frac{m_0^2 \mu_0}{3kT} H_{loc} \implies M = \frac{n m_0^2 \mu_0}{3kT} H_l = \alpha_p H_{loc} \tag{5.25}$$

Using $\chi_m \approx \alpha_p$ and writing the number density of atoms $n = \rho N_A / A$ we have

$$\chi_m(T) = \frac{\rho N_A m_0^2 \mu_0}{3k} \frac{1}{T} \tag{5.26}$$

Which is Curie's law that we defined before, with the constant written out in full in this classical view of microscopic electromagnetism

§§§ 5.4.2.2 Ferromagnets

For ferromagnets the approximation $y << 1$ doesn't hold anymore since m_0 is big, and using Weiss' law for ferromagnets (5.16) and the definition of magnetization, remembering that $L(\infty) = 1$ indicates the saturation of the magnet, we have that the saturation magnetization will simply be $M_s = n m_0$, and we'll get

$$\begin{cases} M(y, H) = M_s L(y) \\ M(y, H) = \dfrac{kT}{m_0 \mu_0 \gamma} y - \dfrac{H}{\gamma} \end{cases} \tag{5.27}$$

Plotting the first equation we get the magnetization in terms of the parameter y, which looks something like this

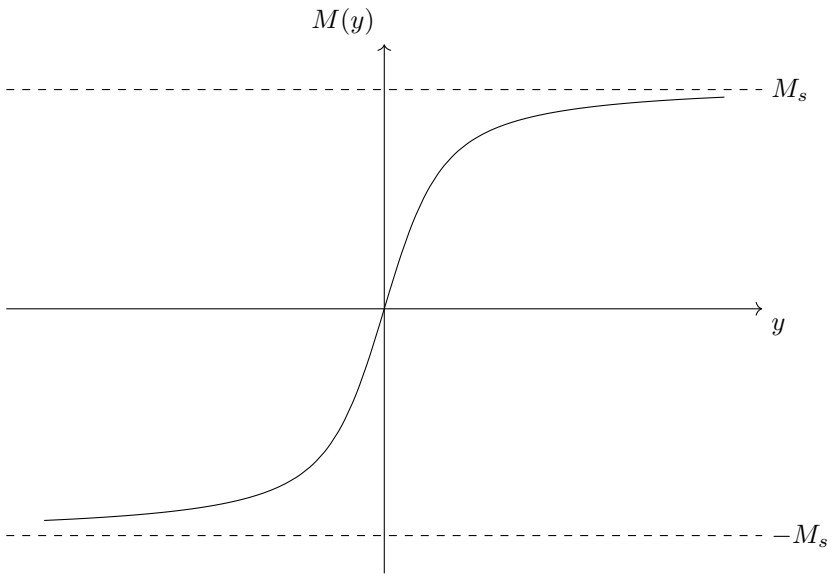


Figure 5.2: Langevin function for Magnetization

The second equation is a line tangent to $ML(y)$ at $y = 0$ intersecting the M axis at $-H/\gamma$, increasing the field H the intersection moves towards the right up until $M = M_s$.
Lowering the field until $H = H_c$ the line becomes tangent to $M_c L(y)$ for which we get two new intersections.
Inverting H (therefore I) the line reaches first M_c then $-M_s$, describing a magnetic hysteresis cycle (not drawn).
Reconsidering he system (5.27) we see that the line has angular coefficient $kT/m_0\mu_0\gamma$, therefore if T is high enough this coefficient is higher than $M_s/3$ of Langevin's curve $M(y)$. In this particular case there is only one intersection point, and therefore the substance becomes paramagnetic (there cannot be an hysteresis cycle).
Considering the derivative of the first and the second we have

$$\frac{kT}{m_0\mu_0 T} \geq \frac{M_s}{3} \implies T \geq \frac{\mu_0\gamma m_0 M_s}{3k} = T_c$$

Which gives the Curie temperature definition again. For $T > T_c$ we can approximate $L(y) \approx y/3$, therefore

$$\begin{aligned} M &= \frac{M_s y}{3} = \frac{n m_0^2 \mu_0}{3kT} H_{loc} = \frac{T_c}{\gamma T} H_{loc} \\ H_{loc} &= H + \gamma M \end{aligned} \tag{5.28}$$

Inserting the second equation in the first we have

$$M = \frac{T_c}{\gamma T}(H + \gamma M) = \frac{T_c}{\gamma (T - T_c)} H$$

Using again $\chi_m = M/H$ we have

$$\chi_m = \frac{1}{\gamma}\frac{T_c}{T - T_c} = \frac{\mu_0 n m_0^2}{3kT_c}\frac{T_c}{T - T_c}$$

And we get via simple algebra the Curie-Weiss law for ferromagnets

$$\chi_m(T) = \frac{\mu_0 m_0^2 n}{3k}\frac{1}{T - T_c}$$

Part III

Electrodynamics and Optics

6 Maxwell's Equations

§ 6.1 Faraday's Law

So far we managed to build four equations for the two fields, in *static* conditions. These are, whenever there no dielectrics and no magnets, are

$$\begin{cases} \partial_i E^i = \dfrac{\rho}{\epsilon_0} \\ \partial_i B^i = 0 \\ \epsilon^i{}_{jk}\partial^j E^k = 0 \\ \epsilon^i{}_{jk}\partial^j B^k = \mu_0 J^i \end{cases} \tag{6.1}$$

Faraday in his works had a major discovery while using two simple circuits.
Consider a closed circuit A with only a galvanometer, and a circuit B with a battery and a switch.
Using the switch for controlling the current flow on the circuit B (which is NOT connected to circuit A), he saw that the galvanometer measures a current while the current in B is changing.
Taking the same setup but keeping the switch closed on B, if either of the two circuits are in motion, there is current flow in A. The same happens if circuit B is substituted by a magnet!
Now take a non-rigid circuit immersed in a region with a constant magnetic field B^i. Deforming the circuit will also induce a current flow on it.
Calling f_{em} the *electromotive force* that drives the current in circuit A, Faraday deduced experimentally that

$$f_{em} = -\frac{\mathrm{d}\Phi}{\mathrm{d}t} \tag{6.2}$$

Where Φ is the magnetic flux passing inside the circuit.
Note that we know, by definition, that this electromotive force must be driven by an electric field E^i, where

$$f_{em} = \oint_C E^i_{ind}\hat{t}_i \mathrm{d}l \tag{6.3}$$

This field *can't* be conservative! Using the definition of the electric field as force per unit charge, and using Lorentz's force law we have a little hint that this induced field is the sum of a pure electric field plus a second field generated by the movement of charges

$$E^i_{ind} = E^i + \epsilon^i{}_{jk} v^j B^k \tag{6.4}$$

Noting that the charges are constrained to the circuit, we can divide the velocity v^i with a component parallel to the circuit $v^i_\parallel$ and a perpendicular component $v^i_\perp$. It's obvious then that

$$f_{em} = \oint_C \left(E^i + \epsilon^i{}_{jk} v^j_\perp B^k \right) \hat{t}_i \mathrm{d}l \tag{6.5}$$

Suppose now that we do not move the circuit, then $v^i = v^i_\parallel$ and $E^i_{ind} = E^i$, where this electric field is for sure not conservative.
All this jargon, condenses itself in one simple but powerful law, *Faraday-Neumann-Lenz's law*, which indicates exactly what Faraday discovered experimentally

Theorem 6.1 (Faraday-Neumann-Lenz, Electromagnetic Induction). *Given a time-dependent magnetic field $B^i(t, x^i)$, an electric field is induced by the variation of its flux, where*

$$\oint_{\partial S} E^i \hat{t}_i \mathrm{d}l = -\frac{\mathrm{d}}{\mathrm{d}t} \iint_S B^i \hat{n}_i \mathrm{d}s$$

Or, in its differential counterpart

$$\epsilon^i{}_{jk} \partial^j E^k = -\frac{\partial B^i}{\partial t} \tag{6.6}$$

Proof. Suppose that there is some circuit ∂S that spans some surface S inside of it, which is immersed in a time dependent magnetic field $B^i(t)$, then (6.2) holds, and therefore

$$\frac{\mathrm{d}\Phi}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t} \iint_{S(t)} B^i(x,t) \hat{n}_i \mathrm{d}s$$

The derivative on the right can be seen as the variation of the surface $S(t)$ when the B field is fixed in time at t_0, plus the integral over the surface $S(t_0)$ of the derivative of B with respect to time, i.e.

$$\frac{\mathrm{d}\Phi}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t_0} \iint_{S(t)} B^i(x,t_0) \hat{n}_i \mathrm{d}s + \iint_{S(t_0)} \frac{\partial B^i}{\partial t} \hat{n}_i \mathrm{d}s$$

Since we know already that (6.2) holds, we have that if f_{ind} is the induced f_{em}, we have that

$$f_{ind} = \oint_{\partial S(t_0)} E^i \hat{t}_i \mathrm{d}l = \oint_{\partial S(t_0)} E^i \hat{t}_i \mathrm{d}l = \iint_{S(t_0)} \frac{\partial B^i}{\partial t} \hat{n}_i \mathrm{d}s$$

But

$$\mathrm{d}\Phi = \iint_S B^i \frac{\mathrm{d}s}{\mathrm{d}t} \mathrm{d}t = \int \oint_{\partial S} \epsilon_{ijk} B^i v^j_D \hat{t}^k \mathrm{d}l \mathrm{d}t \tag{6.7}$$

Where v_D is the velocity in the direction of the movement of the circuit. Therefore, "dividing" by $\mathrm{d}t$, we get

$$\frac{\mathrm{d}\Phi}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t} \iint_{S(t)} B^i(x,t_0) \hat{n}_i \mathrm{d}s = \oint_{\partial S} \epsilon^i{}_{jk} \hat{t}_i v^j_D B^k \mathrm{d}l$$

Since we know that an additional f_{ind} is given by the deformation of the circuit $\partial S(t_0) \to \partial S(t)$, we have in total

$$f_{em} = -\frac{\mathrm{d}\Phi}{\mathrm{d}t} = \oint_{\partial S} E^i + \epsilon^i{}_{jk} \hat{t}_i v^j_D B^k \mathrm{d}l$$

Since $v_D = v_\parallel + v_\perp$ and $v_\parallel \parallel \hat{t}$ we get that since $\mathbf{E}_{ind} = \mathbf{E} + \mathbf{v} \times \mathbf{B}$, $\mathbf{E}_{ind}$ has $\mathbf{B}$ as a source, and we can condense it all in a single integro-differential equation, which is Faraday-Neumann-Lenz's law

$$\nabla \times \mathbf{E} \cdot \hat{\mathbf{n}} \mathrm{d}s = -\frac{\mathrm{d}}{\mathrm{d}t} \iint_S \frac{\partial \mathbf{B}}{\partial t} \cdot \hat{\mathbf{n}} \mathrm{d}s$$

□

§§ 6.1.1 Self-Induction

Having deduced our previous results, the first thing that we might check is how a circuit behaves with itself.
Consider a closed circuit with some current $I(t)$ such that $\partial_t I \approx 0$. If the magnetic permeability of the body is constant we can apply Biot-Savart for evaluating the field, and we have

$$B^i(t) = \frac{\mu_0 I(t)}{4\pi} \oint_C \frac{\epsilon^i{}_{jk} \hat{t}^j r^k}{r^3} \mathrm{d}l \tag{6.8}$$

Evaluating the flux of this field we have, since the current is independent from the integrated variables, that $\Phi \propto I(t)$, and therefore

$$\Phi(t) = LI(t) \qquad L = \frac{\mu_0}{4\pi} \oint_C \iint_S \frac{\epsilon^i{}_{jk} \hat{t}^j r^k}{r^3} \mathrm{d}l\mathrm{d}s$$

The constant L only depends on the geometry of the circuit as it's easy to see from the integral, and it's known as the *self-induction coefficient* or also as *autoinduction coefficient*.
From Faraday's law, since this flux depends on time (through our current $I(t)$), it generates an electromotive force f_L, as follows

$$f_L = -\frac{\mathrm{d}\Phi}{\mathrm{d}t} = -L\frac{\mathrm{d}I}{\mathrm{d}t} \tag{6.9}$$

The autoinduction coefficient has units the following units:

$$[L] = \frac{[\Phi]}{[I]} = \frac{W}{A} = \frac{Vs}{A} = \Omega s = H$$

The SI unit H is known as *Henry* and it's equal to Watts/Ampere. Note that this can also be calculated via Ohm's law, noting that $V = RI$ and that $[RI] = \Omega A$

§§ 6.1.2 Mutual-Induction

Consider now a setup similar to the previous one, but with two circuits C_1 and C_2, which are close enough to each other such that the generated magnetic fluxes through each circuit are not negligible. The fluxes as before will be proportional to the currents, and without evaluating the self-induction of both circuits we have

$$\begin{aligned} \Phi_1(B_2) &= \iint_{S_1} B_2^i \hat{n}_i \mathrm{d}s \propto I_2(t) \\ \Phi_2(B_1) &= \iint_{S_2} B_1^i \hat{n}_i \mathrm{d}s \propto I_1(t) \end{aligned} \tag{6.10}$$

Using the previous considerations, we have then, in index form, that

$$\Phi_i = M_{ij} I_j(t) \tag{6.11}$$

The coefficients M_{ij} are known as the *mutual induction coefficients*. Obviously $M_{ij} = M_{ji}$.

§ 6.2 Magnetic Energy

With what we wrote before, we might consider a circuit with a given self-induction coefficient L and some time-dependent current $I(t)$ flowing through it. We can evaluate the work of that the magnetic force exerts on these charges as follows. Per unit time

$$\frac{\mathrm{d}w}{\mathrm{d}t} = -f_{em} I(t) \tag{6.12}$$

Using Faraday's law we know that

$$\frac{\mathrm{d}\Phi}{\mathrm{d}t} = -f_{em} = -L\frac{\mathrm{d}I}{\mathrm{d}t}$$

Therefore

$$\frac{\mathrm{d}w}{\mathrm{d}t} = L\left(\frac{\mathrm{d}I}{\mathrm{d}t}\right)^2 = \frac{1}{2}LI^2(t) \tag{6.13}$$

We can go forward with this calculus, noting that then, since $\Phi = LI$, and

$$\Phi = \iint \epsilon^i{}_{jk} \partial^j A^k \hat{n}_i \mathrm{d}s$$

Then, using Stokes' theorem

$$\oint A^i \hat{t}_i \mathrm{d}l = LI$$

And therefore, vectorizing the current as $I^i = I\hat{t}^i$

$$w = \frac{1}{2}\oint A^i I_i \mathrm{d}l$$

Or in general

$$w = \frac{1}{2}\iiint A^i J_i \mathrm{d}^3x = \frac{1}{2\mu_0}\iiint \epsilon^i{}_{jk} A_i \partial^j B^k \mathrm{d}^3x \tag{6.14}$$

Where we used Ampere's law to get the last integral. Playing around with the last curl, using vector notation, we have

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

Applying Stokes' theorem on the divergence and noting that the surface integral goes to 0 when we integrate over all space, we end up with the following result

$$w = \frac{1}{2\mu_0}\iiint_{\mathbb{R}^3} B^2 \mathrm{d}^3x$$

The parallelism with the energy of an electric field is astounding. Written side by side we have

$$\begin{aligned} W_{es} &= \frac{1}{2}\int_{\mathbb{R}^3} V\rho \mathrm{d}^3x = \frac{\epsilon_0}{2}\int_{\mathbb{R}^3} E^2 \mathrm{d}^3x \\ W_{ms} &= \frac{1}{2}\int_{\mathbb{R}^3} A^i J_i \mathrm{d}^3x = \frac{1}{2\mu_0}\int_{\mathbb{R}^3} B^2 \mathrm{d}^3x \end{aligned} \tag{6.15}$$

§ 6.3 Maxwell's Equations

So far, we found 2 pairs of coupled differential equations for the electric and magnetic field

$$\begin{cases} \partial_i E^i = \dfrac{\rho}{\epsilon_0} \\ \epsilon^i{}_{jk}\partial^j E^k = -\partial_t B^i \end{cases} \tag{6.16}$$

And

$$\begin{cases} \partial_i B^i = 0 \\ \epsilon^i{}_{jk}\partial^j B^k = \mu_0 J^i \end{cases} \tag{6.17}$$

The second pair of equations holds only if the current field is divergenceless, but using Gauss' law and the current conservation equation we have

$$\partial_i J^i + \partial_t \rho = \partial_i J^i + \epsilon_0 \partial_t \partial_i E^i = 0 \tag{6.18}$$

Grouping the divergences we see that the time derivative of the electric field behaves exactly like a current, commonly called the "displacement current". In order to fix all the equations now we can add this new current in the last couple of the Maxwell equations and get the well known fundamental equations of electromagnetism

$$\begin{cases} \partial_i E^i = \dfrac{\rho}{\epsilon_0} \\ \epsilon^i{}_{jk}\partial^j E^k = -\partial_t B^i \\ \partial_i B^i = 0 \\ \epsilon^i{}_{jk}\partial^j B^k = \mu_0 J^i + \epsilon_0\mu_0\partial_t E^i \end{cases} \tag{6.19}$$

These equations account for moving charges and are absolutely general in nature. It will be seen later that they're also Lorentz invariant, therefore they preserve between Lorentz transformations and therefore are relativistically covariant.
For linear dielectric and magnetic media we can rewrite easily with the already known rules, Maxwell's equations

$$\begin{cases} \partial_i D^i = \rho \\ \epsilon^i{}_{jk}\partial^j D^k = \epsilon\mu\partial_t H^i \\ \partial_i H^i = 0 \\ \epsilon^i{}_{jk}\partial^j H^k = J^i + \partial_t D^i \end{cases} \tag{6.20}$$

§§ 6.3.1 Poynting's Vector, Energy Conservation

As we have seen via previous calculations, the energies of the two separated fields are, calling them U_e and U_m

$$\begin{aligned} U_e &= \frac{\epsilon_0}{2}\int_V E^2 \mathrm{d}^3x \\ U_m &= \frac{1}{2\mu_0}\int_V B^2 \mathrm{d}^3x \end{aligned} \tag{6.21}$$

We can imagine that the energy of the combined electromagnetic field will be a sum of the two, i.e.

$$U_{em} = \frac{1}{2}\int_V \epsilon_0 E^2 + \frac{B^2}{\mu_0}\mathrm{d}^3x \tag{6.22}$$

We try to confirm this using the work done by some particle. Substituting into the force the formula for Lorentz's force we get

$$F^i \mathrm{d}l_i = q\left(E^i + \epsilon^i{}_{jk} v^j B^k\right)\mathrm{d}l_i = q\left(E^i + \epsilon^i{}_{jk} v^j B^k\right) v_i \mathrm{d}t = qE^i v_i \mathrm{d}t = \mathrm{d}W$$

Which is what we expected.
Going to a microscopic consideration we substitute $q = \rho \mathrm{d}^3x$, we get $J^i = \rho v^i$, and therefore, integrating with respect to time we have

$$W = \int_V E^i J_i \mathrm{d}^3x\mathrm{d}t \implies E^i J_i = \frac{\mathrm{d}w}{\mathrm{d}t}$$

But, from Maxwell equations we have

$$\epsilon^i{}_{jk}\partial^j B^k = \mu_0 J^i + \frac{1}{c^2}\frac{\partial E^i}{\partial t} \implies J^i = \frac{\epsilon^i{}_{jk}\partial^j B^k}{\mu_0} - \frac{1}{\mu_0 c^2}\frac{\partial E^i}{\partial t}$$

Where we used $\mu_0\epsilon_0 = c^-2$ (it will be clear later, for now check multiplying those two and see that it adds up)
Therefore, we rewrite $E^i J_i$ as follows

$$E^i J_i = \frac{1}{\mu_0} E_i \epsilon^i{}_{jk}\partial^j B^k - \epsilon_0 E_i \epsilon^i{}_{jk}\partial^j B^k$$

Note that, tho:

$$\partial_i \epsilon^i{}_{jk} E^j \frac{B^k}{\mu_0} = \frac{1}{\mu_0} B_i \epsilon^i{}_{jk}\partial^j E^k - E^i \epsilon^i{}_{jk}\partial^j \frac{B^k}{\mu_0}$$

Therefore

$$E_i \epsilon^i{}_{jk}\partial^j \frac{B^k}{\mu_0} = -\frac{1}{\mu_0} B_i \epsilon^i{}_{jk}\partial^j E^k - \partial_i \epsilon^i{}_{jk} E^j \frac{B^k}{\mu_0} = -\frac{B_i}{\mu_0}\frac{\partial B^i}{\partial t} - \partial_i \epsilon^i{}_{jk} E^j \frac{B^k}{\mu_0}$$

Which, gives us back

$$E^i J_i = -\frac{1}{2}\frac{\partial}{\partial t}\left(\epsilon_0 E^2 + \frac{1}{\mu_0} B^2\right) - \frac{1}{\mu_0}\partial_i \epsilon^i{}_{jk} E^j B^k = \frac{\mathrm{d}w}{\mathrm{d}t} \tag{6.23}$$

We immediately recognize the volumetric density of energy of the electromagnetic field, let's denote it as u_{em}, and we get

$$E^i J_i = -\frac{\partial u_{em}}{\partial t} - \partial_i \epsilon^i{}_{jk} E^j \left(\frac{B^k}{\mu_0}\right) \tag{6.24}$$

We begin to have a better view of the phenomenon, we see a variation of energy on the right plus the divergence of some vector that we define now.

Definition 6.3.1 (Poynting Vector)**.** The *Poynting vector* is a vector defined as follows:

$$S^i = \frac{1}{\mu_0} \epsilon^i{}_{jk} E^j B^k \tag{6.25}$$

It has dimensions of a flux of energy, as we will see.

With the previous definition, everything becomes much clearer in terms of notation, in fact

$$E^i J_i = -\frac{\partial u_{em}}{\partial t} - \partial_i S^i$$

Integrating in a random volume V we get, as said before, our flux of energy!

$$\int_V E^i J_i \mathrm{d}^3 x = -\frac{\partial}{\partial t} \int_V u_{em} \mathrm{d}^3 x - \oiint_{\partial V} S^i \hat{n}_i \mathrm{d}s$$

It's clear that in order to make sense we must sum energies with energies, giving the previously stated dimensions of the Poynting vector as an energy flux.
Rewriting $E^i J_i$ as our work variation we have that it's nothing else than the time derivative of the volumetric density of mechanical energy, and writing $u_{em} + u_{mech}$ as our total energy variation, we have

$$\int_V \frac{\partial u}{\partial t} \mathrm{d}^3 x = -\oiint_{\partial V} S^i \hat{n}_i \mathrm{d}s \tag{6.26}$$

The associated PDE is clearly the conservation of energy of the whole system

$$\frac{\partial u}{\partial t} + \partial_i S^i = 0 \tag{6.27}$$

This shape also gives the real idea of what's Poynting's vector: an energy "current"

§§ 6.3.2 Stress Tensor, Momentum Conservation

Lorentz'force as we have seen, in terms of microscopic evaluations is written (it's a force density in this case) as

$$f^i = \rho E^i + \epsilon^i{}_{jk} J^j B^k \tag{6.28}$$

We rewrite it in terms of fields only using the two following Maxwell equations

$$\begin{aligned} \partial_i E^i = \frac{\rho}{\epsilon_0} \quad &\Longrightarrow \quad \rho = \frac{1}{\epsilon_0} \partial_i E^i \\ \epsilon^i{}_{jk} \partial^j B^k = \mu_0 J^i + \frac{1}{c^2} \frac{\partial E^i}{\partial t} \quad &\Longrightarrow \quad J^i = \frac{1}{\mu_0} \epsilon^i{}_{jk} \partial^j B^k - \frac{1}{\mu_0 c^2} \frac{\partial E^i}{\partial t} \end{aligned} \tag{6.29}$$

Therefore, Lorentz's force becomes, using $\epsilon_0 = (\mu_0 c^2)^{-1}$

$$f^i = \epsilon_0 E^i \partial_j E^j + \epsilon^i{}_{jk} \left(\frac{1}{\mu_0} \epsilon^j{}_{ml} \partial^l B^m - \epsilon_0 \frac{\partial E^j}{\partial t} \right) B^k$$

Or, moving inside the cross product for clarity

$$f^i = \epsilon_0 E^i \partial_j E^j + \frac{1}{\mu_0} \epsilon^i{}_{jk} B^k \epsilon^j{}_{lm} \partial^l B^m - \epsilon_0 \epsilon^i{}_{jk} \partial_t E^j B^k$$

Using the product rule on the time derivative at the last factor we have that

$$\partial_t \epsilon^i{}_{jk} E^j B^k = \epsilon^i{}_{jk} \partial_t (E^j) B^k + \epsilon^i{}_{jk} E^j \partial_t (B^k) \implies \epsilon^i{}_{jk} \partial_t (E^j) B^k = \partial_t \epsilon^i{}_{jk} E^j B^k - \epsilon^i{}_{jk} E^j \partial_t (B^k)$$

From the second Maxwell equation tho we have

$$\epsilon^i{}_{jk} \partial^j E^k = -\partial_t B^i$$

Therefore

$$-\epsilon^i{}_{jk} E^j \partial_t (B^k) = \epsilon^i{}_{jk} E^j \epsilon^k{}_{lm} \partial^l B^m$$

And everything comes back to

$$\epsilon^i{}_{jk} \partial_t (E^j) B^k = \partial_t \left(\epsilon^i{}_{jk} E^j B^k \right) + \epsilon^i{}_{jk} E^j \epsilon^k{}_{lm} \partial^l E^m$$

And inserting it back into the Lorentz force density gives

$$f^i = \epsilon_0 E^i \partial_j E^j + \frac{1}{\mu_0} \epsilon^i{}_{jk} B^k \epsilon^j{}_{lm} \partial^l B^m - \epsilon_0 \frac{\partial}{\partial t} \left(\epsilon^i{}_{jk} E^j B^k \right) + \epsilon_0 \epsilon^i{}_{jk} E^j \epsilon^k{}_{lm} \partial^l E^m \tag{6.30}$$

Or, rearranging the two fields

$$f^i = \epsilon_0 \left(E^i \partial_j E^j - \epsilon^i{}_{jk} \epsilon^k{}_{lm} E^j \partial^l E^m \right) + \frac{1}{\mu_0} \left(B^i \partial_j B^j - \epsilon^i{}_{jk} \epsilon^j{}_{lm} \partial^l B^m B^k \right) - \epsilon_0 \frac{\partial}{\partial t} \left(\epsilon^i{}_{jk} E^j B^k \right) \tag{6.31}$$

Where we used $\partial_i B^i = 0$ in order to symmetrize the shape of the equation.
Using the properties of the Levi-Civita symbol we have that (note that $g_{ij} = \delta_{ij}$ in this metric)

$$\epsilon^i{}_{jk} \epsilon^k{}_{lm} = \delta^i_l \delta_{jm} - \delta_{jl} \delta^i_l \tag{6.32}$$

We have (using a generic vector here, it can be either B^i or E^i)

$$\epsilon^i{}_{jk} \epsilon^k{}_{lm} A^j \partial^l A^m = A_m \partial^i A^m - A^l \partial_l A^i$$

So

$$f^i = \epsilon_0 \left(E^i \partial_j E^j + E^l \partial_l E^i - E^m \partial^i E_m \right) + \frac{1}{\mu_0} \left(B^i \partial_j B^j + B^l \partial_l B^i - B^m \partial^i B_m \right) - \frac{1}{c^2} \frac{\partial S^i}{\partial t}$$

Looking closely we see that the big mess inside the parentheses is simply

$$\partial_j \left(E^i E^j \right) - \frac{1}{2} \partial_j \left(E^k E_k \right) = \partial_j \left(E^i E^j - \frac{1}{2} \delta^{ij} E^k E_k \right)$$

Seen this, we define the following symmetric rank-2 tensor σ^{ij}

Definition 6.3.2 (Maxwell Stress Tensor)**.** The Maxwell stress tensor is defined as follows:

$$\sigma^{ij} = \epsilon_0 \left(E^i E^j - \frac{1}{2} \delta^{ij} E^k E_k \right) + \frac{1}{\mu_0} \left(B^i B^j - \frac{1}{2} \delta^{ij} B^k B_k \right) \tag{6.33}$$

It's a rank 2 tensor and it's obviously symmetric

The Lorentz force density becomes then the following simply

$$f^i = \frac{\partial \sigma^{ij}}{\partial x^k} - \frac{1}{c^2} \frac{\partial S^i}{\partial t} \tag{6.34}$$

7 Potentials and Fields

§ 7.1 Maxwell's Equation for Potentials

As we have seen already, Maxwell's equations are the following

$$\begin{cases} \partial_i E^i = \dfrac{\rho}{\epsilon_0} \\ \epsilon^i{}_{jk}\partial^j E^k = -\dfrac{\partial B^i}{\partial t} \\ \partial_i B^i = 0 \\ \epsilon^i{}_{jk}\partial^j B^k = \mu_0 J^i + \dfrac{1}{c^2}\dfrac{\partial E^i}{\partial t} \end{cases} \tag{7.1}$$

If we wanted to write the potential formulation of this equation, we must know that in general, the potentials might be different. In fact, Coulomb's law and Biot-Savart only work in the static case, i.e. where $\partial_t E = \partial_t B = 0$.
Using that $\partial_i B^i = 0$ from the third equation we can say for sure that

$$\epsilon^i{}_{jk}\partial^j A^k = B^i$$

And inserting it in the second we get

$$\epsilon^i{}_{jk}\partial^j E^k = -\frac{\partial}{\partial t}\epsilon^i{}_{jk}\partial^j A^k$$
$$\epsilon^i{}_{jk}\partial^j \left(E^k + \frac{\partial A^k}{\partial t} \right) = 0$$

The second line immediately tells us that the vector field inside must be the gradient of some scalar field!
Using $\partial^i V = -E^i$ then we can rewrite the electric potential as a sum of the time variation of the vector potential and the scalar potential, which gives us

$$E^i(x^i, t) = -\frac{\partial V}{\partial x_i} - \frac{\partial A^i}{\partial t} \tag{7.2}$$

Now that we have the potentials for the dynamic case we know that the Poisson equation for the electric field then becomes

$$\partial_i E^i = -\partial_i \partial^i V - \frac{\partial}{\partial t}\partial_i A^i = \frac{\rho}{\epsilon_0}$$

Or, written better

$$\partial_i \partial^i V + \frac{\partial}{\partial t} \partial_i A^i = -\frac{\rho}{\epsilon_0} \tag{7.3}$$

Also, for the equivalent vectorial Poisson equation for the A^i field

$$\epsilon^i{}_{jk} \epsilon^k{}_{lm} \partial^l A^m = \mu_0 J^i - \frac{1}{c^2} \frac{\partial}{\partial t} \left(\frac{\partial V}{\partial x_i} - \frac{1}{c^2} \frac{\partial A^i}{\partial t} \right)$$

Rewriting the first double cross product as

$$\partial^i \partial_j A^j - \partial_j \partial^j A^i$$

And bringing the time derivative of the gradient of V to the left, while grouping it inside the ∂^i operator (it's linear), and bringing with it the second time derivative on A^i we have (note also that I changed sign on both sides)

$$\partial_j \partial^j A^i - \frac{\partial}{\partial x_i} \left(\frac{1}{c^2} \frac{\partial V}{\partial t} + \partial_j A^j \right) - \frac{1}{c^2} \frac{\partial^2 A^i}{\partial t^2} = -\mu_0 J^i \tag{7.4}$$

Both together give us the most general possible way to formulate Maxwell equations for potentials, which reduce to two coupled non-homogeneous second order PDEs

$$\begin{cases} \partial_j \partial^j V + \dfrac{\partial}{\partial t} \partial_j A^j = -\dfrac{\rho}{\epsilon_0} \\ \partial_j \partial^j A^i - \dfrac{1}{c^2} \dfrac{\partial^2 A^i}{\partial t^2} - \dfrac{\partial}{\partial x_i} \left(\dfrac{1}{c^2} \dfrac{\partial V}{\partial t} + \partial_j A^j \right) = -\mu_0 J^i \end{cases} \tag{7.5}$$

§§ 7.1.1 Gauge Freedom

What we've learned before about electromagnetic potentials is that *they're gauge-modifiable*. Depending on what we really need we might choose between any given gauge, since Maxwell's equation are gauge-invariant.
The first gauge we will use is the most common one, it's useful when dealing with magnetostatics or when we really need to find V. It's *Coulomb's gauge*.
Here we set the divergence of A^i to zero, and the first equation of (7.5) reduces back to a Poisson's equation. The second simplifies a bit, but it's not easy to solve…

$$\begin{cases} \partial_i \partial^i V = -\dfrac{\rho}{\epsilon_0} \\ \partial_j \partial^j A^i - \dfrac{1}{c^2} \dfrac{\partial^2 A^i}{\partial t^2} - \dfrac{1}{c^2} \dfrac{\partial^2 V}{\partial x_i \partial t} = -\mu_0 J^i \end{cases} \tag{7.6}$$

Another thing to note here is that V *cannot* be observable, moving charges change ρ which changes V istantaneously, it's not Lorentz invariant.
The second most important gauge we can define it's Lorenz's[1] gauge, which defines the divergence of A as follows

$$\partial_i A^i = \frac{1}{c^2} \frac{\partial V}{\partial t} \tag{7.7}$$

[1] Lorenz, not Lorentz, apparently

Then, reinserting it back to (7.5) we get by immediate substitution two uncoupled non-homogeneous wave equations

$$\begin{cases} \partial_j \partial^j V - \dfrac{1}{c^2}\dfrac{\partial^2 V}{\partial t^2} = -\dfrac{\rho}{\epsilon_0} \\ \partial_j \partial^j A^i - \dfrac{1}{c^2}\dfrac{\partial^2 A^i}{\partial t^2} = -\mu_0 J^i \end{cases} \tag{7.8}$$

This is a *relativistic wave equation* with sources ρ and J^i.
Note that if we define the four-gradient as the 4-vector composed by the following components ($\mu = 0, \cdots, 3$)

$$\partial_\mu = \left(\frac{1}{c}\partial_t, -\partial_i\right) \tag{7.9}$$

We have that, formally

$$\partial_\mu \partial^\mu = \frac{1}{c^2}\partial_t^2 - \partial_i \partial^i = \Box$$

Where the box operator is known as the *D'Alambertian*, which is the equivalent of the Laplacian in 4 spacetime dimensions. Therefore we can also write

$$\begin{cases} \Box V = \partial^\mu \partial_\mu V = \dfrac{\rho}{\epsilon_0} \\ \Box A^i = \partial^\mu \partial_\mu A^i = \mu_0 J^i \end{cases} \tag{7.10}$$

§ 7.2 Retarded Potentials

Using (7.10) and setting the time derivatives as 0, we get back Poisson's equations for both potentials, for which we know already the general solution for a volume V.

$$V(x^j) = \frac{1}{4\pi\epsilon_0}\int_V \frac{\rho(\tilde{x}^j)}{r}\mathrm{d}^3\tilde{x}$$
$$A^i(x^j) = \frac{\mu_0}{4\pi}\int_V \frac{J^i(\tilde{x}^j)}{r}\mathrm{d}^3\tilde{x}$$

We can say, from the previous equations, that the interaction travels at speed c, therefore we might imagine the time progression of the interaction as "retarded in time" by a factor of r/c. We then define the *retarded time* as

$$t_r = t - \frac{r}{c} \tag{7.11}$$

We therefore can imagine a solution to those equations as

$$\begin{aligned} V(x^j, t_r) &= \frac{1}{4\pi\epsilon_0}\int_V \frac{\rho(\tilde{x}^j, t_r)}{r}\mathrm{d}^3\tilde{x} \\ A^i(x^j, t_r) &= \frac{\mu_0}{4\pi}\int_V \frac{J^i(\tilde{x}^j, t_r)}{r}\mathrm{d}^3\tilde{x} \end{aligned} \tag{7.12}$$

Note that we could imagine this solution only due to the mathematical shape of the equation, it cannot be done the same way for the fields.
Now let's check if this idea we had is a solution for the relativistic equations. Noting that

$$\partial_i t_r = -\frac{\hat{x}^i}{c}$$

We have, after using the chain rule

$$\frac{\partial \rho}{\partial x^i} = -\frac{1}{c}\frac{\partial \rho}{\partial t}\hat{x}^i$$

And therefore

$$\partial_i V = -\frac{1}{4\pi\epsilon_0}\int_V \frac{1}{c}\frac{\partial \rho}{\partial t}\frac{\hat{x}^i}{r} - \rho\frac{\hat{x}^i}{r^2}\mathrm{d}^3x$$

Applying again the del operator we get

$$\partial^i\partial_i V = -\frac{1}{4\pi\epsilon_0}\int_V \frac{1}{c}\left(\frac{\hat{x}^i}{r}\frac{\partial^2\rho}{\partial t\partial x^i} + \frac{\partial\rho}{\partial t}\frac{\partial_i\hat{x}^i}{r} + \frac{\partial\rho}{\partial t}\hat{x}^i\partial_i\left(\frac{1}{r}\right)\right) + \left(\frac{\hat{x}^i}{r^2}\frac{\partial\rho}{\partial x^i} + \rho\frac{\partial}{\partial x^i}\left(\frac{\hat{x}^i}{r^2}\right)\right)\mathrm{d}^3x$$

But

$$\frac{\partial^2\rho}{\partial t\partial x^i} = -\frac{1}{c}\frac{\partial^2\rho}{\partial t^2}\hat{x}^i$$

And

$$\partial_i\left(\frac{\hat{x}^i}{r}\right) = \frac{1}{r^2}, \qquad \partial_i\left(\frac{\hat{x}^i}{r^2}\right) = 4\pi\delta^3(x^i)$$

Therefore, finally

$$\partial^i\partial_i V = \frac{1}{4\pi\epsilon_0 c^2}\int_V \frac{1}{r}\frac{\partial^2\rho}{\partial t^2}\mathrm{d}^3x - \frac{\rho}{\epsilon_0}\delta^3(x^i) \tag{7.13}$$

Seeing immediately on the right the time second time derivative with respect to ct of V, bringing it to the left and playing with minuses we get again the awaited Maxwell equation.

$$\Box V = \frac{\rho}{\epsilon_0}$$

The calculation for A^i is completely analogous. Note that we could also have chosen an *advanced time* t_a defined as

$$t_a = t + \frac{r}{c}$$

Everything comes back to the two Maxwell equations, but the physical sense gets lost since the potentials we found don't respect causality, they sense the change *before* it actually happens in the chosen reference frame.

§ 7.3 Jefimenko's Equations

Given the two retarded potentials defined in (7.12), we could imagine to determine the electric and magnetic field generated by both. Since the retarded potentials, as we have shown, solve *generally*

Maxwell's equations (7.10), the fields will also solve generally Maxwell's equations for the fields. We begin by finding E^i. We know that

$$E^i = -\partial^i V - \partial_t A^i$$

Therefore

$$E^i = -\frac{1}{4\pi\epsilon_0}\int_V \partial^i\left(\frac{\rho}{r}\right)\mathrm{d}^3x - \frac{\mu_0}{4\pi}\int_V \frac{1}{r}\frac{\partial J^i}{\partial t}\mathrm{d}^3x \tag{7.14}$$

From the previous calculations, we already know that

$$\partial^i\left(\frac{\rho}{r}\right) = -\frac{1}{c}\frac{\partial\rho}{\partial t}\hat{x}^i - \rho\frac{\hat{x}^i}{r^2}$$

And we get easily the first Jefimenko equation for the E field

$$E^i(x^j,t) = \frac{1}{4\pi\epsilon_0}\int_V\left(\frac{1}{c}\frac{\partial\rho}{\partial t} + \frac{\rho}{r}\right)\frac{\hat{x}^i}{r}\mathrm{d}^3x - \frac{\mu_0}{4\pi}\int_V\frac{1}{r}\frac{\partial J^i}{\partial t}\mathrm{d}^3x$$

Using $\epsilon_0 = (\mu_0 c^2)^{-1}$ we can group everything in a clearer equation

$$E^i(x^j,t) = \frac{1}{4\pi\epsilon_0}\int_V\left(\frac{1}{c}\frac{\partial\rho}{\partial t} + \frac{\rho}{r}\right)\frac{\hat{x}^i}{r} - \frac{1}{rc^2}\frac{\partial J^i}{\partial t}\mathrm{d}^3x \tag{7.15}$$

For B^i the calculations are slightly harder due to the presence of the curl, but with some discipline are doable. We have that

$$B^i = \epsilon^i{}_{jk}\partial^j A^k = \frac{\mu_0}{4\pi}\int_V\frac{1}{r}\epsilon^i{}_{jk}\partial^j J^k + \epsilon^i{}_{jk}J^i\partial^k\left(\frac{1}{r}\right)\mathrm{d}^3x$$

But, by definition we have

$$\partial_i J^k = \left(\partial_t J^k\partial_k\right)t_r = -\frac{1}{c}\frac{\partial J^k}{\partial t}\frac{\partial r}{\partial x^k} = \frac{1}{c}\frac{\partial J^k}{\partial t}\hat{x}_k$$

So, the cross product is simply

$$\epsilon^i{}_{jk}\partial^j J^k = \frac{1}{c}\epsilon^i{}_{jk}\partial_t J^j\hat{x}^k$$

The second part instead comes immediately from the gradient of r^{-1}, and we have

$$B^i(x^j,t) = \frac{\mu_0}{4\pi}\int_V\frac{1}{r}\epsilon^i{}_{jk}\left(\frac{J^j}{r} + \frac{1}{c}\frac{\partial J^j}{\partial t}\right)\hat{x}^k\mathrm{d}^3x \tag{7.16}$$

Both Jefimenko equations grouped are, therefore (and finally)

$$\begin{cases} E^i(x^k,t) = \dfrac{1}{4\pi\epsilon_0}\displaystyle\int_V\frac{1}{r}\left(\frac{1}{c}\frac{\partial\rho}{\partial t}\hat{x}^i + \rho\frac{\hat{x}^i}{r} - \frac{1}{c^2}\frac{\partial J^i}{\partial t}\right)\mathrm{d}^3\tilde{x} \\ B^i(x^k,t) = \dfrac{\mu_0}{4\pi}\displaystyle\int_V\frac{1}{r}\epsilon^i{}_{jk}\left(\frac{J^j}{r} + \frac{1}{c}\frac{\partial J^j}{\partial t}\right)\hat{x}^k\mathrm{d}^3\tilde{x} \end{cases} \tag{7.17}$$

8 Electromagnetic Waves

§ 8.1 The Wave Equation

Maxwell's equations are of utmost importance in studying the behavior of electromagnetic field, due to their versatility and generality. It's due to Maxwell himself that we now treat light as electromagnetic radiation, particularly, electromagnetic waves. The reasoning behind Maxwell's work comes *directly* from his equation. Consider a location of space in which there are no charges nor currents, i.e. no *sources*. For such system, Maxwell's equations are[1]

$$\begin{cases} \nabla \times \mathbf{E} = -\dfrac{\partial \mathbf{B}}{\partial t} \\ \nabla \times \mathbf{B} = \dfrac{1}{c^2}\dfrac{\partial \mathbf{E}}{\partial t} \end{cases} \tag{8.1}$$

Where, the divergence equations are both equally zero

$$\nabla \cdot \mathbf{E} = \nabla \cdot \mathbf{B} = 0 \tag{8.2}$$

It's important to remember that

$$c^2 = \frac{1}{\mu_0 \epsilon_0}$$

Is the "*speed of light*" for reasons that will be clear after a couple of manipulations. For reasons of symmetry of the two curl equations we use the following substitution

$$\mathbf{B} = \mu_0 \mathbf{H} \tag{8.3}$$

for which, the only nonzero equations are

$$\begin{cases} \nabla \times \mathbf{E} = -\mu_0 \dfrac{\partial \mathbf{H}}{\partial t} \\ \nabla \times \mathbf{H} = \epsilon_0 \dfrac{\partial \mathbf{E}}{\partial t} \end{cases} \tag{8.4}$$

[1] From now on, here we will use the standard boldface vector notation for ease of reading. Index notation will be used in the chapter on crystals and in some other section

The main problem of these equations is that they're still coupled between eachother, and there are still the previous divergence equations, but this can be solved quickly. Remembering the following operator identity

$$\nabla \times \nabla \times [\quad] = \nabla (\nabla \cdot [\quad]) - \nabla^2 [\quad]$$

We get, by taking the curl of both equations, and reinserting them to the right hand side

$$\begin{cases} \nabla \times \nabla \times \mathbf{E} = -\mu_0 \epsilon_0 \dfrac{\partial^2 \mathbf{E}}{\partial t^2} \\ \nabla \times \nabla \times \mathbf{H} = -\mu_0 \epsilon_0 \dfrac{\partial^2 \mathbf{H}}{\partial t^2} \end{cases} \tag{8.5}$$

Or, inserting the identity and (8.2)

$$\begin{cases} \nabla^2 \mathbf{E} = \dfrac{1}{c^2} \dfrac{\partial^2 \mathbf{E}}{\partial t^2} \\ \nabla^2 \mathbf{H} = \dfrac{1}{c^2} \dfrac{\partial^2 \mathbf{H}}{\partial t^2} \end{cases} \tag{8.6}$$

Or, more compactly

$$\begin{cases} \Box \mathbf{E} = 0 \\ \Box \mathbf{H} = 0 \end{cases} \tag{8.7}$$

Where

$$\Box = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2 \tag{8.8}$$

Is known as the *D'Alembertian* operator. The equation we wrote is a wave equation for waves traveling at $v = c$.

§§ 8.1.1 Dielectric Wave Equation

If we consider now ourselves inside some media, we have to take account of both *polarization* and *magnetization*, given by the presence of atoms inside the medium. These atoms will absorb some special frequencies, which will be known as *resonance frequencies*.
Far from the resonant frequencies, the medium is known as a transparent and non-absorbent medium. Maxwell's equations are the usual complete ones:

$$\begin{cases} \nabla \cdot \mathbf{E} = \dfrac{\rho}{\epsilon_0} \\ \nabla \times \mathbf{E} = -\dfrac{\partial \mathbf{B}}{\partial t} \\ \nabla \cdot \mathbf{B} = 0 \\ \nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \dfrac{\partial \mathbf{E}}{\partial t} \end{cases} \tag{8.9}$$

Consider now a real electromagnetic wave, it will be composed of multiple frequencies, denoted with ω. An electromagnetic wave will be denoted as *monochromatic* if and only if it's composed by a single frequency, (note that in nature there are no monochromatic waves). In this ideal case, the electric field is:

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_\omega(\mathbf{r}) e^{-i\omega t} \tag{8.10}$$

With the exponential coming from the wave equation itself.
Since we are dealing with dielectrics, we gotta consider charge polarization $\mathbf{P}$, which will also be decomposed in frequencies, therefore, inserting the D field in our calculations

$$\mathbf{D}_\omega = \epsilon_0 \mathbf{E}_\omega + \mathbf{P}_\omega = \epsilon_\omega \mathbf{E}_\omega \tag{8.11}$$

Where we used the following two relations

$$\begin{cases} \mathbf{P}_\omega = \epsilon_0 \chi_\omega \mathbf{E}_\omega \\ \epsilon_\omega = \epsilon_0 (1 + \chi_\omega) \end{cases} \tag{8.12}$$

Noting that here, in general, ϵ_ω depends on position, the first Maxwell equation for dielectrics becomes slightly more complicated

$$\nabla \cdot \mathbf{D}_\omega = \mathbf{E}_\omega \cdot \nabla \epsilon_\omega + \epsilon_\omega \nabla \cdot \mathbf{E}_\omega = 0$$

Solving for E_ω we get

$$\nabla \cdot \mathbf{E}_\omega = -\frac{\mathbf{E}_\omega \cdot \nabla \epsilon_\omega}{\epsilon_\omega} \tag{8.13}$$

Evaluating the time derivatives for the monochromatic fields we have

$$\frac{\partial \mathbf{E}}{\partial t} = -i\omega \mathbf{E}$$
$$\frac{\partial \mathbf{H}}{\partial t} = -i\omega \mathbf{H}$$

Inserting everything into Maxwell's equations we get the set of equations for monochromatic waves in general non-magnetic media

$$\begin{cases} \nabla \cdot \mathbf{E}_\omega = -\dfrac{\mathbf{E}_\omega \cdot \nabla \epsilon_\omega}{\epsilon_\omega} \\ \nabla \times \mathbf{E}_\omega = i\omega \mu_0 \mathbf{H}_\omega \\ \nabla \cdot \mathbf{H}_\omega = 0 \\ \nabla \times \mathbf{H}_\omega = -i\omega \epsilon_\omega \mathbf{E}_\omega \end{cases} \tag{8.14}$$

Now, using the same technique we used before for finding the wave equation, we get

$$\nabla \times \nabla \times \mathbf{E}_\omega = -\omega^2 \mu_0 \epsilon_\omega \mathbf{E}_\omega$$
$$\nabla \times \nabla \times \mathbf{H}_\omega = -\omega^2 \mu_0 \epsilon_\omega \mathbf{H}_\omega$$

Therefore, inserting the divergence equations and taking care of the minus signs

$$\nabla^2 \mathbf{E}_\omega + \nabla \left(\frac{\mathbf{E}_\omega \cdot \nabla \epsilon_\omega}{\epsilon_\omega} \right) = \omega^2 \mu_0 \epsilon_\omega \mathbf{E}_\omega$$
$$\nabla^2 \mathbf{H}_\omega = \omega^2 \mu_0 \epsilon_\omega \mathbf{H}_\omega$$

Definition 8.1.1 (Refraction Index)**.** We define the *refraction index* n_ω as follows

$$n_\omega(\mathbf{r}) = \sqrt{\frac{\epsilon_\omega(\mathbf{r})}{\epsilon_0}} \tag{8.15}$$

Hence

$$\epsilon_\omega = \epsilon_0 n_\omega^2$$

Inserting the previous definition into the divergence of E_ω we see that

$$\frac{\mathbf{E}_\omega \cdot \nabla \epsilon_\omega}{\epsilon_\omega} = 2\mathbf{E}_\omega \cdot \nabla \log(n_\omega)$$

And, the right hand side becomes

$$\omega^2 \mu_0 \epsilon_\omega \mathbf{E}_\omega = \frac{\omega^2}{c^2} n_\omega^2 \mathbf{E}_\omega$$

Definition 8.1.2 (Wavenumber)**.** We define the *vacuum wavenumber* k_0 as follows

$$k_0 = \frac{\omega}{c} = \frac{2\pi}{\lambda} \tag{8.16}$$

Reuniting both definitions and the simplification we then get, for the equation on E

$$\nabla^2 \mathbf{E}_\omega + 2\nabla\left(\mathbf{E}_\omega \cdot \nabla \log(n_\omega)\right) = k_0^2 n_\omega^2 \mathbf{E}_\omega$$

Now, in order to ease calculations in our range of frequencies (or wavelenghts, or wavenumbers also) we check if defined a characteristic length l_n which indicates the spatial scale of variation of n_ω (remember that it depends on space position), we have that:

$$\nabla^2 E_\omega \propto \frac{E_\omega}{l_n \lambda} \frac{\Delta n_\omega}{n_\omega} \propto k_0^2 n_\omega^2 E_\omega$$
$$\|\nabla \nabla \cdot \mathbf{E}_\omega\| \propto \frac{E_\omega}{l_n^2} \frac{\Delta n_\omega}{n_\omega}$$

And since in optical ranges of light $l_n >> \lambda$ and $\Delta n_\omega << n_\omega$ we can discard immediately the divergence term, and get two symmetric wave equations for the wave in a generic dielectric (nonmagnetic or transparent) medium

$$\begin{cases} \nabla^2 \mathbf{E}_\omega = k_0^2 n_\omega^2 \mathbf{E}_\omega \\ \nabla^2 \mathbf{H}_\omega = k_0^2 n_\omega^2 \mathbf{H}_\omega \end{cases} \tag{8.17}$$

These equations are known as *Helmholtz equations* for the single chromatic part of the E, H fields. Note that these equations are exactly the equations solved by a single Fourier component of the transformed wave equation (n appears since we are in some media and $\epsilon_r \neq 0$).
Note that via Fourier transforms it's possible to go back to the already known wave equation, where the speed of propagation is not c but it's $c/n_\omega = u$. It's then also obvious that in vacuum $n_\omega = 1$, which is also clear from the definition of the refraction index.

§ 8.2 Vectorial Behavior of Electromagnetic Waves

The general solution to the wave equation (8.6) can be written both as a real function or a complex exponential. The latter one, although non "real", eases a lot calculations and therefore will be the favored approach. Using all previous definitions the general result is

$$\begin{aligned}\mathbf{E} &= \mathbf{E}_0 e^{i\mathbf{k}\cdot\mathbf{r}-i\omega t}\\ \mathbf{H} &= \mathbf{H}_0 e^{i\mathbf{k}\cdot\mathbf{r}-i\omega t}\end{aligned} \tag{8.18}$$

Evaluating the curl, the divergence and the time derivative for these solutions it's possible to write the following operator equation

$$\begin{aligned}\hat{\partial}_t &\to -i\omega\\ \hat{\nabla} &\to i\mathbf{k}\end{aligned} \tag{8.19}$$

Where $\mathbf{k}$ is known as the *wavevector*, which is the vector with magnitude $\|\mathbf{k}\| = nk_0$ And Maxwell's equations can be rewritten as

$$\begin{cases}\mathbf{k}\cdot\mathbf{E} = 0\\ \mathbf{k}\cdot\mathbf{H} = 0\\ \mathbf{k}\times\mathbf{E} = \omega\mu_0\mathbf{H}\\ \mathbf{k}\times\mathbf{H} = -\omega\epsilon\mathbf{E}\end{cases} \tag{8.20}$$

Therefore, $\mathbf{k}, \mathbf{E}, \mathbf{H}$ are three mutually orthogonal vectors. By definition, then, $\mathbf{k}$ must be oriented parallel to the direction of motion.
Also from the last three equations, taken the norm of the two and solved the equations for E/H we have

$$\frac{E}{H} = n\sqrt{\frac{\epsilon_0}{\mu_0}} = \frac{n}{Z_0} \tag{8.21}$$

With $Z_0^2 = \mu_0/\epsilon_0$ being the *free space impedance* which has value of $Z_0 \approx 377\Omega$. Note also that by definition of the Poynting vector $\mathbf{S}$, we also have that $\mathbf{k} \parallel \mathbf{S}$, which implies that, if we define the irradiance as $I = \|\mathbf{S}\|$, that:

$$\mathbf{S} = I\frac{\mathbf{k}}{k} = I\hat{\mathbf{k}} \tag{8.22}$$

Or, using what we found before

$$I = \frac{n}{2Z_0}E_0^2 \tag{8.23}$$

For plane waves it's actually better if we take the time average of the fields using phasor notation (complex exponential notation) for the fields, we have

$$\langle\mathbf{S}(t)\rangle = \mathbf{S} = \mathfrak{Re}\langle\mathbf{E}_0 e^{i\mathbf{k}\cdot\mathbf{r}-i\omega t}\rangle \times \mathfrak{Re}\langle\mathbf{H}_0 e^{i\mathbf{k}\cdot\mathbf{r}-i\omega t}\rangle$$

Since $\mathfrak{Re}(z) = \frac{1}{2}(z+\overline{z})$ and the complex conjugate is distributive, the calculation boils down to simplifying the following expression

$$\frac{1}{4}\left\langle\left(\mathbf{E}_0\times\mathbf{H}_0 e^{2i\mathbf{k}\cdot\mathbf{r}-2i\omega t} + \overline{\mathbf{E}}_0\times\mathbf{H}_0 = \mathbf{E}_0\times\overline{\mathbf{H}}_0 + \overline{\mathbf{E}}_0\times\overline{\mathbf{H}}_0 e^{-2i\mathbf{k}\cdot\mathbf{r}+2i\omega t}\right)\right\rangle$$

Noting that the average value for the real part of the exponential is 0, we have

$$\mathbf{S} = \frac{1}{2}\mathbf{E}_0 \times \overline{\mathbf{H}}_0 \tag{8.24}$$

Where we omitted both the real part operator and the time average.

§§ 8.2.1 Polarization

The vectorial nature of waves comes up in most part with the phenomenon of *polarization*, which is simply the "favored" direction of oscillation of the wave. Waves can also be non-polarized, as is the case for natural light, when there is no well defined oscillation direction.
The simplest polarization state obtainable is *linear polarization*. Setting $\mathbf{k} \parallel \hat{\mathbf{z}}$ we have $\mathbf{E}, \mathbf{H}$ orthogonal and coplanar on the xy plane.
Linear polarization is then achieved when $\mathbf{E}$ oscillates with a constant angle from the chosen x axis. This can be expressed mathematically as:

$$\mathbf{E} = E_x\hat{\mathbf{x}} + E_y\hat{\mathbf{y}} \tag{8.25}$$

Note how here, in general, $\mathbf{E} \in \mathbb{R}^2$.
The instrument used to generate linearly polarized light is the *linear polarizer*. This object is built in a way such that it transmits light only in one orientation. The associated axis is known as the *transmission axis*.
Chosen a 2D orthonormal reference system $\hat{\mathbf{t}}, \hat{\mathbf{s}}$ in which the t axis is parallel to the transmission axis, then, we must have

$$\mathbf{E}_{in} = E_t\hat{\mathbf{t}} + E_s\hat{\mathbf{s}}$$
$$\mathbf{E}_{out} = E_t\hat{\mathbf{t}}$$

Since $E_t = E_{in}\cos\theta$ then the irradiance of the outgoing field is

$$I_{out} = I_{in}\cos^2\theta \tag{8.26}$$

This behavior is known as *Malus' law*.
In case that the incoming light isn't polarized and it can't be described with the previous decomposition, taken the time average, we get

$$I_{out} = \frac{1}{2}I_{in}$$

Due to the superposition principle, it's not hard to imagine a mixture of polarized and unpolarized light. The *degree of polarization* P of this light can be evaluated using Malus' law, and it will be equal to

$$P = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{8.27}$$

Or, also, as the fraction of polarized irradiance

$$P = \frac{I_{pol}}{I_{pol} + I_{unp}} \tag{8.28}$$

There is one more possible (general) state of polarization, in which the field is totally described as a complex vector. In this state the two components of the field are dephased by exactly $\phi = \pi/2$, adding

the phase to the exponential, we get a new factor of i on one of the two components.
The field can then be described by a complex vector in the following way

$$\mathbf{E} = E_0\hat{\mathbf{x}} + iE_1\hat{\mathbf{y}} \tag{8.29}$$

This configuration is known as *elliptical polarization* or *circular polarization* when $E_1 = E_0$, i.e.

$$\mathbf{E} = E_0(\hat{\mathbf{x}} + i\hat{\mathbf{y}}) \tag{8.30}$$

This special polarization state has two version, called *right hand* and *left hand* polarization, depending on whether the dephasing between the components is $+i$ or $-i$. Left hand elliptical (circular) polarization is defined by a positive phase difference of i.
The instrumentation used to produce elliptical or circular polarization is known as *quarter wave plates*, optical devices which exhibit a special property known as *birefringence*, i.e., they have two different refraction indexes, n_f, n_s where the subscripts mean "slow" and "fast" because $n_f < n_s$. The objects are specially made in a way such that the indexes are disposed in two mutually orthogonal axes.
The quarter wave denomination comes from the criterion used to choose the thickness of the slabs. In general, it depends from the wavelength of the beam that gets polarized with the following formula

$$d = \frac{\lambda_0}{4(n_s - n_f)} \tag{8.31}$$

Note that a quarter wave plate isn't enough to generate circularly polarized light. In order to make both components equal after the dephasing, it's necessary to linearly polarize light at an angle of $\pi/4$, so that when it traverses the quarter wave plate it's perfectly diagonal between the axes.

§§ 8.2.2 Jones Calculus

Due to the vectorial nature of polarization it's not hard to imagine a new notation which uses the power of linear algebra.
A generic wave will be described as follows

$$\mathbf{E} = \begin{pmatrix} E_x e^{i\phi_x} \\ E_y e^{i\phi_y} \end{pmatrix} \tag{8.32}$$

This vector is known as a *Jones vector*.
Then, the two basic polarization states are, in general

$$\mathbf{E}_{lin} = \begin{pmatrix} A \\ B \end{pmatrix} = A\begin{pmatrix} 1 \\ 0 \end{pmatrix} + B\begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{8.33}$$

Linear, when $A, B \in \mathbb{R}$, decomposed as a sum of horizontally and vertically polarized waves.

$$\mathbf{E} = \begin{pmatrix} A \\ \pm iB \end{pmatrix} = A\begin{pmatrix} 1 \\ 0 \end{pmatrix} \pm B\begin{pmatrix} 0 \\ i \end{pmatrix} \tag{8.34}$$

Elliptical, with $A, B \in \mathbb{C}$ in general decomposed into a sum of dephased linearly polarized light. Circularly polarized light can be interpreted as a special case with $A = B$. All these polarization states can be treated with usual vector operations, and with it, it's possible to determine the final polarization state of two superimposed waves.

The action of *optically active objects* such as linear polarizers and waveplates, can then be described by multiples of rotation matrices, which act linearly on the polarization states. The general action of an object is then described by a 2×2 complex *Jones matrices*.
Using composition it's also possible to determine the action of multiple objects on the polarization state.
Some common optical objects have the following matrix reprsentations (when these is more thae one written I'm writing different orientations of the fast/transmission axis $[0, \pi, \pi/4]$)
Linear polarizer:

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad \frac{1}{2}\begin{pmatrix} 1 & \pm 1 \\ \pm 1 & 1 \end{pmatrix} \tag{8.35}$$

Quarter wave plate:

$$\begin{pmatrix} 1 & 0 \\ 0 & -i \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \quad \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & \pm i \\ \pm i & 1 \end{pmatrix} \tag{8.36}$$

Half wave plate:

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{8.37}$$

Phase retarder:

$$\begin{pmatrix} e^{i\phi_x} & 0 \\ 0 & e^{i\phi_y} \end{pmatrix} \tag{8.38}$$

Circular polarizer (quarter wave + diagonal polarizer):

$$\frac{1}{2}\begin{pmatrix} 1 & \pm i \\ \mp i & \end{pmatrix} \tag{8.39}$$

Another thing we can borrow from algebra and use it with Jones calculus is orthogonality and eigenstates. Two polarization states are said to be *orthogonal* if

$$\mathbf{E}_1 \cdot \mathbf{E}_2 = \begin{pmatrix} \overline{A} & \overline{B} \end{pmatrix}\begin{pmatrix} C \\ D \end{pmatrix} = \overline{A}C + \overline{B}D = 0 \tag{8.40}$$

A *polarization eigenstate* instead is the eigenvector of a Jones matrix, i.e. the only thing that can change when light with this kind of polarization passes through the object is its phase and/or its amplitude. Note how, since Jones vectors require *purely* polarized states, we cannot define nonpolarized light.

§ 8.3 Boundary Conditions for Electromagnetic Waves

§§ 8.3.1 Snell's Law

Consider a plane boundary separating two different optical media, with refraction indexes n_1, n_2. It's empirically proven from Huygens that the wave will be both transmitted and reflected at the boundary. These three waves will have the following exponential dependence

$$e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$$

$$e^{i(\mathbf{k}_R\cdot\mathbf{r}-\omega t)}$$

$$e^{i(\mathbf{k}_T\cdot\mathbf{r}-\omega t)}$$

Where $\mathbf{k}_R, \mathbf{k}_T$ are respectively the wavevectors for the reflected and transmitted wave.
Said, without loss of generality, $t = 0$ at the boundary, due to the continuity of the field we must have

$$e^{i\mathbf{k}\cdot\mathbf{r}} = e^{i\mathbf{k}_R\cdot\mathbf{r}} = e^{i\mathbf{k}_T\cdot\mathbf{r}}$$

Writing in terms of k_0 which is the same, we have

$$\begin{aligned}\mathbf{k} &= n_1\mathbf{k}_0\\ \mathbf{k}_R &= n_1\mathbf{k}_0\\ \mathbf{k}_T &= n_2\mathbf{k}_0\end{aligned}$$

Said θ the incidence angle, θ_R the reflection angle and φ the transmission angle, writing explicitly the norm of the scalar product, we get what's commonly known as *Snell's law*

$$\begin{aligned}\sin\theta_R &= \sin\theta\\ n_1\sin\theta &= n_2\sin\varphi\end{aligned} \tag{8.41}$$

The first one indicates that $\theta = \theta_R$, while the second can be rewritten as

$$\sin\theta = n\sin\varphi \tag{8.42}$$

Where $n = n_2/n_1 = \sin\theta/\sin\varphi$ is the *relative refraction index* of the medium.

§§ 8.3.2 Fermat's Principle

All of this can be mathematically explained using *Fermat's principle*.
Taken two points A, B, Fermat's principle states that light will travel in the path for which the traveling time will be minimized. Said $n \equiv n(\mathbf{r}(t))$ we define the following functional

$$t[\mathbf{r}] = \int_{t_A}^{t_B} \frac{n(\mathbf{r}(t))}{c}\mathrm{d}t \tag{8.43}$$

Where we used $\mathrm{d}t = u\mathrm{d}s$.
We also define the *optical path* as

$$l_{AB} = \int_A^B n(s)\mathrm{d}s \tag{8.44}$$

Suppose now that A, B are in two different media with n_A, n_B as refraction indexes. Putting ourselves in the coordinate system of the incidence point, with the y axis parallel to the boundary, and with A, B distant respectively d_1, d_2 from the origin, we have that the total optical path is

$$l = d_1 + d_2 = \sqrt{a^2 + x^2} + \sqrt{b^2 + (d-x)^2}$$

Where a, b are the y distances of the two points. Dividing by n/c we have

$$t(x) = \frac{n_A}{c}\sqrt{a^2 + x^2} + \frac{n_B}{c}\sqrt{b^2 + (d-x)^2}$$

Applying the fundamental principle of the calculus of variations, we take the derivative and search for the extremal

$$\frac{\partial t}{\partial x} = \frac{n_A x}{c\sqrt{a^2+x^2}} - \frac{n_B(d-x)}{c\sqrt{b^2+(d-x)^2}} = \frac{n_A x\sqrt{b^2+(d-x)^2} - n_B(d-x)\sqrt{a^2-x^2}}{c\sqrt{a^2+x^2}\sqrt{b^2+(d-x)^2}} = 0$$

Since the denominator is always not zero, everything boils down to

$$n_A^2 x^2 \left[b^2 + (d-x)^2\right] = n_B^2 (a^2 + x^2)(d - x^2)$$

Now, taken the same coordinate system as the previous section, we write

$$\begin{aligned} x &= d_1 \sin\theta \\ b &= d_2 \cos\varphi \\ d - x &= d_2 \sin\varphi \\ a^2 + x^2 &= d_1^2 \end{aligned}$$

Then, by substitution we get Snell's law for transmission

$$n_A^2 d_1^2 \sin^2\theta \left[d_2^2 \cos^2\varphi + d_2^2 \sin^2\varphi\right] = n_B d_1^2 d_2^2 \sin^2\varphi \implies n_A \sin\theta = n_B \sin\varphi$$

Putting B in the same region as A (so $n_B = n_A = n$, $\varphi \to \theta_R$) we have the reflection law

$$d_1^2 d_2^2 \sin^2\theta = d_1^2 d_2^2 \sin^2\theta_R \implies \sin\theta = \sin\theta_R \tag{8.45}$$

§§ 8.3.3 Fresnel Equations

For now we only determined part of the behavior of the waves, since we are still missing what happens to the amplitudes.
Starting from equation (8.20) we rewrite the equations as

$$\begin{aligned} \mathbf{E} &= -\frac{1}{\epsilon\omega}\mathbf{k} \times \mathbf{H} \\ \mathbf{H} &= \frac{1}{\mu\omega}\mathbf{k} \times \mathbf{E} \end{aligned} \tag{8.46}$$

Now, the behavior at the boundary is described exactly from the boundary conditions (or jump conditions) of Maxwell's equations, therefore the polarization of the fields becomes important in the determination of the behavior! We will firstly consider cases where the $\mathbf{E}$ field is either parallel (p or π polarization, from German Parallel) or transverse (s or σ polarization, from German Senkrecht) to the *incidence plane*, i.e. the plane where all three waves lay.
Fresnel's idea was to find the ratio between the incoming amplitude and the reflected or transmitted wave. These ratios are known as *Fresnel coefficients*, and are strongly dependent on wave polarization. They are indicated as r, t
Consider firstly s polarization. The continuity relations are

$$\begin{aligned} \mathbf{E} \cdot \hat{\mathbf{t}} + \mathbf{E}_R \cdot \hat{\mathbf{t}} &= \mathbf{E}_T \cdot \hat{\mathbf{t}} \\ \mathbf{H} \cdot \hat{\mathbf{t}} - \mathbf{H}_R \cdot \hat{\mathbf{t}} &= \mathbf{H}_T \cdot \hat{\mathbf{t}} \end{aligned} \tag{8.47}$$

Calculating the dot products we have for $\mathbf{H}$

$$\begin{aligned} \mathbf{H} \cdot \hat{\mathbf{t}} &= H \cos\theta \\ \mathbf{H}_R \cdot \hat{\mathbf{t}} &= -H_R \cos\theta \\ \mathbf{H}_T \cdot \hat{\mathbf{t}} &= H_T \cos\varphi \end{aligned}$$

Therefore, using (8.46) we get the following system of equations

$$\begin{cases} E + E_R = E_T \\ kE\cos\theta - kE\cos\theta = k_T E_T \cos\varphi \end{cases} \tag{8.48}$$

Where $k = n_1 k_0$ and $k_T = n_2 k_0$.
Dividing by k_0 then we get

$$\begin{cases} E + E_R = E_T \\ n_1(E - E_R)\cos\theta = n_2 E_T \cos\varphi \end{cases}$$

Substituting the first equation into the second and dividing by E we have

$$\left.\frac{E_R}{E}\right|_s = r_s = \frac{n_1\cos\theta - n_2\cos\varphi}{n_2\cos\varphi + n_1\cos\theta}$$

Using $n = n_2/n_1$ we have the first coefficient, r_s

$$r_s = \frac{\cos\theta - n\cos\varphi}{\cos\theta + n\cos\varphi} \tag{8.49}$$

We can also write

$$E_R = E_T - E$$

Therefore

$$n_1(2E - E_T)\cos\theta = n_2 E_T \cos\varphi$$

Rearranging and dividing by E

$$\left.\frac{E_T}{E}\right|_s = t_s = \frac{2n_1\cos\theta}{n_1\cos\theta + n_2\cos\varphi}$$

Or

$$t_s = \frac{2\cos\theta}{\cos\theta + n\cos\varphi} \tag{8.50}$$

Now, we switch to the orthogonal polarization state, p polarization. Here, we have the same conditions for $\mathbf{H}$ as we had for $\mathbf{E}$ before and vice-versa, yielding

$$\begin{cases} H - H_R = H_T \\ E\cos\theta + E_r\cos\theta = E_t\cos\varphi \end{cases}$$

Which becomes, using $H = kE$

$$\begin{cases} kE - kE_R = k_T E_T \\ E\cos\theta + E_R\cos\theta = E_T\cos\varphi \end{cases}$$

Writing again $k = nk_0$ and dividing by k_0 we can write either

$$E_T = \frac{n_1}{n_2}(E - E_R)$$
$$E_R = E - \frac{n_2}{n_1}E_T$$

Inserting the first of the two into the system we get, after dividing by E

$$\left.\frac{E_R}{E}\right|_p = r_p = \frac{\cos\theta - \frac{n_1}{n_2}\cos\varphi}{-\frac{n_1}{n_2}\cos\varphi - \cos\theta}$$

Rearranging and inserting the relative refraction index

$$r_p = \frac{\cos\varphi - n\cos\theta}{\cos\varphi + n\cos\theta} \tag{8.51}$$

Using the second instead we have

$$2E\cos\theta = E_T\left(\cos\varphi + \frac{n_2}{n_1}\cos\theta\right)$$

Which, after rearrangement, gives

$$t_p = \frac{2\cos\theta}{\cos\varphi + n\cos\theta} \tag{8.52}$$

Therefore, reuniting the results obtained from both polarizations we have

$$\begin{aligned} r_s &= \frac{\cos\theta - n\cos\varphi}{\cos\theta + n\cos\varphi} \\ t_s &= \frac{2\cos\theta}{\cos\theta + n\cos\varphi} \\ r_p &= \frac{\cos\varphi - n\cos\theta}{\cos\varphi + n\cos\theta} \\ t_p &= \frac{2\cos\theta}{\cos\varphi + n\cos\theta} \end{aligned} \tag{8.53}$$

Using Snell's law we can eliminate also the relative refraction index using $n = \sin\theta\csc\varphi$, giving

$$\begin{aligned} r_s &= \frac{\cos\theta - \sin\theta\cos\varphi\csc\varphi}{\cos\theta + \sin\theta\cos\varphi\csc\varphi} = \frac{\cos\theta\sin\varphi - \sin\theta\cos\varphi}{\cos\theta\sin\varphi + \sin\theta\cos\varphi} \\ t_s &= \frac{2\cos\theta}{\cos\theta + \sin\theta\cos\varphi\csc\varphi} = \frac{2\cos\theta\sin\varphi}{\cos\theta\sin\varphi + \sin\theta\cos\varphi} \\ r_p &= \frac{\cos\varphi - \sin\theta\cos\theta\csc\varphi}{\cos\varphi + \sin\theta\cos\varphi\csc\varphi} = \frac{\cos\varphi\sin\varphi - \sin\theta\cos\theta}{\cos\varphi\sin\varphi + \sin\theta\cos\theta} \\ t_p &= \frac{2\cos\theta}{\cos\varphi + \cos\theta\sin\theta\csc\varphi} = \frac{2\cos\theta\sin\varphi}{\cos\varphi\sin\varphi + \sin\theta\cos\theta} \end{aligned} \tag{8.54}$$

Using the following trigonometric identities we can simplify things even more

$$\begin{aligned} \sin\theta\cos\phi - \cos\theta\sin\varphi &= \sin(\theta - \varphi) \\ \sin\theta\cos\theta - \cos\varphi\sin\varphi &= \sin(\theta - \varphi)\cos(\theta + \varphi) \\ \cos\varphi\sin\varphi + \sin\theta\cos\theta &= \sin(\theta + \varphi)\cos(\theta - \varphi) \end{aligned}$$

And in the end obtain what are known as *Fresnel equations* for reflection and refraction of electromagnetic waves

$$\begin{aligned} r_s &= -\frac{\sin(\theta-\varphi)}{\sin(\theta+\varphi)} \\ t_s &= \frac{2\cos\theta\sin\varphi}{\sin(\theta+\varphi)} \\ r_p &= -\frac{\tan(\theta-\varphi)}{\tan(\theta+\varphi)} \\ t_p &= \frac{2\cos\theta\sin\varphi}{\sin(\theta+\varphi)\cos(\theta-\varphi)} \end{aligned} \tag{8.55}$$

§§§ 8.3.3.1 Fresnel Equations for Irradiance and General Polarization

From equation (8.23) we can define the *luminous power of a beam* as

$$W = SI = \frac{n_i S}{2Z_0}E^2 \tag{8.56}$$

Where S is the surface of the cross-section of the beam.
It's not hard to then generalize this concept to reflected and refracted beams. Inserting n_1, n_2 in the previous formula, where both indexes are defined exactly as before we get

$$\begin{aligned} W &= \frac{n_1 S}{2Z_0}E^2 \\ W_R &= \frac{n_1 S_R}{2Z_0}E_R^2 \\ W_T &= \frac{n_2 S_T}{2Z_0}E_T^2 \end{aligned}$$

With S_R, S_T being the cross-sections of the respective beams.
We define the *reflectance* R and the *transmittance* T as the ratio of incoming vs reflected/transmitted power

$$\begin{aligned} R &= \frac{W_R}{W} = \frac{n_1 S_R}{2Z_0}\frac{2Z_0}{n_1 S}r^2 = r^2 \\ T &= \frac{W_T}{W} = \frac{n_2 S\cos\varphi}{2Z_0}\frac{2Z_0}{n_1 S\cos\theta}t^2 = \frac{n\cos\varphi}{\cos\theta}t^2 \end{aligned} \tag{8.57}$$

The polarization here can be either p or s. In general tho, light can be also be polarized in some random direction which can be described by a mix of p and s polarizations.
In the case of general polarization, we have

$$\begin{aligned} \mathbf{E} &= E_s\hat{\mathbf{s}} + E_p\hat{\mathbf{p}} \\ \mathbf{E}_R &= r_s E_s\hat{\mathbf{s}} + r_p E_p\hat{\mathbf{p}} \\ \mathbf{E}_T &= t_s E_s\hat{\mathbf{s}} + t_p E_p\hat{\mathbf{p}} \end{aligned}$$

Squaring and multiplying by n_i/Z_0 we get both I_R and I_T

$$I_R = \frac{n_1 S_R}{Z_0}\left(r_s^2 I_s + r_p^2 I_p\right)$$
$$I_T = \frac{n_2 S_T}{Z_0}\left(t_s^2 I_s + t_p^2 I_p\right)$$

Dividing by I_0 we have then

$$\begin{aligned} R &= \frac{I_R}{I_0} = r_s^2\frac{I_R}{I_0} + r_p^2\frac{I_T}{T_0} = R_s\frac{I_s}{I_0} + R_p\frac{I_p}{I_0} \\ T &= \frac{n\cos\varphi}{\cos\theta}\left(t_s^2\frac{I_s}{I_0} + t_p^2\frac{I_p}{I_0}\right) = T_s\frac{I_s}{I_0} + T_p\frac{I_p}{I_0} \end{aligned} \tag{8.58}$$

§ 8.4 Special Angles

Taken the equations (8.55) using simple mathematical analysis it's clear that there are some special angles for which there is some special behavior. These angles are known in literature as

- Brewster angle
- Near grazing incidence angle
- Normal incidence angle

§§ 8.4.1 Brewster Angle

The first one is the so called *Brewster angle*. Taken Fresnel's equations for the reflection coefficients

$$r_s = -\frac{\sin(\theta-\varphi)}{\sin(\theta+\varphi)}$$
$$r_p = -\frac{\tan(\theta-\varphi)}{\tan(\theta+\varphi)}$$

It's clear that r_s can never be 0 for $\theta \in [0, \pi/2]$, which is not the case for r_p. The angle, θ_B, for which $r_p(\theta_B) = 0$ is Brewster's angle.
At this incidence angle, all light reflected is s-polarized. From the tangent formula we have

$$r_p(\theta_B) = 0 \implies \varphi = \frac{\pi}{2} - \theta_B$$

From Snell's law we have that

$$\sin(\theta_B) = n\sin\left(\frac{\pi}{2} - \theta_B\right) = n\cos(\theta_B)$$

Therefore

$$n = \frac{\sin\theta_B}{\cos\theta_B}$$

Which gives

$$\theta_B = \arctan(n) \tag{8.59}$$

From the definition of relative index of refraction we can also say that

$$\theta_B = \begin{cases} > \dfrac{\pi}{4} & n_1 < n_2 \\ \dfrac{\pi}{4} & n_1 = n_2 \\ < \dfrac{\pi}{4} & n_1 > n_2 \end{cases} \tag{8.60}$$

§§ 8.4.2 Normal Incidence and Near Grazing Incidence

Another two special angles are at the limits of the interval, at $\theta = 0$ (normal incidence) and $\theta = \pi/2$ (grazing incidence).
At $\theta = 0$ we have

$$\sin\varphi = 0$$

I.e.

$$\begin{cases} r_s = \dfrac{1-n}{1+n} \\ t_s = \dfrac{2}{1+n} \\ r_p = \dfrac{n-n^2}{n+n^2} = r_s \\ t_p = \dfrac{2}{1+n} = t_s \end{cases} \tag{8.61}$$

At normal incidence then, the polarization of the wave is irrelevant, and the transmitted/reflected waves' amplitude will depend only on the properties of the material.
At $\theta = \pi/2$ instead we get

$$\sin\varphi = \frac{1}{n} \implies \cos\varphi = \frac{1}{n}\sqrt{n^2-1}$$

Which, in terms of Fresnel coefficients

$$\begin{cases} r_s = -\dfrac{\sqrt{n^2-1}}{\sqrt{n^2-1}} = -1 \\ r_p = 1 \\ t_s = t_p = 0 \end{cases} \tag{8.62}$$

I.e. the wave is only reflected, and the difference between the two polarization states is only a phase shift of π.

§ 8.5 Total Internal Reflection

For each angle of incidence between 0 and $\pi/2$ it's possible to distinguish two kinds of reflection

- External reflection, when $n > 1$
- Internal reflection, when $n < 1$

This distinction is clearer if we rewrite $\cos\varphi$ in terms of $\sin\theta$, as follows from Snell's law

$$\cos\varphi = \frac{1}{n}\sqrt{n^2 - \sin^2\theta}$$

It's then clear that in general, $\cos\varphi \in \mathbb{C}$, i.e. there exists an angle θ_c for which, when $\theta > \theta_c$ the cosine is complex. This angle is known as the *critical angle*, defined by

$$\sin\theta_c = n \implies \theta_c = \arcsin(n) \tag{8.63}$$

In the regime of $\theta \geq \theta_c$, all Fresnel coefficients are complex and, particularly, it can be proven that

$$\begin{aligned} r_s\overline{r_s} = r_p\overline{r_p} = 1 \\ t_s = t_p = 0 \end{aligned} \tag{8.64}$$

This gives the name to the regime that we're studying, *total internal reflection*, or TIR as we'll call it. The nonzero Fresnel coefficients are then r_s, r_p, which can be written explicitly complex using $\sqrt{n^2 - \sin^2\theta} = i\sqrt{\sin^2\theta - n^2}$

$$\begin{aligned} r_s &= \frac{\cos\theta - i\sqrt{\sin^2\theta - n^2}}{\cos\theta + i\sqrt{\sin^2\theta - n^2}} \\ r_p &= \frac{-n^2\cos\theta + i\sqrt{\sin^2\theta - n^2}}{n^2\cos\theta + i\sqrt{\sin^2\theta - n^2}} \end{aligned} \tag{8.65}$$

§§ 8.5.1 Fiber Optics

The main technological usage of TIR are fiber optics. These cables are built similarly to coaxial cables, where the cable (with refraction index n_1) is covered by a coating (with refraction index n_2) with materials which give $n = n_1/n_2 < 1$. In order to be in TIR regime, the beam must be incident at a particular angle, such that $\theta > \theta_c$. For purely technical reasons this angle is known as the *acceptance angle* α.
Define an angle related to this α as:

$$\frac{\pi}{2} - \alpha' = \theta_c$$

From the definition of θ_c we have then

$$\cos\alpha' = n$$

And therefore, also

$$\sin\alpha' = \frac{1}{n_1}\sqrt{n_1^2 - n_2^2}$$

We then define α as follows

$$\sin\alpha = n_1\sin\alpha' = \sqrt{n_1^2 - n_2^2} \tag{8.66}$$

Which implies

$$\alpha = \arcsin\left(\sqrt{n_1^2 - n_2^2}\right) \tag{8.67}$$

§§§ 8.5.1.1 Evanescent Waves

Consider now the field inside a fiber optic cable (in TIR regime in general). The field inside the cable is the field transmitted from the first reflection, and is

$$\mathbf{E}_T(\mathbf{r},t) = \mathbf{E}_T e^{i\mathbf{k}_T\cdot\mathbf{r}-i\omega t}$$

But, we have that

$$\mathbf{k}_T\cdot\mathbf{r} = k_T x\sin\varphi + k_T y\cos\varphi$$

But, from Snell's law, and the fundamental equation of trigonometry, that

$$\cos\varphi = \frac{1}{n}\sqrt{n^2-\sin^2\theta} = \frac{i}{n}\sqrt{\sin^2\theta-n^2}$$

Where we considered that we are in TIR, i.e. $\theta > \theta_C$ and $\sin^2\theta > n^2$ Therefore, said

$$a = \frac{k_T}{n}\sqrt{\sin^2\theta-n^2}$$
$$k_x = \frac{k_T\sin\theta}{n}$$

Which, if inserted into the field equation, give

$$\mathbf{E}_T(x,y,t) = \mathbf{E}_T e^{-a|y|}e^{ik_x x-i\omega t} \tag{8.68}$$

Measuring the field along the y axis, there's a part of the wave which decays exponentially, known as the *evanescent wave*. Note that the evanescent wave decays really rapidly, in fact, for visible light, $a \propto 10^6\ \mathrm{m}^{-1}$. It's clear that this wave will be really faint even at really close distances, and it's measurable only by really sensible detectors.

§§ 8.5.2 Phase Shift in TIR

One consequence of TIR is that, even if the amplitude of the wave isn't changed after each reflection ($|r|=1$), the complex nature of the Fresnel coefficients adds in a phase on the wave.
From basic complex analysis, since both $r_s, r_p \in \mathbb{C}$, we can write both the coefficients as a constant ($|r|$) times a complex exponential, which will have the added phase as its argument. Therefore, in general

$$r_s = e^{-i\delta_s} = \frac{ae^{-i\alpha}}{ae^{i\alpha}}$$
$$r_p = -e^{-i\delta_p} = -\frac{be^{-i\beta}}{be^{i\beta}}$$

Using the following fact

$$\tan(\theta) = \frac{\mathfrak{Im}\left\{re^{-i\theta}\right\}}{\mathfrak{Re}\left\{re^{-i\theta}\right\}}$$

We can write, using equations (8.65), and noting that $\delta_s = 2\alpha$, $\delta_p = 2\beta$ that

$$\tan(\alpha) = \tan\left(\frac{\delta_s}{2}\right) = \frac{\sqrt{\sin^2\theta-n^2}}{\cos\theta}$$
$$\tan(\beta) = \tan\left(\frac{\delta_p}{2}\right) = \frac{\sqrt{\sin^2\theta-n^2}}{n^2\cos\theta}$$

Which gives, as a relative phase difference $\Delta = \beta - \alpha$

$$\tan\left(\frac{\delta_p - \delta_s}{2}\right) = \tan\left(\frac{\Delta}{2}\right) = \frac{\cos\theta\sqrt{\sin^2\theta - n^2}}{\sin^2\theta} \tag{8.69}$$

Which gives

$$\Delta = 2\arctan\left(\frac{\cos\theta\sqrt{\sin^2\theta - n^2}}{\sin^2\theta}\right) \tag{8.70}$$

§§ 8.5.3 Fresnel's Rhomb

The results obtained from the previous section show that it's possible to polarize light using TIR and the right incidence angle. One optical instrument that serves this purpose is the so called *Fresnel rhomb*, a literal transparent rhomb that transforms linearly polarized light to circularly polarized light.
The rhomb has the two major sides inclined by 54 degrees (air-air). In this configuration, linearly polarized light at 45 degrees endures two total internal reflections and gets transmitted as circularly polarized light.
For each reflection, in this configuration, we get

$$\Delta = \frac{\pi}{4}$$

Summed up for both reflections, we get a total phase change between s and p polarized light of

$$\Delta_{tot} = \frac{\pi}{2}$$

Which is what defines circular polarization.

§ 8.6 Reflection and Transmission Matrices

As we saw before, polarization states can be described using a comfortable polarization basis. In the case of reflection and refraction this basis is obviously given by the versors $\hat{\mathbf{s}}$ and $\hat{\mathbf{p}}$, parallel to the perpendicular (s) or the parallel (p) direction with respect to the incidence plane. This system is obviously orthogonal, and the basis vectors are clearly described as

$$\hat{\mathbf{p}} = \begin{pmatrix}1\\0\end{pmatrix} \qquad \hat{\mathbf{s}} = \begin{pmatrix}0\\1\end{pmatrix} \tag{8.71}$$

A generally polarized wave therefore can be described as a linear composition of these two vectors, i.e.

$$\mathbf{E} = \begin{pmatrix}A\\B\end{pmatrix} = A\begin{pmatrix}1\\0\end{pmatrix} + B\begin{pmatrix}0\\1\end{pmatrix} = A\hat{\mathbf{p}} + B\hat{\mathbf{s}}$$

The reflection and the refraction of a beam, considered their linear behavior with respect to the generic wave, can then be described as a matrix. This matrix is

$$r^i_j = \begin{pmatrix}-r_p & 0\\0 & r_s\end{pmatrix} \tag{8.72}$$

Analogously, the transmission matrix is

$$t_j^i = \begin{pmatrix} t_p & 0 \\ 0 & t_s \end{pmatrix} \tag{8.73}$$

From these results it's possible to write the transmission and reflection matrices for special angles, especially for near grazing incidence and normal incidence.

- Normal incidence

$$r_j^i(0) = \frac{1-n}{1+n}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \tag{8.74}$$

- Near grazing incidence/grazing incidence

$$r_j^i\left(\frac{\pi}{2}\right) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \tag{8.75}$$

For TIR the results are slightly different, due to the complex nature of r. As we wrote before, we have

$$\begin{aligned} r_s &= e^{-i\delta_s} \\ r_p &= -e^{-i\delta_p} \end{aligned}$$

Therefore

$$r_j^i = \begin{pmatrix} -e^{-i\delta_p} & 0 \\ 0 & e^{-i\delta_s} \end{pmatrix} = e^{-i\delta_p}\begin{pmatrix} -1 & 0 \\ 0 & e^{i\Delta} \end{pmatrix} \tag{8.76}$$

With this definition, the dephasement ensued after TIR is really vivid.
Note that applying this matrix to a generally polarized wave $\mathbf{E}$ we have in general an elliptically polarized outgoing wave, as

$$E_R^i = r_j^i E^j = e^{-i\delta_p}\begin{pmatrix} -1 & 0 \\ 0 & e^{i\Delta} \end{pmatrix}\begin{pmatrix} A \\ B \end{pmatrix} = e^{-i\delta_p}\begin{pmatrix} -A \\ Be^{i\Delta} \end{pmatrix} \tag{8.77}$$

Note that there is a generic phase change outside the resulting wave (which doesn't affect the polarization state), while there's a dephasement on the second component, which depends directly on $\delta_p - \delta_s$, that factor, on the other hand, is the one that will change the polarization state into, generally, an elliptical state.
In the most general case, we can define the *Fresnel-Jones matrices*, which are the most general expression of reflection-refraction problems of polarized waves

$$\begin{aligned} r_j^i &= -\begin{pmatrix} \tan(\theta-\varphi)\cot(\theta+\varphi) & 0 \\ 0 & \sin(\theta-\varphi)\csc(\theta+\varphi) \end{pmatrix} \\ t_j^i &= \frac{2\cos\theta\sin\phi}{\sin(\theta+\varphi)}\begin{pmatrix} \sec(\theta-\varphi) & 0 \\ 0 & 1 \end{pmatrix} \end{aligned} \tag{8.78}$$

§ 8.7 Theory of Multilayer Films

The laws of reflection and refraction can be used in multiple ways in order to create special optically active instruments, like antireflecting lenses for glasses, highly-reflecting lenses, heat-reflecting or heat-transmitting surfaces and lenses which reflect only one specific wavelength of light.
These peculiar instruments, which some of them are also really common (see antireflecting glasses or blue-light protection glasses) are made by superimposing various thin strata of material with different refraction indexes n_i.

§§ 8.7.1 Transfer Matrix

In order to study the physics behind, we'll start with only two layers, which separate an "incoming" zone with index n_0 from an "outgoing" zone with index n_T, with index n_1 in between.
In this case, we will have, after applying simple logic and the idea of reflection and transmission of light:

- An incoming wave, *outside* of the film, where the reflection index is some value n_0, and the wave is described by $(\mathbf{k}, \mathbf{E}, \mathbf{H})$
- A wave reflected from the first layer (it's still outside of the layer, hence the refraction index is n_0), described by $(\mathbf{k}_R, \mathbf{E}_R, \mathbf{H}_R)$
- A wave transmitted from the first layer, here the refraction index is n_1, described by $(\mathbf{k}_1, \mathbf{E}_1, \mathbf{H}_1)$
- A wave reflected from the second layer, (here the refraction index is *still* n_1), described by $(\mathbf{k}'_R, \mathbf{E}'_R, \mathbf{H}'_R)$
- A wave transmitted from the second layer, finally outside the film, where the refraction index is n_T and the wave will be $(\mathbf{k}_T, \mathbf{E}_T, \mathbf{H}_T)$

If we choose, for ease of calculation, the origin at the first layer, the equations for this layer will be similar to what we saw before for a single reflection and refraction. The same goes for the second layer, remembering tho that after having traveled a distance l we will have $E' = Ee^{ikl}$.
Everything considered then we will have a system of four coupled equations in four unknowns, which we will separate into two coupled systems that represent the two different layers. We chose, for ease of calculation, to evaluate only normal reflection.

$$\begin{cases} E + E_R = E_1 + E'_R \\ n_0 (E - E_R) = n_1 (E_1 - E'_R) \end{cases} \qquad \begin{cases} E_1 e^{ikl} + E'_R e^{-ikl} = E_T \\ n_1 \left(E_1 e^{ikl} - E'_R e^{-ikl}\right) = n_T E_T \end{cases} \tag{8.79}$$

Where we used $H_i = (\mu\omega)^{-1} n_i k_0 E_i$ on the second equation of both systems. We begin by solving for the second layer. From the second equation we have

$$E_T = \frac{n_1}{n_T} \left(E_1 e^{ikl} - E'_R e^{-ikl}\right)$$

Then, inserting it in the first equation

$$E_1 e^{ikl} + E'_R e^{-ikl} = \frac{n_1}{n_T} \left(E_1 e^{ikl} - E'_R e^{-ikl}\right)$$

Reuniting the terms with the same amplitude we get

$$\left(1-\frac{n_1}{n_T}\right)E_1e^{ikl}=-\left(1+\frac{n_1}{n_T}\right)E'_Re^{-ikl}$$

Or, developing the sum, multiplying by the inverse of the result on the left and on the right and dividing by E'_R, we have

$$\frac{E_1}{E'_R}=\frac{n_1+n_T}{n_1-n_T}e^{-2ikl}\qquad\frac{E'_R}{E_1}=\frac{n_1-n_T}{n_1+n_T}e^{2ikl}$$

Taken again the first equation of the second layer in (8.79) and dividing everything by E_1 and E'_R, we find ourselves a new solvable system

$$\begin{cases}e^{ikl}+\dfrac{E'_R}{E_1}e^{-ikl}=\dfrac{E_T}{E_1}\\ \dfrac{E_1}{E'_R}e^{ikl}+e^{-ikl}=\dfrac{E_T}{E'_R}\end{cases}$$

Inserting what we found before we have, for the first equation

$$e^{ikl}+\left(\frac{n_1-n_T}{n_1+n_T}e^{2ikl}\right)e^{-ikl}=\frac{E_T}{E_1}$$

Or, solving by regrouping the exponentials

$$\frac{E_T}{E_1}=\frac{2n_1}{n_1+n_T}e^{ikl}$$

Analogously for the second equation we get

$$\frac{E_T}{E'_R}=\frac{2n_1}{n_1-n_T}e^{-ikl}$$

Rewriting E_1, E'_R in terms of E_T, the whole system (8.79) becomes

$$\begin{cases}E+E_R=E_1+E'_R\\ n_0\left(E-E_R\right)=n_1\left(E_1-E'_R\right)\end{cases}\qquad\begin{cases}E'_R=\dfrac{n_1-n_T}{2n_1}E_Te^{ikl}\\ E_1=\dfrac{n_1+n_T}{2n_1}E_Te^{-ikl}\end{cases}\tag{8.80}$$

Inserting the results of the second system into the first we get

$$\left\{E+E_R=\left(\frac{n_1-n_T}{2n_1}e^{-ikl}+\frac{n_1-n_T}{2n_1}e^{ikl}\right)E_Tn_0\left(E+E_R\right)\quad=n_1\left(\frac{n_1-n_T}{2n_1}e^{-ikl}-\frac{n_1-n_T}{2n_1}e^{ikl}\right)E_T\right.$$

Regrouping on the right with respect to the refraction indexes we get on the first equation

$$E+E_R=\frac{1}{2n_1}\left[n_1\left(e^{-ikl}+e^{ikl}\right)+n_T\left(e^{-ikl}-e^{ikl}\right)\right]E_T$$

Using the usual identities for sines and cosines and bringing in the factor on the left we get

$$E + E_R = \left(\cos(kl) - i\frac{n_T}{n_1}\sin(kl)\right) E_T$$

And analogously, for the second equation

$$n_0\,(E - E_R) = (-in_1\sin(kl) + n_T\cos(kl))\,E_T$$

Remembering the definition of r and t, we can rewrite the system in terms of both Fresnel parameters dividing by E

$$\begin{cases} 1 + r = \left(\cos kl - i\frac{n_T}{n_1}\sin(kl)\right) t \\ n_0\,(1 - r) = (-in_1\sin(kl) + n_T\cos(kl))\,t \end{cases} \tag{8.81}$$

This system can clearly be rewritten in matrix terms as follows

$$\begin{pmatrix} 1 \\ n_0 \end{pmatrix} + \begin{pmatrix} 1 \\ -n_0 \end{pmatrix} r = \begin{pmatrix} \cos(kl) & -\frac{i}{n_1}\sin(kl) \\ -in_1\sin(kl) & \cos(kl) \end{pmatrix} \begin{pmatrix} 1 \\ n_T \end{pmatrix} t \tag{8.82}$$

Here, the left hand side represent the discontinuity equations for the first layer, while the right hand side represents the discontinuity equations on the second layer. The matrix, known as the *transfer matrix*, "transfers" the equation from the first boundary to the second, where the discontinuity equations are applied.
Generalizing to multilayer films composed by n layers, we can immediately find the transfer matrix of the whole system using simple matrix multiplication. Said M_j^i the transfer matrix of this system, then we have, if the matrix of the $k-$th layer is $M_{j_k}^{i_k}$ that

$$\prod_{k=1}^{n} M_{j_k}^{i_k} = M_j^i = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{8.83}$$

Therefore, the general Fresnel equation becomes

$$\begin{pmatrix} 1 \\ n_0 \end{pmatrix} + \begin{pmatrix} 1 \\ -n_0 \end{pmatrix} r = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} 1 \\ n_T \end{pmatrix} t \tag{8.84}$$

This equation is readily solvable for both r, t. Rewriting everything as a system of two equations, we have

$$\begin{cases} 1 + r = (A + Bn_T)\,t \\ n_0(1 - r) = (C + Dn_T)\,t \end{cases}$$

From the first equation we get $r(t)$, which inserted in the second equation gives

$$2n_0 - n_0\,(A + n_T B)\,t = (C + n_T)t$$

Which solved gives

$$t = \frac{2n_0}{n_0 A + n_0 n_T B + n_T D + C} \tag{8.85}$$

Reinserting it into the first equation, we have also the result for r

$$r = \frac{n_0 A + n_0 n_T B - n_T D - C}{n_0 A + n_0 n_T B + n_T D + C} \tag{8.86}$$

If the elements of the transmission matrix are known, then it's possible to find the reflectance and transmittance of the complete multilayer by remembering that $R = \|r\|^2$, $T = \|t\|^2$

§§ 8.7.2 Antireflecting Films

One of the applications of multilayer theory we indicated before are *antireflecting films*, which are the reason that we can build and use antireflecting lenses in cameras and glasses, and for the watch enthusiasts, also have antireflecting sapphire crystals.
Considering this real-world usage, we will suppose that the antireflecting film is placed over a glass (or sapphire) surface with refraction index n_T, and on the other side the light is immersed in air, with refraction index $n_0 = 1$, while the layer of antireflecting material has refraction index n_1.
The transfer matrix of the system is

$$M_j^i = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} \cos(kl) & -\frac{i}{n_1}\sin(kl) \\ -in_1 \sin(kl) & \cos(kl) \end{pmatrix}$$

The equation for the reflection coefficient is as before (8.86), which, after substituting the matrix elements and regrouping the common factor is

$$r = \frac{(1 - n_T)\cos(kl) - i\left(\frac{n_T}{n_1} - n_1\right)\sin(kl)}{(1 + n_T)\cos(kl) - i\left(\frac{n_T}{n_1} + n_1\right)\sin(kl)}$$

Multiplying and dividing both denominator and numerator we get the much simpler result

$$r = \frac{n_1(1 - n_T)\cos(kl) - i(n_T - n_1^2)\sin(kl)}{n_1(1 + n_T)\cos(kl) - i(n_T + n_1^2)\sin(kl)} \tag{8.87}$$

Now, here the reflectance is not necessarily 0, but we can manipulate the parameters of the layer in order to get it to become null for the desired frequency.
The easiest to mechanically manipulate is the thickness l of the layer. In fact, if we take $l = \lambda/4$ we have

$$kl = \frac{k\lambda}{4} = \frac{2\pi\lambda}{4\lambda} = \frac{\pi}{2} \tag{8.88}$$

Inserting this awesome result we get for r and R

$$r = \frac{n_T - n_1^2}{n_T + n_1^2}, \qquad R = \frac{\left(n_T - n_1^2\right)^2}{\left(n_T + n_1^2\right)^2} \tag{8.89}$$

Now, we have a constraint on n_1, i.e. the composition of the layer. In order to have 0 reflectance at the chosen frequency λ we must find a material such that

$$n_1 = \sqrt{n_T} \tag{8.90}$$

In common use for real-world application, like antireflecting glasses (i.e. $n_T \approx 1.5$), is magnesium fluoride, where $n_{MgF_2} = n_1 \approx 1.35$. This combination doesn't respect precisely the previous constraint, but permits a reduction of the reflectance of the glasses by 99%. For negating completely the reflection of a chosen wavelength, multiple layers are used, where specifically they're alternated high-low reflectance films thick $\lambda/4$, with only 3 layers it's possible to block 3 different wavelengths. CHECK PROPERLY, I wrote $\lambda/4$ and not $\lambda_0/4$! It's not an error, the thickness depends directly on the material and its composition, since it also changes the refraction index, and therefore λ *will* differ.

§§ 8.7.3 High Reflectance Films

The opposite result of what we got with antireflecting films, is that of high-reflectance films, special films built by alternating layers thick $l = \lambda/4$ of high and low refracting indexes.
Said n_H the refraction index of the highly refracting material and n_L the refraction index of the low-refracting material (obviously $n_H > n_L$) we have, since they're thick exactly $\lambda/4$, that the transfer matrix of such system is

$$\left(M_{H/L}\right)^i_j = \begin{pmatrix} 0 & -\frac{i}{n_{H/L}} \\ -in_{H/L} & 0 \end{pmatrix} \tag{8.91}$$

For the first high-low couple of layers the transfer matrix is simply the product of the two

$$(M_{HL})^i_j = \begin{pmatrix} 0 & -\frac{i}{n_L} \\ -in_L & 0 \end{pmatrix} \begin{pmatrix} 0 & -\frac{i}{n_H} \\ -in_H & 0 \end{pmatrix} = \begin{pmatrix} -\frac{n_H}{n_L} & 0 \\ 0 & -\frac{n_L}{n_H} \end{pmatrix}$$

Alternating $2N$ stacks of high-low reflectance layer, the transfer matrix is found simply using matrix powers, i.e.

$$M^i_j = \left(M^N_{HL}\right)^i_j = (-1)^N \begin{pmatrix} (n_H/n_L)^N & 0 \\ 0 & (n_L/n_H)^N \end{pmatrix} \tag{8.92}$$

From (8.86) then, we have that the reflection coefficient for a high-reflectance film is, supposed for ease of calculation $n_0 = n_T = 1$

$$r = \frac{\left(\frac{n_H}{n_L}\right)^N - \left(\frac{n_L}{n_H}\right)^N}{\left(\frac{n_H}{n_L}\right)^N + \left(\frac{n_L}{n_H}\right)^N} \tag{8.93}$$

Which gives a reflectance of

$$R = \frac{\left(n_H^{2N} - n_L^{2N}\right)^2}{\left(n_H^{2N} + n_L^{2N}\right)} \tag{8.94}$$

The reflectance clearly goes to unity with $N \to \infty$, i.e., the more stacks we put the more the film reflects a specific frequency. In the case of layers of zinc sulfide (ZnS, $n_H = 2.3$) and magnesium fluoride ($\mathrm{MgF_2}$, $n_L = 1.35$), with a stack of 8 layers (4 layers of each, or 4 couples of high-low layers), the reflectance is $R \simeq 0.97$, which is higher than that of silver, in the visible range. A stack of 30 layers gives $R \gtrsim 0.999$. The highly-reflected wavelenght band can be broadened by the simple combination of multiple thicknesses of the layers

9 Coherence and Interference

§ 9.1 Interference

Suppose that we have a single point-like source S which emits an electromagnetic wave $\mathbf{E}$, which passes through two point apertures S_1, S_2, and then converges again to a point P on which we put some detector.
The starting wave is described as the sum of two single waves coming out the pinholes via the superposition principle, i.e.

$$\mathbf{E} = \mathbf{E}_{(1)} + \mathbf{E}_{(2)}$$

Where, in general, the two wavelets can be described generally as usual

$$\begin{aligned} \mathbf{E}_{(1)} &= \mathbf{E}_1 e^{i\mathbf{k}_1 \cdot \mathbf{r} - i\omega t + i\phi_1} \\ \mathbf{E}_{(2)} &= \mathbf{E}_2 e^{i\mathbf{k}_2 \cdot \mathbf{r} - i\omega t + i\phi_2} \end{aligned} \tag{9.1}$$

Due to the constraint of the system we have that $\mathbf{k}_1 = \mathbf{k}_2$, therefore the sum of the two gives the general field, which in general depends only on the phase difference of the two wavelets.
These two wavelets are said to be *mutually coherent*, if the phase difference between the two is constant, i.e.

$$\Delta\phi = \phi_1 - \phi_2 = \text{ const} \tag{9.2}$$

This definition comes up directly when we evaluate the irradiance of the measured field $\mathbf{E}$. In fact, we have

$$I = \sqrt{\frac{\mu}{\epsilon}} E^2 = \sqrt{\frac{\epsilon}{\mu}} \left[E_1^2 + E_2^2 + 2\mathfrak{Re}\left\{ \mathbf{E}_{(1)} \cdot \mathbf{E}_{(2)} \right\} \right]$$

Remembering that $\sqrt{\epsilon/\mu} = nZ_0^{-1}$, and rewriting $nE_i^2/Z_0 = I_i$ we have that the total irradiance measured is (in the general case with two different waves)

$$I = I_1 + I_2 + 2\sqrt{\frac{\mu}{\epsilon} I_1 I_2} \cos\left(\mathbf{k}_1 \cdot \mathbf{r} - \mathbf{k}_2 \cdot \mathbf{r} + \phi_1 - \phi_2 \right) \tag{9.3}$$

The last term, which depends on the root of the product of the intensities of the single waves, is called the *interference term*. This term is the only of the three that depends on the "angle" θ, defined as

$$\Delta(\mathbf{r}) = (\mathbf{k}_1 - \mathbf{k}_2) \cdot \mathbf{r} + \phi_1 - \phi_2 \tag{9.4}$$

This last interference term is clearly dependent on the polarization of the two waves. In fact, it can be exactly zero when the scalar product of the two fields of the single waves is zero. I.e. when their polarization is orthogonal.
In laboratories, for obvious reasons we never measure the instantaneous irradiance, but an average. Also, the interfering waves must not have the same frequencies and can come from different sources, therefore the total irradiance we measure will instead be described by the following equation

$$I = \langle I \rangle = \lim_{T\to\infty} \frac{1}{T}\sqrt{\frac{\epsilon}{\mu}} \int_0^T \left(\mathbf{E}_{(1)} + \mathbf{E}_{(2)}\right) \cdot \left(\mathbf{E}_{(1)} + \mathbf{E}_{(2)}\right)^\dagger \mathrm{d}t \tag{9.5}$$

The interference term I_{int} is therefore defined as follows

$$I_{int} = 2\mathbf{E}_1 \cdot \mathbf{E}_2 \sqrt{\frac{\epsilon}{\mu}} \lim_{T\to\infty} \frac{1}{T} \int_0^T \cos\left((\mathbf{k}_1 - \mathbf{k}_2)\cdot\mathbf{r} - (\omega_1 - \omega_2)t + \phi_1 - \phi_2\right) \mathrm{d}t \tag{9.6}$$

The integral at the end is zero only in a handful of cases, in fact

$$I_{int} = 0 \implies \begin{cases} \mathbf{E}_1 \perp \mathbf{E}_2 \\ \omega_1 \neq \omega_2 \\ \phi_1 - \phi_2 \neq \text{ const} \end{cases} \tag{9.7}$$

§§ 9.1.1 Double Slit Interferometer

The first experiment with interference was prepared by Thomas Young in the early 1800s. This experimental setup, better known as the *double slit experiment* is set up, ideally, as a point source which emits a single wave; this wave then comes through two slits and then a detecting screen shows the interference pattern.
In general we can say that the two sections of the experiment (one where lays the source, and one where lays the screen) have different refraction indexes $n_1\, n_2$, for better emphasizing this we write $\mathbf{k} = n\mathbf{k}_0$. Since the electromagnetic wave comes from a single source, we have to impose that the two wavelets coming out from the slits into the screen have the same color, i.e. $\omega_1 = \omega_2 = \omega$, therefore, the interference term will be

$$I_{int} = 2\mathbf{E}_1 \cdot \mathbf{E}_2 \sqrt{\frac{\epsilon}{\mu}} \lim_{T\to\infty} \frac{1}{T} \int_0^T \cos\left(\mathbf{k}_0\left(n_1\mathbf{r}_1 - n_2\mathbf{r}_2\right) + \Delta\phi\right) \mathrm{d}t$$

Due to the different position of the two slits, and since their distance from the measuring point on the screen is not necessarily equal, we have that the two single wavelets will be described by two different $\mathbf{r}$ vectors.
Now, since the two incoming wavelets are not necessarily parallel, we can write the following

$$\mathbf{E}_1 \cdot \mathbf{E}_2 = \sqrt{I_1 I_2}\cos\delta$$

And therefore, in general, the irradiance will be

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\delta\cos\Delta \tag{9.8}$$

Where we omitted the time average (as we will do going forward from now, for notational ease).
Now, due to the origin of the two waves, we know that I_1, I_2 don't depend on time and are constants, but it's clear that depending on the position of the measuring point on the screen that $I_{min} \leq I \leq I_{max}$, all determined by the last cosine factor of the interference term.
Since $-1 \leq \cos\Delta \leq 1$ we will have

$$I = \begin{cases} I_{max} & \Delta = 2m\pi \\ I_{min} & \Delta = (2m+1)\pi \end{cases} \qquad m \in \mathbb{Z} \tag{9.9}$$

This result can be reproduced also thinking in strictly trigonometric terms. Said L the distance from the slits to the screens and said d the distance from the half-point between the slits and the measuring point on the screen, we have, called x' the measuring point, that, firstly

$$I = I_0 + 2I_0\cos\Delta = I\left(1 + 2\cos\Delta\right) = 4I\cos^2\left(\frac{\Delta}{2}\right)$$

And then that

$$\frac{\Delta}{2} \approxeq \frac{\pi d}{\lambda_0}\frac{x'}{L} \tag{9.10}$$

I.e.

$$I_t = 4I_0\cos^2\left(\frac{\pi d}{\lambda_0}\frac{x'}{L}\right)$$

Then, the maxima and minima of the irradiance measured on the screen can be described all in terms of wavelength and distance from the screen, i.e.

$$I = \begin{cases} I_{max} & \Delta = \dfrac{m\lambda_0}{d} \\ I_{min} & \Delta = \dfrac{\lambda_0}{2d}(2m+1) \end{cases} \qquad m \in \mathbb{Z} \tag{9.11}$$

§§ 9.1.2 Michelson-Morley Interferometer

Another experiment demonstrating the interference of electromagnetic waves was made in the later years of the 1800s by Michelson and Morley. This experiment was of critical importance also in other branches, like mechanics, in fact it proved that without doubt there was no ether in space, but rather gave the foundation to the special relativistic Lorentz transformations.
This experiment is composed by a point-like source, which passes through a *beam splitter*, i.e. an optically active object which divides an incoming beam into two separate beams.
This beam splitter is oriented in a way such that the two outgoing beams are orthogonal between each other, and after a *different* distance for each (d_1, d_2) they meet again at the beam splitter, which aligns them back up again and sends them to a receiver.
Said d_i the optical path traveled by the two waves we have that

$$2d = 2(d_1 - d_2)$$

Then, the total irradiance at the receiver will be

$$I = 4I_0\cos^2\left(\frac{\pi}{\lambda}2d\right) \tag{9.12}$$

Then, as we did before with the double slit interferometer, we will have by finding the maximum and the minimum of the cosine

$$I = \begin{cases} I_{max} & 2d = m\lambda \\ I_{min} & 2d = \left(m + \frac{1}{2}\right)\lambda \end{cases} \qquad m \in \mathbb{Z} \tag{9.13}$$

This kind of interferometer is still widely in use in the world of modern physics, in fact it's the same kind of interferometer used in LIGO and VIRGO experiments (albeit much more modern). These interferometers use arms long in the order of kilometers for detecting the slightest change in the interference pattern given by the oscillation of the optical path length at the passage of a gravitational wave. Here the point source chosen is a high-power laser.
Another usage of the Michelson-Morley interferometer will see the light in the future with the launch of the LISA constellation of satellites, which will use huge distances for detecting the faintest gravitational waves.

§ 9.2 Partial Coherence

§§ 9.2.1 Correlation

It's time to consider the most generic case possible, and the nature of interference itself.
Consider two beams coming from a point-like source S, with the same frequency and polarization. Chosen a point R on which the two beams rejoin after two different paths l_1, l_2, we know for a fact (since electromagnetic waves travel at a constant speed $u = c/n$), that one of the two beams will arrive at a later time at the point.
Said $l_1 < l_2$, t being the time needed to cross the path long l_1 and $t + \tau$ being the time needed to cross the path l_2, we have

$$\begin{aligned} \mathbf{E}_{(1)}(t) &= \mathbf{E}_1 e^{i\mathbf{k}\cdot\mathbf{r} - i\omega t + i\phi_1(t)} \\ \mathbf{E}_{(2)}(t+\tau) &= \mathbf{E}_2 e^{i\mathbf{k}\cdot\mathbf{r} - i\omega t + i\phi_2(t+\tau)} \end{aligned} \tag{9.14}$$

Where we consider the randomness of phase by noting that it's time dependent.
By definition of what we said so far, then

$$\tau = \frac{1}{u}|l_2 - l_1| = \frac{\Delta l}{u} \tag{9.15}$$

And, therefore the irradiance at the point R will then be defined by the usual equation

$$I = I_1 + I_2 + 2Z\left\langle \mathfrak{Re}\left\{\mathbf{E}_1(t)\cdot\mathbf{E}_2^\dagger(t+\tau)\right\}\right\rangle$$

Definition 9.2.1 (Correlation Function)**.** From the last equation it's possible to define a new integral function, which we will call the *correlation function* or the *mutual correlation*. This function is defined as a function of τ, as follows

$$\Gamma_{12}(\tau) = \lim_{T\to\infty} \frac{1}{T}\sqrt{\frac{\epsilon}{\mu}} \int_0^T \mathbf{E}_1(t)\mathbf{E}_2^\dagger(t+\tau)\mathrm{d}t \tag{9.16}$$

From the definition, it's clear that

$$\Gamma_{12}(\tau) = \sqrt{\frac{\epsilon}{\mu}} \left\langle \mathbf{E}_1(t)\mathbf{E}_2^\dagger(t+\tau) \right\rangle$$

We also define the *autocorrelation function* as

$$\Gamma(\tau) = \Gamma_{ii}(\tau) = \sqrt{\frac{\epsilon}{\mu}} \left\langle \mathbf{E}_i(t)\mathbf{E}_i^\dagger(t+\tau) \right\rangle \tag{9.17}$$

Also, by substitution, we can say that

$$\Gamma_{12}(\tau) = \frac{1}{2} I_{int}$$

And also that

$$\Gamma_{11}(0) = I_1$$
$$\Gamma_{22}(0) = I_2$$

Another useful function derived by the correlation function is the normalized version, known as the *degree of correlation* $\gamma_{ij}(\tau)$.
It's defined as follows

$$\gamma_{ij}(\tau) = \frac{\Gamma_{ij}(\tau)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}} \tag{9.18}$$

From the previous definitions it's possible then to see how the interaction term depends on the correlation between the two waves, and specifically, on the phase difference between the two. Being γ a complex function, we can write

$$\mathfrak{Re}\left\{\gamma_{ij}(\tau)\right\} = |\gamma|_{ij}(\tau)\cos\left(\Delta\phi_{ij}\right)$$

Then, the interaction term can be defined in the most general way as follows

$$I_{int} = 2\sqrt{I_1 I_2}|\gamma_{12}(\tau)|\cos\left(\Delta\phi_{12}\right) \tag{9.19}$$

The definition of the autocorrelation as a normalized term of measure of the correlation of two waves, gives rise to the definitions of *perfect incoherence*, *partial coherence* and *perfect coherence*, respectively when $|\gamma_{12}| = 1$, $|\gamma_{12}| = 0$ and $0 < |\gamma_{12}| < 1$.
Therefore, the range of interference is exactly in the set

$$I_{int} \in \left[-2\sqrt{I_1 I_2}|\gamma_{12}|, \sqrt{I_1 I_2}|\gamma_{12}|\right] \tag{9.20}$$

Definition 9.2.2 (Visibility)**.** Another useful definition is the *visibility of fringes* $\mathcal{V}$, a constant comprised between $0, 1$, defined as the ratio of the difference of irradiance between the peaks and the shadows of the interference pattern and the sum of the two intensities

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} = \frac{2\sqrt{I_1 I_2}|\gamma_{12}(\tau)|}{I_1 + I_2} \tag{9.21}$$

Note also that if $I_1 = I_2 = I_0$, clearly

$$\mathcal{V} = |\gamma_{12}(\tau)|$$

Which implies that the maximum visibility will be obtained when the two waves are in the regime of *total coherence*.

§§ 9.2.2 Coherence Time and Coherence Length

From what we have seen before, the degree of correlation is strictly tied to the signal and its properties, but especially to its phase.
Consider a quasimonochromatic wave ($\Delta\omega \approx 0$). The field is then described as follows

$$\mathbf{E}(\mathbf{r},t) = \mathbf{E}_0(\mathbf{r})e^{-i\omega t + i\phi(t)} \tag{9.22}$$

The phase function $\phi(t)$ is as we said before a random function of time. Physically, we can see this function as a composition of multiple Heaviside step functions, or, more clearly, it describes intervals of coherence ($\phi(t) =$ const) and instants of decoherence.
Said τ_0 the *coherence time*, i.e. the time needed for $\phi(t)$ to change from one constant value to the other, we can begin to analyze the behavior of $\gamma(\tau)$ in different occasions.
Since

$$\mathbf{E}(\mathbf{r},t) = \mathbf{E}_0(\mathbf{r})e^{-i\omega t}e^{i\phi(t)}$$
$$\mathbf{E}(\mathbf{r},t+\tau) = \mathbf{E}_0(\mathbf{r})e^{-i\omega(t+\tau)}e^{i\phi(t+\tau)}$$

The scalar product of $\mathbf{E}(t)$ and $\mathbf{E}(t+\tau)$ is then (omitting the spatial dependence, since it does not interfere with our calculations)

$$\mathbf{E}(t)\cdot\mathbf{E}^\dagger(t+\tau) = E_0^2 e^{-i\omega\tau}e^{i(\phi(t)-\phi(t+\tau))}$$

Therefore, the degree of correlation depends only on the difference of the two phases

$$\gamma(\tau) = \frac{1}{\langle E^2\rangle}\left\langle \mathbf{E}(t)\cdot\mathbf{E}^\dagger(t+\tau)\right\rangle = e^{-i\omega\tau}\left\langle e^{i(\phi(t))-\phi(\tau)}\right\rangle$$

Since we defined $\phi(t)$ as a periodic (random) step function with a period of τ_0, the expected value is

$$\langle\phi(t)-\phi(t+\tau)\rangle = \begin{cases} 0 & \tau > \tau_0 \ \vee \ 0 < t < \tau_0 - \tau \\ \Delta\phi & \tau_0 - \tau < t < \tau_0 \end{cases} \tag{9.23}$$

Or, in common words, it's zero if we're evaluating the coherence when $\phi(t+\tau)$ has changed already to another random value, or vice-versa when $\phi(t)$ has not yet reached the coherence time τ_0. It's equal to a constant value $\Delta\phi$ only and only when we're checking in an interval which is not greater than the coherence time.
Considering only the first interval of coherence, then

$$\gamma(\tau) = \frac{e^{-i\omega\tau}}{\tau_0}\left[\int_0^{\tau_0-\tau}\mathrm{d}t + e^{i\Delta\phi}\int_0^{\tau_0}\mathrm{d}t\right] = \left(\frac{\tau_0-\tau}{\tau_0} + \frac{\tau}{\tau_0}e^{i\Delta\phi}\right)e^{-i\omega\tau}$$

Or, evaluating the integrals we have, in general, for a single cycle (or, in common terms, in a single *wave train*) the coherence is strictly tied to the coherence time τ_0, with the equation

$$\gamma(\tau) = \begin{cases} \dfrac{\tau_0-\tau}{\tau_0}e^{-i\omega\tau} & \tau < \tau_0 \\ 0 & \tau > \tau_0 \end{cases} \tag{9.24}$$

Note that, if we take the evaluation of the integral and check it's expectation value for $T \to \infty$, the expected value is zero, due to the random variation of phase $\Delta\phi$. Therefore, a wave will tend to

decoherence for big periods.
Note also that, since we're not considering two different waves, if there are no attenuation phenomena, the absolute value of the degree of correlation is the visibility of fringes, i.e.

$$\mathcal{V} = 1 - \frac{\tau}{\tau_0} \tag{9.25}$$

Considered everything and evaluated the real part of what we found before we have

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\mathfrak{Re}\left\{\frac{\tau_0 - \tau}{\tau_0}e^{-i\omega\tau}\right\} \quad \tau < \tau_0 \tag{9.26}$$

Which, if we develop the last operation on the right becomes

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\left(1 - \frac{\tau}{\tau_0}\right)\cos(\omega\tau) \quad \tau < \tau_0 \tag{9.27}$$

Noting that on the interference term we have the real part of the degree of self-correlation, we can say with ease that the interference pattern will be present only when τ is less than the coherence time τ_0, therefore indicating that peak-shadow patterns only appear with coherent light.
The parameters *coherence length* and *coherence time* are two intrinsic parameters of the wave, which indicate single coherent packets of light ,or *wave trains*, where the first can be understood experimentally as the length of the wave train.
Note that also, if $\Delta d > l_0 = c\tau_0$, $\Delta\tau > \tau_0$, i.e. if the wave is non-coherent, then $I_{int} = 0$, and the visibility of patterns is null ($\mathcal{V} = 0$).

§ 9.3 Coherence and Fourier Calculus

§§ 9.3.1 Line Width and Power Spectrum

When dealing with electromagnetic waves it's important to note that in nature monochromatic sources *do not* exist in general.
Every single source that emits electromagnetic waves at some frequency ω_0, has *always* some spread around the emission frequency called the *line width* $\Delta\omega$, given from dispersion.
This line width is strictly tied to the coherence time τ_0 of the source.
Consider now a generic wave train $f(t)$ with coherence time τ_0. Its time dependence is generally described by a complex exponential in the following way

$$f(t) = \begin{cases} e^{-i\omega_0 t} & -\frac{\tau_0}{2} < t < \frac{\tau_0}{2} \\ 0 & |t| \geq \frac{\tau_0}{2} \end{cases} \tag{9.28}$$

The study of this single wave train is way simpler in the ω space. Applying the Fourier transformation to our wave train we get

$$\hat{f}(\omega) = \hat{\mathcal{F}}\left[f\right](\omega) = \frac{1}{i(\omega - \omega_0)\sqrt{2\pi}}\left[e^{i(\omega-\omega_0)\frac{\tau_0}{2}} - e^{-i(\omega-\omega_0)\frac{\tau_0}{2}}\right]$$

Noting the sine on the right, we have finally

$$\hat{f}(\omega) = \sqrt{\frac{2}{\pi}} \frac{\sin\left[(\omega - \omega_0)\frac{\tau_0}{2}\right]}{\omega - \omega_0} \qquad |t| < \frac{\tau_0}{2} \tag{9.29}$$

From this, we define the *power spectrum* $\hat{F}(\omega)$ of the wave in ω space as

$$\hat{F}(\omega) = \left|\hat{f}(\omega)\right|^2$$

Which, in for this train wave is

$$\hat{F}(\omega) = \frac{2}{\pi} \frac{\sin^2\left[(\omega - \omega_0)\frac{\tau_0}{2}\right]}{\omega - \omega_0} \tag{9.30}$$

Now, in order to find the line width, we search for the zeroes of the power spectrum, which in this case are

$$\sin\left[(\omega_k - \omega_0)\frac{\tau_0}{2}\right] = 0 \implies \omega_k = \frac{2k\pi}{\tau_0} + \omega_0$$

The distance between two consecutive orders of shadows (ω_k, ω_{k+1}), we have

$$\Delta\omega = \frac{2\pi}{\tau_0} \implies \Delta\nu = \frac{1}{\tau_0} \tag{9.31}$$

I.e., the line width is *exactly* the inverse of the coherence time. Note that a perfectly coherent (ideal) electromagnetic wave must have an infinite coherence time, therefore the line width must be zero. This is obtained only if the power spectrum is a delta distribution around the emission frequency ω_0

$$\hat{F}(\omega) = \delta(\omega - \omega_0)$$

This result, gives us a way to experimentally measure the average coherence time and length of a wave. By definition we have

$$\langle\tau_0\rangle = \frac{1}{\Delta\nu} \qquad \langle l_0\rangle = c\langle\tau_0\rangle = \frac{c}{\Delta\nu} \tag{9.32}$$

Since $\Delta\nu/\nu = \Delta\lambda/\lambda$ we have

$$\Delta\nu = \frac{\Delta\lambda}{c\lambda^2}$$

I.e., if we pass from frequencies to wavelengths, we can estimate the coherence length with the following expression

$$\langle l_0\rangle = \frac{\lambda^2}{\Delta\lambda} \tag{9.33}$$

§§§ 9.3.1.1 Power Spectra and Interference

Given some wave train, how can we find the interference pattern from the power spectrum? This comes easily as a result of the following theorem

Theorem 9.1 (Wiener-Khinchin). *Given an electromagnetic wave with power spectrum $G(\omega) = |g(\omega)|^2$, the autocorrelation function is given by*

$$\Gamma(\tau) = \hat{\mathcal{F}}^{-1}\left[G(\omega)\right](\tau) \tag{9.34}$$

I.e., the autocorrelation function of a wave is the inverse Fourier transform of the power spectrum

Proof. Said $E(t)$ the inverse Fourier transform of $g(\omega)$, we have

$$\Gamma(\tau) = \int_{\mathbb{R}} E(t)\overline{E}(t+\tau)\mathrm{d}t = \frac{1}{2\pi}\int_{\mathbb{R}}\left[\int_{\mathbb{R}} g(\omega)e^{-i\omega t}\mathrm{d}\omega\right]\overline{\left[\int_{\mathbb{R}} g(\omega')e^{-i\omega'(t+\tau)}\mathrm{d}\omega'\right]}\mathrm{d}t$$

Due to the independence of the variables, using Fubini's theorem on the exchange of integrals we can rewrite everything as follows

$$\Gamma(\tau) = \frac{1}{2\pi}\iiint_{\mathbb{R}^3} g(\omega)\overline{g(\omega')}e^{-i(\omega-\omega')t}e^{i\omega'\tau}\mathrm{d}\omega\mathrm{d}\omega'\mathrm{d}t$$

Integrating the first exponential with respect to t it transforms exactly to a delta distribution, and therefore

$$\Gamma(\tau) = \frac{1}{2\pi}\iint_{\mathbb{R}^2} g(\omega)\overline{g(\omega')}\delta(\omega-\omega')e^{i\omega'\tau}\mathrm{d}\omega\mathrm{d}\omega'$$

Applying the delta distribution on the integral in ω' we have then

$$\Gamma(\tau) = \frac{1}{2\pi}\int_{\mathbb{R}} g(\omega)\overline{g(\omega)}e^{i\omega\tau}\mathrm{d}\omega$$

Using $z\overline{z} = |z|^2$ we obtain the power spectrum, and the theorem is proven

$$\Gamma(\tau) = \frac{1}{2\pi}\int_{\mathbb{R}} G(\omega)e^{i\omega\tau}\mathrm{d}\omega = \hat{\mathcal{F}}^{-1}\left[G(\omega)\right](\tau)$$

□

§ 9.4 Multiple Beam Interference

§§ 9.4.1 Laser Cavities

So far we treated only the interference of a single wave with itself and of two different waves (also the same wave split in two) interfering between each other.
Now we will treat the case of multiple beams (or a single beam interfering with itself) interfering with each other. Experimentally it can be shown (with instruments like an etalon or a Fabry-Perot interferometer) that multiple beams interfering with each other *will* show an interference pattern on a given screen.
Consider a material thick d with a refraction index n_2 and a single wave with amplitude E_0 incoming from outside the material at some angle θ, where the outside has refraction index n_1. The result, as we know already, will be a reflection and a refraction.
Said t_1, r_1 the Fresnel coefficient for $n_1 \to n_2$ and t_2, r_2 the Fresnel coefficients $n_2 \to n_2$ we have that, at the first reflection we will get a reflected beam $E_{r,1}$ with amplitude $r_1 E_0$ and transmitted wave with

amplitude $E_1 = t_1 E_0$. The transmitted wave will reach again the boundary of the material (on the other side) and the process will be repeated with a wave with amplitude E_1.
At the end we will have a series of transmissions and reflections, with $n-$th term

$$\begin{aligned} E_{T,n} &= t_1 t_2 r_2^{2n} E_0 \\ E_{R,n} &= t_1 t_2 r_2^{2n+1} E_0 \end{aligned} \tag{9.35}$$

On each reflection, there will be a phase shift equal to the wavenumber times the optical path traveled. For one process (*two* reflections) we have that, if the reflection angle *inside* the medium is θ, the phase shift will be

$$\delta = 2kd\cos\theta = \frac{4n_2\pi}{\lambda_0} d\cos\theta \tag{9.36}$$

Where λ_0 is the vacuum wavelength. Considering *every* reflection ($n \to \infty$) together with the phase shift we have that the final amplitude of the series of reflected waves from the material

$$E_R = r_1 E_0 + t_1 t_2 r_2 E_0 e^{i\delta} \sum_{j=0}^{\infty} r_2^{2n} e^{in\delta} \tag{9.37}$$

The infinite sum on the right is a convergent geometric sum, which gives the following final result for the amplitude of the last wave of the reflected series

$$E_R = E_0 \left[r_1 + \frac{t_1 t_2 r_2 e^{i\delta}}{1 - r_2^2 e^{i\delta}} \right]$$

It's provable that $r_1 = -r_2$ and $t_1 t_2 = 1 - r_1^2$, then, rearranging the term on the right we have

$$E_R = E_0 \left[r_1 - \frac{t_1 t_2 r_1 e^{i\delta}}{1 - r_1^2 e^{i\delta}} \right] = E_0 \left[\frac{r_1 \left(1 - r_1^2 e^{i\delta}\right) - t_1 t_2 r_1 e^{i\delta}}{1 - r_1^2 e^{i\delta}} \right] = E_0 r_1 \left[\frac{1 - (r_1^2 + t_1 t_2) e^{i\delta}}{1 - r_1^2 e^{i\delta}} \right]$$

Using $t_1 t_2 = T_1$ and $r_1^2 = R_1$ and that $r_1^2 + t_1 t_2 = 1$ we can write that the final amplitude of the reflected series of waves is

$$E_R = E_0 r_1 \frac{1 - e^{i\delta}}{1 - R e^{i\delta}} \tag{9.38}$$

Or, using $R = |E_R/E_0|^2$, we could also write that the reflection coefficient of the slab is

$$R = \left| \frac{E_R}{E_0} \right|^2 = R_1 \frac{\left|1 - e^{i\delta}\right|^2}{\left|1 - R e^{i\delta}\right|}^2$$

Which, using $|z|^2 = z\overline{z}$ as usual for complex numbers, becomes

$$R = R_1 \frac{(1 - e^{i\delta})(1 - e^{-i\delta})}{(1 - R e^{i\delta})(1 - R e^{-i\delta})} = 2R \frac{1 - \cos\delta}{1 + R^2 - 2R\cos\delta}$$

Further simplifying, using

$$1 - \cos\delta = 2\sin^2\left(\frac{\delta}{2}\right)$$

We have

$$R = \frac{4R\sin\left(\frac{\delta}{2}\right)}{1+R^2-2R-4R\sin^2\left(\frac{\delta}{2}\right)} = \frac{4R}{(1-R)^2}\frac{\sin^2\left(\frac{\delta}{2}\right)}{1+\frac{4R}{(1-R)^2}\sin^2\left(\frac{\delta}{2}\right)}$$

Which, defined the *Finesse coefficient* as

$$F = \frac{4R}{(1-R)^2} \tag{9.39}$$

Becomes

$$R = \frac{F\sin^2\left(\frac{\delta}{2}\right)}{1+F\sin^2\left(\frac{\delta}{2}\right)} \tag{9.40}$$

The function on the right (divided by F) is known as the *Airy function*.
It's possible to find easily also the transmission coefficient of the slab T, using

$$T = 1-R = \frac{1}{1+F\sin^2\left(\frac{\delta}{2}\right)}, \qquad F = \frac{4(1-T)}{T^4} \tag{9.41}$$

The peaks and the shadows of the image on the screen can be calculated by optimization calculus, using the function

$$I_R(\delta) = I_0 F\frac{1}{1+F\sin^2\left(\frac{\delta}{2}\right)} = RI_0$$

Deriving with respect to delta then

$$\frac{\partial I_R}{\partial\delta} = I_0 F\frac{\sin\left(\frac{\delta}{2}\right)\cos\left(\frac{\delta}{2}\right)}{\left(1+F\sin^2\left(\frac{\delta}{2}\right)^2\right)} = 0$$

Which implies that the maxima and minima of the irradiance are at the following values of δ

$$\begin{cases}\delta_m = 2m\pi & max\left\{R\right\}\\ \delta'_m = (2m+1)\pi & \max\left\{T\right\}\end{cases} \tag{9.42}$$

Remembering that δ is a function of the wavelength and optical path, we have

$$\delta(k) = 2kd\cos(\theta), \quad \delta(\lambda) = \frac{4\pi d}{\lambda}\cos\theta, \quad \delta(\omega) = \frac{2\omega n}{c}d\cos\theta$$

Noting the dependence on the frequency of δ, it's clear that the slab is useful for distinguishing between waves with different frequencies, as different frequencies will have different peaks, precisely, noting that at different *peak orders*, defined by the whole number $m \in \mathbb{Z}$

$$\omega_m = \frac{m\pi c}{dn}$$

And that, developing the transmission irradiance around these peaks, i.e. at $\omega - \omega_m$, we have

$$I_T = I_0T^2\frac{1}{T^2+4(1-T)\sin^2\left(\frac{dm}{\pi c}(\omega-\omega_m)\right)} \approxeq I_0\frac{\frac{\sigma^2}{4}}{\frac{\sigma^2}{4}+(\omega-\omega_m)^2} \tag{9.43}$$

Where σ is a constant which depends only on the properties of the material

$$\frac{\sigma^2}{4} = \frac{4T^2c^2}{d^2n^2}\frac{1}{4(1-T)}$$

This is *exactly* as if we studied the behavior of a cavity. Experimentally it can be interpreted as a *laser cavity*.
Summing for each transmission, and using the cavity limits $0 \leq m \leq N-1$ noting that what we found is exactly the power spectrum of the cavity, from the Wiener-Khinchin theorem that the self correlation of the beam is simply the Fourier transform of the power spectrum, Therefore

$$\gamma(\tau) = c\sum_{m=0}^{N-1}\hat{\mathcal{F}}\left[\frac{\frac{\sigma^2}{4}}{\frac{\sigma^2}{4}+(\omega-\omega_m)^2}\right](\tau) = c\sum_{m=0}^{N-1}e^{\frac{|\sigma|\tau}{2}}e^{i\omega_m\tau}$$

Which, after summation, gives

$$\gamma(\tau) = ce^{\frac{|\sigma|\tau}{2}}\frac{1-e^{iN\omega_m\tau}}{1-e^{i\omega_m\tau}} \quad c\in\mathbb{C} \tag{9.44}$$

The visibility of fringes is simply $\mathcal{V} = |\gamma| = \sqrt{\gamma\bar{\gamma}}$. Evaluating the parenthesis multiplications and writing the correct trigonometric functions we have (inglobating a $\sqrt{2}$ in a constant k

$$\mathcal{V}(\tau) = ke^{\frac{|\sigma|\tau}{2}}\sqrt{\frac{1-\cos(N\omega_m\tau)}{1-\cos(\omega_m\tau)}} = ke^{\frac{|\sigma|\tau}{2}}\left|\frac{\sin\left(\frac{N\omega_m\tau}{2}\right)}{\sin\left(\frac{\omega_m\tau}{2}\right)}\right| \tag{9.45}$$

Normalizing everything, and using $|\gamma(0)| = 1$, we have $k = N^{-1}$, which gives finally

$$\mathcal{V}(\tau) = \frac{e^{\frac{|\sigma|\tau}{2}}}{N}\left|\sin\left(\frac{N\omega_m\tau}{2}\right)\csc\left(\frac{\omega_m\tau}{2}\right)\right| \tag{9.46}$$

§§ 9.4.2 Fabry-Perot Instruments

Fabry-Perot interferometers utilize the results obtained from the study of multi-beam interference with broad sources of light. They're usually used to determine the frequency of waves.
Their construction is similar to a laser cavity, where the slab gets substituted by two semi-transparent mirrors which replicate the cavity. In this case, since there is air inside these mirrors, we have

$$r_1 = r_2 = r, \quad t_1 = t_2 = t$$

As before, the path difference between the n-th and the $n+1$-th reflection is

$$d = 2nd\cos\theta$$

Therefore, the phase difference in a single double reflection is

$$\delta = \frac{4\pi d}{\lambda}\cos\theta$$

Hence, the final outgoing reflected and trasmitted amplitude are

$$\begin{cases} E_R = rE_0 + t^2E_0 \sum_{n=0}^{\infty} r^{2n+1}e^{in\delta} \\ E_T = t^2E_0 \sum_{n=0}^{\infty} r^{2n}e^{in\delta} \end{cases} \tag{9.47}$$

Noting that $r^2 < 1$ and that $r^2 = R,\ t^2 = T$, we have

$$\begin{aligned} E_T &= \frac{t^2E_0}{1 - Re^{i\delta}} \\ I_T &= \frac{T^2I_0}{|1 - Re^{i\delta}|^2} \end{aligned} \tag{9.48}$$

As with the laser cavity we have that $\left|1 - Re^{i\delta}\right|^2 = (1-R)^2(1 + F\sin^2(\delta/2))$, with F being the finesse of the instrument, defined in (9.39).
Considered also the absorbment of some of the irradiance, i.e. noting that $R + T + A = 1$, with A being the absorption coefficient, and considering also that $r \in \mathbb{C}$ will bring a phase shift itself, precisely, for each reflection some value $\delta_r/2 \in [0, 2\pi)$ we get that, writing $\Delta = \delta + \delta_r$ that firstly

$$\frac{T^2}{(1-R)^2} = \left(\frac{1 - R - A}{1 - R}\right)^2$$

And therefore, for our realistic interferometer

$$I_T = I_0 \left(1 - \frac{A}{1-R}\right)^2 \frac{1}{1 + F\sin^2\left(\frac{\Delta}{2}\right)} \tag{9.49}$$

As usual, we find the maxima of the Airy function, which correspond to

$$\Delta = 2N\pi = \frac{4\pi d}{\lambda}\cos\theta - \delta_r \tag{9.50}$$

The integer $N \in \mathbb{N}$ is known as the *order of interference* of the beams, which indicates the optical distance difference of two beams with different λ.
At this maxima we have

$$I_{max} = \frac{T^2I_0}{(1-R)^2} = I_0\left(1 - \frac{A}{1-R}\right)^2 \tag{9.51}$$

These instruments are used to measure the wavelength of a source with high precision. A real Fabry-Perot instrument takes the light coming from a broad source of light, collimates it towards the two mirrors described before and then with another lense collimates the outgoing rays to a single point in the measuring screen. There are two configurations for a Fabry-Perot instrument, one is the *etalon* in which the mirrors are fixed in position, and another, known as the interferometer, where the mirrors can be moved in order to change the phase difference δ obtained by the multiple reflections.

§§§ 9.4.2.1 Resolution Power of Fabry-Perot Instruments

In order to actually measure the wavelenghts (or frequencies) of the broad source in question, we gotta understand what it means to *resolve* two lines in the interference pattern obtained.
For a simpler evaluation, consider $A = \delta_r = 0$, and call Γ the line width at half height at the peak. Consider two lines, one at order m and one at order $m+1$.
In order to call the two peaks resolved, we employ *Taylor's criterion*, which states that two lines are resolved if they, at maximum, intersecate at the half-height point, where the irradiance is $I_0/2$. Therefore if $I = I_0/2$ at the half height point, we have then, using (9.51)

$$\delta = 2\pi m + \frac{\Gamma}{2} = \Delta$$

Where we moved from the peak of exactly half line width, bringing ourselves at the intersection point of the two lines.
Applying the aforementioned criterion, we get then

$$\frac{I_0}{2} = \frac{T^2 I_0}{1 + F\sin^2\left(m\pi + \frac{\Gamma}{4}\right)} \implies \frac{1}{2} = \left(1 - \frac{A}{1-R}\right)^2 \frac{1}{1 + F\sin^2\left(m\pi + \frac{\Gamma}{4}\right)}$$

Imposing $A = 0$, and noting that $m\pi + \Gamma/4 << 1$ we have, after some algebraic juggling

$$1 = F\sin^2\left(m\pi + \frac{\Gamma}{4}\right) \approx \frac{F\Gamma^2}{16}$$

Solving for gamma we get that a line will be resolved (for Taylor), *if and only if* its line width is exactly equal to the following

$$\Gamma = \frac{4}{\sqrt{F}} \tag{9.52}$$

I.e., if the maximas are at a distance $d_m > \Gamma$. It's clear that, due to the definition of finesse, the minimum resolution distance is tied only to the instrument and not to the properties of the wave

§§§ 9.4.2.2 Spectral Resolution

The "goodness" of a measure with a Fabry-Perot instrument is evaluated with what is known as *Resolving Power*, RP. This value is a pure number defined by the ratio of the measured wavelength (or frequency) of the measured wave with the smallest resolvable interval of wavelength (frequency) as

$$RP = \frac{\lambda}{\Delta\lambda_{min}} = \frac{\nu}{\Delta\nu_{min}} = \frac{\omega}{\Delta\omega_{min}} \tag{9.53}$$

It's clear that if the instrument can measure a smaller interval, then the resolution power will be greater. We can tie this value with what we wrote before noting that, close to the maximum of the line

$$\delta = \frac{4\pi d}{\lambda}\cos\theta = 2\pi m \implies \frac{2d}{\lambda}\cos\theta = m \qquad m \in \mathbb{Z}$$

Or, in terms of only wavelengths

$$m\lambda = 2d\cos\theta$$

Differentiating, we can also say that

$$\begin{cases} m\Delta\lambda = 2d\sin\theta\Delta\theta \\ \Delta\delta = \dfrac{4\pi d}{\lambda}\sin\theta\Delta\theta \end{cases}$$

Or, solving

$$2\pi\Delta\delta = 2\pi m\frac{\Delta\lambda}{\lambda} = \frac{4}{\sqrt{F}}$$

Therefore, we have then

$$\frac{\Delta\lambda}{\lambda} = \frac{\pi}{2m}\frac{1}{\sqrt{F}} \tag{9.54}$$

Which, inserted into the formula for the RP, we have that

$$RP = \frac{\pi}{2}m\sqrt{F} \tag{9.55}$$

Therefore, the resolving power is also linearly dependent to the order of interference.
Another important part of spectral analysis of waves using Fabry-Perot instruments is the distance between two maxima. This is commonly known as the *Free Spectral Range* of the instrument.
By definition, we have $\delta_m = 2\pi m$

$$\begin{aligned} \delta_{m+1} - \delta_m &= 2\pi \\ 4\pi d\left(\frac{1}{\lambda_{m+1}} - \frac{1}{\lambda_m}\right)\cos\theta &= 2\pi \end{aligned} \tag{9.56}$$

The free spectral range, or FSR is then defined

$$FSR = \frac{1}{\frac{1}{\lambda_{m+1}-\lambda_m}} = \frac{\lambda^2}{2d\cos\theta}$$

Since we're talking about maxima, we have that $m\lambda = 2d\cos\theta$, which inserted into the previous equation gives

$$FSR = \frac{\lambda}{m} \tag{9.57}$$

Therefore, the FSR gets smaller with higher orders Defining also the *reflecting finesse* as

$$\mathcal{F} = \frac{\pi}{2}\sqrt{F} \tag{9.58}$$

We can also redefine the RP as

$$RP = m\mathcal{F} = m\pi\frac{\sqrt{R}}{1-R} \tag{9.59}$$

Therefore tying closely the resolving power of the instrument to its physical properties.
In general, then, given $m \in \mathbb{Z}$ the interference order of the wave in study, we can summarize what we wrote in maths as follows

- The resolving power (RP) is tied to the physical properties of the instrument and grows with m

- The distance between peaks (FSR) gets smaller with greater m

It's clear that even if the order of interference grows and with it the resolving power, the distance between peaks will reduce, and there will be a point where the peaks won't be resolved anymore. This fact sigillates etalons to a single specific group of measures, while interferometers can be used with a wider range of tests.

10 Diffraction

§ 10.1 Fresnel-Kirchhoff Theory

§§ 10.1.1 Huygens Principle

The general idea behind the theory of diffraction comes from a simple fact. Given a sharp object, the shadow casted from it is not sharp as predicted from geometric optics.
The smearing of the boundary of the shadow comes from a phenomenon known as *diffraction*. It can be explained summarily with Huygens principle

Principle (Huygens)**.** Given a generic wave, its propagation can be described by taking each point of the wavefront and counting it as a source of a spherical wave.
The sum of all the spherical wavelets will define the wave at a later time

Counting all the wavelets when the wave encounters the object, we can see how they envelop the object and propagate around it, giving the smearing effect we see on the shadow.

§§ 10.1.2 Kirchhoff Integral Formula

Huygens' principle can be rewritten mathematically using Green's identities.
Firstly, the chosen electromagnetic wave due to symmetry considerations can be approximated to a scalar function satisfying the following equation

$$\Box_u \psi = \frac{1}{u^2}\frac{\partial^2 \psi}{\partial t^2} - \nabla^2 \psi = 0 \tag{10.1}$$

By definition, $\psi \in C^2$ and it's said to be *harmonic*. Taken a second harmonic function φ, then we have that Green's second identity holds

$$\iiint_V \left(\psi \nabla^2 \varphi - \varphi \nabla^2 \psi\right) \mathrm{d}^3 x = \oiint_{\partial V} \left(\psi \nabla \varphi \cdot \hat{\mathbf{n}} - \varphi \nabla \psi \cdot \hat{\mathbf{n}}\right) \mathrm{d}^2 s \tag{10.2}$$

A corollary of this identity comes in handy

Corollary 10.1.1. Given f, g two harmonic functions, then

$$\oiint_{\partial V} \left(f \nabla g \cdot \hat{\mathbf{n}} - g \nabla f \cdot \hat{\mathbf{n}}\right) \mathrm{d}^2 s = 0 \tag{10.3}$$

It's clear then that the two functions ψ, φ satisfy this corollary, helping us in our evaluations. Now, let's bring ourselves to a real case. Said V a set containing the source of the wave ψ and letting φ be the spherical wavelets indicated by Huygens, described as follows

$$\varphi(r,t) = \frac{\varphi_0}{r} e^{ikr - i\omega t} \tag{10.4}$$

Said without loss of generality that the source of the wave ψ is the point $\{r = 0\} \in V$, i.e. we have

$$\lim_{r \to 0} \psi(r,t) = \pm\infty$$

Or, in other terms, Green's second identity cannot be applied, since the function diverge at the origin. Said $\tilde{V} = V \setminus B_\epsilon(0)$ a new set which excludes a ball of radius $\epsilon > 0$ around the origin, then, we can apply Green's identity and its corollary. Note also that

$$\partial\tilde{V} = \partial V \setminus \partial B_\epsilon(0)$$

Thus

$$\iiint_{\tilde{V}} \psi\nabla^2\varphi - \varphi\nabla^2\psi \mathrm{d}^3x = \oiint_{\partial V} \psi\nabla\varphi\cdot\hat{\mathbf{n}} - \varphi\nabla\psi\cdot\hat{\mathbf{n}}\mathrm{d}^2s - \oiint_{\partial B_\epsilon(0)} \psi\nabla\varphi\cdot\hat{\mathbf{n}} - \varphi\nabla\psi\cdot\hat{\mathbf{n}}\mathrm{d}^2s = 0 \tag{10.5}$$

Using the previous definition to our spherical wavelets, we have by definition on our integral over the boundary of the sphere with radius ϵ

$$\varphi_0 \oiint_{\partial B_\epsilon(0)} (\cdots)\mathrm{d}^2s = \varphi_0 \iint_{4\pi} \left(\psi \left.\frac{\partial}{\partial r}\right|_\epsilon \left(\frac{e^{ikr-i\omega t}}{r}\right) - \frac{e^{ik\epsilon - i\omega t}}{\epsilon} \left.\frac{\partial\psi}{\partial r}\right|_\epsilon \right) \epsilon^2 \mathrm{d}\Omega \tag{10.6}$$

Writing the derivative and evaluating the limit for $\epsilon \to 0$, i.e. accounting for the source point, we get

$$\lim_{\epsilon\to 0} \iint_{4\pi} \left(\epsilon e^{ik\epsilon - i\omega t} \left.\frac{\partial\psi}{\partial r}\right|_\epsilon - \psi(\epsilon,t)\left(ik\epsilon e^{ik\epsilon - i\omega t} - e^{ik\epsilon - i\omega t}\right)\right) \mathrm{d}\Omega = 4\pi\psi(0,t)e^{-i\omega t} \tag{10.7}$$

Which, inserted into the initial integral gives *Kirchhoff's Integral Formula*, which relates the wave at the source with the wave at the boundary.

$$\psi(0,t) = -\frac{e^{-i\omega t}}{4\pi} \oiint_{\partial V} \frac{e^{ikr}}{r} \nabla\psi\cdot\hat{\mathbf{n}} - \hat{\mathbf{n}}\cdot\nabla\left(\frac{e^{ikr}}{r}\right)\psi(r,t)\mathrm{d}^2s \tag{10.8}$$

As usual, $I \propto |\psi|^2$. In literature, the function ψ is known as the "disturbance"

§§ 10.1.3 Fresnel-Kirchhoff Integral

The Kirchhoff integral that we found before, we can describe generally the problem of wave propagation. The application of the same to the problem of diffraction was developed by Fresnel.
Given a generic aperture Σ, a source S distant r' from it, with a measuring point P on the other side of the aperture at a distance r, we take the set V in a way such that its boundary is composed by the diffraction aperture Σ and a semisphere of radius R containing the measuring point P. We assume:

- ψ and $\nabla\psi$ are negligible with respect to the integral on the aperture

- ψ and $\nabla\psi$ are the same with or without the aperture

Then, on the aperture, the optical disturbance is described as follows

$$\psi(r',t) = \frac{\psi_0}{r'}e^{ikr'-i\omega t}$$

Then, said ψ_P the wave at the point P, we have

$$\psi_P = -\frac{\psi_0 e^{-i\omega t}}{4\pi}\oiint_{\partial V}\left(\frac{e^{ikr}}{r}\nabla'\left(\frac{e^{ikr'}}{r'}\right)\cdot\hat{\mathbf{n}} - \frac{e^{ikr'}}{r'}\nabla\left(\frac{e^{ikr}}{r}\right)\cdot\hat{\mathbf{n}}\right)\mathrm{d}^2 s$$

Said S the semispherical part of ∂V, and noting that the integral on it vanishes for $R\to\infty$, and noting that on the aperture

$$\begin{aligned}\hat{\mathbf{n}}\cdot\nabla &= \cos(\hat{\mathbf{n}},\mathbf{r})\frac{\partial}{\partial r}\\ \hat{\mathbf{n}}\cdot\nabla &= \cos(\hat{\mathbf{n}},\mathbf{r}')\frac{\partial}{\partial r'}\end{aligned} \tag{10.9}$$

Where $\cos(.,.)$ is the angle between the two vectors, we have, after applying ∂_r

$$\psi_P = -\frac{\psi_0 e^{-i\omega t}}{4\pi}\iint_\Sigma \frac{e^{ik(r+r')}}{rr'}\left[\left(ik-\frac{1}{r'}\right)\cos(\hat{\mathbf{n}},\mathbf{r}') - \left(ik-\frac{1}{r}\right)\cos(\hat{\mathbf{n}},\mathbf{r})\right]\mathrm{d}^2 s \tag{10.10}$$

In the situation where $r, r' >> \lambda$ we can neglect the second order terms, and we get Fresnel-Kirchhoff's integral

$$\psi_P = -\frac{ik\psi_0 e^{-i\omega t}}{4\pi}\iint_\Sigma \frac{e^{ik(r+r')}}{rr'}\left(\cos(\mathbf{n},\mathbf{r}') - \cos(\mathbf{n},\mathbf{r})\right)\mathrm{d}^2 s \tag{10.11}$$

This integral is the mathematical expression of Huygens' principle.
This can be discerned by taking semicircular aperture with radius r, centered on the source of the wave S. Noting that $\cos(\hat{\mathbf{n}},\mathbf{r}') = -1$ we get

$$\psi_P = \frac{ik}{4\pi}\iint_\Sigma \psi_\Sigma \frac{e^{ikr-i\omega t}}{r}\left(\cos(\hat{\mathbf{n}},\mathbf{r})+1\right)\mathrm{d}^2 s$$

Where

$$\psi_\Sigma(r',t) = \frac{\psi_0 e^{ikr'}}{r'}$$

This can be interpreted as seeing that the aperture generates spherical "Huygens' wavelets" at each point $\mathrm{d}^2 s$, s

$$\psi_H = \frac{1}{r}\psi_\Sigma e^{ikr-i\omega t}$$

Then, the KF integral becomes a summation of all these wavelets times a correction, known as the "obliquity factor" given by the cosine

$$\psi_P(r,t) = \frac{ik}{4\pi}\iint_\Sigma \psi_H\left(\cos(\hat{\mathbf{n}},\mathbf{r})+1\right)\mathrm{d}^2 s \tag{10.12}$$

The imaginary unit that multiplies the integral, if written as a phasor, clearly indicates that there is also a phase shift of the wave after diffraction by $\pi/2$, which wasn't theorized by Huygens together with the obliquity factor.

§§ 10.1.4 Babinet Principle

Consider an aperture A which produces a disturbance ψ_P at some point P. Supposed that the aperture is described by two complementary apertures A_1, A_2, from the formula of the KF integral and the properties of integrals, we have that

$$\psi_P = \psi_1 + \psi_2$$

Where ψ_i is the disturbance created by the i-th aperture. Then, if $\psi_P = 0$, we have that the disturbance created by the two complementary apertures are equal and dephased by π radians, while the irradiance is exactly the same.
This principle is known as *Babinet's principle*, and indicates how we can determine the pattern of an object. Given a spherical aperture, the pattern will be the same as for a spherical particle complementary to the aperture, plus a π dephasing.

§ 10.2 Fraunhofer Diffraction

In general, it's not easy to solve the diffraction integral, thus it's usually approximated in two major categories: Fresnel and Fraunhofer diffraction:

- Fraunhofer diffraction, when the source and/or the measuring point are far away from the source (*Far field approximation*)
- Fresnel diffraction, when the source and/or the measuring point are close to the source (*Close field approximation*

Consider the case of a source distant d' and a measuring point distant d from the aperture plane, respectively vertically displaced by h, h' from the center of the aperture. If the aperture is long δ and the two points are respectively r', r from the center of it, we have that the variation Δ of $r + r'$ is

$$\Delta = \sqrt{d'^2 + (h' + \delta)^2} + \sqrt{d^2 + (h + \delta)^2} - \sqrt{d'^2 + h'^2} - \sqrt{d^2 + h^2} \tag{10.13}$$

Approximating it to the second order, remembering that

$$\sqrt{1 + x^2} \approx 1 + \frac{x^2}{2} + \mathcal{O}(x^3)$$

We have, at the second order

$$\Delta \approx \left(\frac{h'}{d'} + \frac{h}{d}\right)\delta + \frac{1}{2}\left(\frac{1}{d'} + \frac{1}{d}\right)\delta^2 \tag{10.14}$$

The term δ^2 essentially describes the curvature of the wave, and it will be used for distinguishing between Fraunhofer and Fresnel diffraction. As we said before, the regime of Fraunhofer diffraction is obtained when the "field is far away", thus we can neglect the wave curvature and work only with purely plane waves. This is obtained when

$$\frac{1}{2}\left(\frac{1}{d'} + \frac{1}{d}\right)\delta^2 << \lambda$$

This same result can be obtained in the laboratory using a collimating lens and a focusing lens which will illuminate the aperture with a (obviously coherent) wave and focus the pattern on the focal plane. When the plane wave approximation is satisfied, we can easily say that

- The obliquity factor is approximately constant on Σ
- The factor $e^{ikr'}/r'$ is approximately constant on Σ
- The factor $1/r$ is approximately constant on Σ

Thus, taking out all the constants that multiply the integral as C, we have that the FK integral for Fraunhofer diffraction takes a really simple shape, as

$$\psi_P = C \iint_\Sigma e^{ikr} \mathrm{d}^2 s \tag{10.15}$$

§§ 10.2.1 Single Slit Diffraction

Consider a narrow 1D slit wide b, then if we're in the Fraunhofer regime, with the source distant r_0 and the focal point placed at some inclination θ with respect of the wave, we have that, if y is the distance from the center of the slit

$$\begin{aligned} r &= y \sin\theta + r_0 \\ \mathrm{d}^2 s &= L \mathrm{d}y \end{aligned}$$

Then

$$\psi_P = C \int_{-\frac{b}{2}}^{\frac{b}{2}} e^{iky\sin\theta + ikr_0} L \mathrm{d}y \tag{10.16}$$

Bringing outside Le^{ikr_0} and incorporating it into the constant C, the integral can be easily calculated, giving

$$\psi_P \frac{LCe^{ikr_0}}{ik\sin\theta}\left(e^{\frac{1}{2}ikb\sin\theta} - e^{-\frac{1}{2}ikb\sin\theta}\right)$$

Expressing the exponentials as a sine, we have

$$\psi_P = \frac{2LCe^{ikr_0}}{k} \frac{sin\left(\frac{1}{2}kb\sin\theta\right)}{\sin\theta} = \frac{2bLCe^{ikr_0}}{k} \operatorname{sinc}\left(\frac{1}{2}kb\sin\theta\right) = C' \operatorname{sinc}\beta \tag{10.17}$$

Where

$$\begin{aligned} \beta &= \frac{1}{2}kb\sin\theta \\ C' &= CbLe^{ikr_0} \end{aligned} \tag{10.18}$$

Said $I_0 = |C'|^2 = |CbL|^2$, the irradiance of the single slit pattern is

$$I = I_0 \operatorname{sinc}^2\beta \tag{10.19}$$

By the definition $I = I(\theta)$ and $I_0 = I(0)$ is the maximum of our irradiance. The zeroes of the irradiance function happen instead when

$$\operatorname{sinc}^2(\beta) = 0 \implies \beta = \pm m\pi \qquad m \geq 1$$

Or, expanding β into its definition

$$I(\theta) = 0 \implies \sin\theta = \frac{2m\pi}{kb} = m\frac{\lambda}{b} \tag{10.20}$$

I.e. for a given wavelenght λ of light the width of the pattern changes inversely to the slit width b. Note also how $I_0 \propto b^2$, and how the pattern we found is exactly the same pattern that we'd get from the interferometric evaluation

§§ 10.2.2 Rectangular Aperture

For the rectangular aperture the evaluation of the integral is pretty much analogous. Said b the width of the rectangle in the y direction and a the width in the x direction, with the substitution

$$r = r_0 + x\sin\varphi + y\sin\theta$$

Our integral becomes

$$\psi_P = C\int_{-\frac{b}{2}}^{\frac{b}{2}}\int_{-\frac{a}{2}}^{\frac{a}{2}} e^{ikr_0 + iky\sin\varphi + ikx\sin\theta} ab\mathrm{d}x\mathrm{d}y \tag{10.21}$$

Using Fubini's theorem and bringing out the constants, this integral is simply the product of two single slits

$$\psi_P = Cabe^{ikr_0}\int_{-\frac{b}{2}}^{\frac{b}{2}} e^{iky\sin\theta}\mathrm{d}y\int_{-\frac{a}{2}}^{\frac{a}{2}} e^{ikx\sin\varphi}\mathrm{d}x \tag{10.22}$$

Which gives

$$\psi_P = abCe^{ikr_0}\left(\frac{2\sin\left(\frac{1}{2}kb\sin\theta\right)}{k\sin\theta}\right)\left(\frac{2\sin\left(\frac{1}{2}ka\sin\varphi\right)}{k\sin\varphi}\right) = Cab\,\mathrm{sinc}\,\alpha\,\mathrm{sinc}\,\beta \tag{10.23}$$

Where

$$\alpha = \frac{1}{2}ka\sin\phi$$
$$\beta = \frac{1}{2}kb\sin\theta$$

Said $I_0 = I(0,0) = |Cab|$, the irradiance is

$$I(\theta\varphi) = I_0\,\mathrm{sinc}^2\left(\frac{1}{2}ka\sin\varphi\right)\mathrm{sinc}^2\left(\frac{1}{2}kb\sin\theta\right) \tag{10.24}$$

The diffraction pattern will be as one of two slits going on the x and y axes of the diffraction plane with a square maxima around $\theta = \varphi = 0$, Thus

$$\begin{aligned}\alpha_z = \pm n\pi &\implies \sin\varphi = n\frac{\lambda}{a}\\ \beta_z = \pm m\pi &\implies \sin\theta = m\frac{\lambda}{b}\end{aligned} \tag{10.25}$$

As for the single slit

§§ 10.2.3 Circular Aperture

For a circular aperture with radius R the Fraunhofer integral is all but obvious. Said as usual

$$r = r_0 + y\sin\theta$$
$$\mathrm{d}^2 s = 2\sqrt{R^2 - y^2}\,\mathrm{d}y$$

The integral is

$$\psi_P = 2Ce^{ikr_0}\int_{-R}^{R} e^{ik\sin\theta}\sqrt{R^2-y^2}\mathrm{d}y \tag{10.26}$$

With the substitution

$$u = \frac{y}{R}$$
$$\rho_k = kR\sin\theta$$

The integral becomes a special integral solved by a Bessel function of the first order

$$\psi_P(\rho) = Ce^{ikr_0}\int_{-1}^{1} e^{i\rho u}\sqrt{1-u^2}\mathrm{d}u = 2C\pi R\frac{J_1(\rho)}{\rho}$$

Said $I_0 = I(0) = |C\pi R|^2$, the irradiance describes an Airy disk from $\rho = 0$ till the first zero, then concentric circles corresponding to the higher orders.
From the irradiance function $I(\theta)$ defined as

$$I(\theta) = I_0\frac{4J_1^2\left(kR\sin\theta\right)}{k^2R^2\sin^2\theta} \tag{10.27}$$

We get that the first zero corresponds to

$$J_1(\rho_0) = 0 \implies \rho_0 \approx 3.832$$

Thus, expanding ρ and said $D = 2R$

$$\sin\theta = 1.22\frac{\lambda}{D} \tag{10.28}$$

This is the dimension of the first peak, and it's therefore also what we'd use to determine if two diffraction patterns are resolved or not by an instrument with aperture diameter D. This value is commonly known as the *diffraction limit* of the instrument at the given wavelength. Note that it's bigger than the diffraction limit of the single slit. Said $\Delta\theta$ the distance between the peaks of the two patterns, we define the *Rayleigh criterion for resolution* as

$$\Delta\theta \geq D_L \tag{10.29}$$

As for our definition of diffraction limit D_L, it's clear then that

$$D_L = \begin{cases} \dfrac{\lambda}{b} & \text{Single Slit, Rectangular Aperture} \\ 1.22\dfrac{\lambda}{2R} & \text{Circular Aperture} \end{cases} \tag{10.30}$$

§§ 10.2.4 Multiple Slit Diffraction

§§§ 10.2.4.1 Double Slit Diffraction

Consider now two equal parallel slits long b, separated by a distance h. As for the problem of the single slit, it can be evaluated in one single dimension. In order to evaluate the Fraunhofer integral, we employ the same substitution we did for the single slit, noting tho that since the slits that we have to integrate over are two, and the integration set is

$$\Sigma = [0, b] \cup [h, h + b]$$

Therefore, we have

$$\psi_P = Ce^{ikr_0} \int_0^b e^{iky \sin\theta} b \mathrm{d}y + Ce^{ikr_0} \int_h^{h+b} e^{iky \sin\theta} b \mathrm{d}y \tag{10.31}$$

Solving the integral we have

$$\psi_P = \frac{Cbe^{ikr_0}}{ik \sin\theta} \left(e^{ikb \sin\theta} - 1\right) \left(1 + e^{ikh \sin\theta}\right)$$

Writing

$$e^{ikh \sin\theta} + 1 = 2e^{\frac{1}{2}kh \sin\theta} \cos\left(\frac{1}{2}kh \sin\theta\right), \qquad e^{ikb \sin\theta} - 1 = 2ie^{\frac{1}{2}kb \sin\theta} \sin\left(\frac{1}{2}kb \sin\theta\right)$$

We have that the solution can be described as follows

$$\psi_P(\theta) = 2bCe^{ikr_0 + ik\left(\frac{h}{2} + \frac{b}{2}\right) \sin\theta} \operatorname{sinc}\left(\frac{1}{2}kb \sin\theta\right) \cos\left(\frac{1}{2}kh \sin\theta\right) \tag{10.32}$$

Or, in term of irradiance

$$I(\theta) = I_0 \operatorname{sinc}^2\left(\frac{1}{2}kb \sin\theta\right) \cos^2\left(\frac{1}{2}kh \sin\theta\right) \tag{10.33}$$

This pattern is clearly a modulation of the single slit pattern. The maxima will be for

$$\cos^2\left(\frac{1}{2}kh \sin\theta\right) = 0 \implies \sin\theta = \frac{2m\pi}{kh} = m\frac{\lambda}{h} \tag{10.34}$$

§§ 10.2.5 Diffraction Gratings

The previous idea can be developed further by creating an aperture made by N equal slits, each long b and distant h between each other. The diffraction integral will then be a finite sum of single slit integrals, as follows

$$\psi_P = Cbe^{ikr_0} \sum_{n=0}^{N} \int_{nh}^{nh+b} e^{iky \sin\theta} L \mathrm{d}y \tag{10.35}$$

The integral inside is easily solvable, giving

$$\psi_P(\theta) = \frac{Cbe^{ikr_0}}{ik\sin\theta}\left(e^{ikb\sin\theta}\sum_{n=0}^{N}e^{iknh\sin\theta} - \sum_{n=0}^{N}e^{iknh\sin\theta}\right)$$

Or, rewriting the right hand side and explicitly summing, we have

$$\psi_P(\theta) = \frac{Cbe^{ikr_0}}{ik\sin\theta}\left(e^{ikb\sin\theta}-1\right)\frac{1-e^{ikNh\sin\theta}}{1-e^{ikh\sin\theta}}$$

Rewriting everything in terms of sines and cosines we have

$$\psi_P(\theta) = 2Cbe^{ikr_0}\operatorname{sinc}\left(\frac{1}{2}kb\sin\theta\right)\frac{\sin\left(\frac{1}{2}Nkh\sin\theta\right)}{\sin\left(\frac{1}{2}kh\sin\theta\right)} \tag{10.36}$$

And in terms of irradiance

$$I(\theta) = I_0\operatorname{sinc}^2\left(\frac{1}{2}kb\sin\theta\right)\frac{\sin^2\left(\frac{1}{2}Nkh\sin\theta\right)}{N^2\sin^2\left(\frac{1}{2}kh\sin\theta\right)} \tag{10.37}$$

As before, we have a modulated single-slit pattern, where we normalized the result dividing by N^2

§§§ 10.2.5.1 Resolving Power of a Diffraction Grating

As we have seen before, the diffraction pattern of a diffraction grating is given by the irradiance function (10.37). The maximas are determined by the last factor, and the primary maxima are found for

$$\frac{1}{2}kh\sin\theta = n\pi \implies \sin\theta = n\frac{\lambda}{h} \tag{10.38}$$

Secondary maximas occur instead for

$$\frac{1}{2}Nkh\sin\theta = (2n+1)\,\pi \implies \sin\theta = \frac{2n+1}{2N}\frac{\lambda}{h} \tag{10.39}$$

While minima occur for

$$\frac{1}{2}Nkh\sin\theta = n\pi \implies \sin\theta = n\frac{\lambda}{Nh} \tag{10.40}$$

The angular distance between the peak and the minimum can be found via differentiation, noting that the argument of the sine at the numerator must be equal to π, therefore

$$\Delta\left(\frac{1}{2}Nkh\sin\theta\right) = \frac{1}{2}Nkh\cos\theta\Delta\theta = \pi$$

This implies that

$$\Delta\theta = \frac{\lambda}{Nh}\sec\theta \tag{10.41}$$

Supposing that we have a big enough number of slits N that we can approximate $\Delta\theta < \epsilon$, using the equation for primary maxima that we found before and differentiating with respect to λ we have that, given the minimal difference of two wavelenghts $\Delta\lambda$, their angular separation will be the following

$$\Delta\theta = n\frac{\Delta\lambda}{h}\sec\theta \tag{10.42}$$

Since we already found the minimal angular separation between a peak and a minima, i.e. our diffraction limit for the grating, equating we have

$$\frac{n}{h}\Delta\lambda\sec\theta = \frac{\lambda}{Nh}\sec\theta \implies \frac{\lambda}{\Delta\lambda} = \frac{nNh}{h}\frac{\sec\theta}{\sec\theta}$$

By definition of resolving power RP, then

$$RP = \frac{\lambda}{\Delta\lambda} = nN \tag{10.43}$$

This simple solution, clearly shows the power of using diffraction gratings. Their resolving power is directly proportional to the fringe order n and to the number of slits in the grating N.

§§§ 10.2.5.2 Types of Gratings

There are two major categories of gratings

- Transmission gratings (transparent)
- Reflection gratings (metallic)

They're both created by incising grooves on the chosen material. A typical grating usually has a groove density of $600\text{grooves}/\text{mm}$ over 10cm of length. Thus, the theoretical RP of this grating is $RP_T \simeq 60000n$. Practically, due to absorption and other non conservative effects the experimental RP is around 90% the theoretical RP. The shape of the grooves is also important, e.g. if the grooves are sawtooth shaped, it's possible to make light appear at only one order n, increasing the efficiency of the grating. The essential requirement is to have grooves distanced by a fraction of wavelenght. Cheaper replicas can be built by plastic molding.
Reflection gratings are usually made plane or concave, where concave reflection gratings make sure that light is precisely collimated.

§ 10.3 Fresnel Diffraction

Fresnel's approximation of the Kirchhoff-Fresnel (KF) integral simply evaluates the maximum distance difference of r, r' to the second order, thus evaluating wave curvature. For this reason this approximation is known as the *close field approximation*. Thus, at the aperture we have

$$\Delta \approx \left(\frac{h'}{d'} + \frac{h}{d}\right)\delta + \frac{1}{2}\left(\frac{1}{d'} + \frac{1}{d}\right)\delta^2 \tag{10.44}$$

Due to the close field nature of Fresnel diffraction, it's easily observable in laboratories.

§§ 10.3.1 Fresnel Zones

Consider a plane aperture illuminated by a point source S, and suppose that the surface is perpendicular to the line connecting the source to the measuring point P. Consider a third point Q on the surface,

distant R from the point where the line between S and P intersects the surface.
Said O the intersection point, we define

$$|SO| = h'$$
$$|OP| = h$$
$$|SQ| = r$$
$$|QP| = r'$$

Thus, as before

$$r + r' = \sqrt{h^2 + R^2} + \sqrt{h'^2 + R^2} \approx h + h' + \frac{1}{2}\left(\frac{1}{h} + \frac{1}{h'}\right) R^2$$

Suppose now that we draw a sequence of values of R and therefore multiple points Q such that the difference of $r + r'$ that we'll indicate with Δ, between a value and its successive is exactly

$$\Delta_{n,n+1} = \frac{1}{2}\lambda$$

By definition then

$$\Delta_{n,n+1} = \frac{1}{2}\left(\frac{1}{h} + \frac{1}{h'}\right)\left(R_{n+1}^2 - R_n^2\right) = \frac{1}{2}\lambda$$

Thus, evaluating everything, we have

$$R_{n+1}^2 - R_n^2 = \lambda\left(\frac{1}{h} + \frac{1}{h'}\right)^{-1} = \lambda f \tag{10.45}$$

Where f is the aperture's "focal length". Noting that $R_0 = 0$, we get $R_1 = \sqrt{\lambda f}$, and via induction, we get

$$R_n = \sqrt{n\lambda f} \tag{10.46}$$

These radii define what are known as *Fresnel zones*. Note that also, the area of these zones is constant. Said A_n the area of the $n-$th zone we have

$$A_n = \pi\left(R_{n+1}^2 - R_n^2\right) = \pi\lambda f = \pi R_1^2 \tag{10.47}$$

These zones are also quite small, in the optical range, i.e. $\lambda \approx 600$ nm and $h = h' \approx 60$ cm we get $R_1 \approx 4$ mm, note also that $R_n \propto \sqrt{n}$, therefore this radius grows relatively slowly. Just imagine that $R_{100} \approx 4$ cm in this case.
Noting that we can consider the total disturbance on the measuring point as the sum of the disturbances of the single zone, noting that for Babinet's principle there is a phase inversion of π, then, if Σ contains N zones we have

$$|\psi| = \sum_{n=1}^{N}(-1)^{n+1}|\psi_n| \tag{10.48}$$

Therefore, if the zones contained are exactly N, we have

$$|\psi_p| = \begin{cases} 0 & N \mod 2 = 0 \\ \sim |\psi_1| & N \mod 2 = 1 \end{cases}$$

Considering also the obliquity factor in the KF integral we also must have

$$|\psi_n| \leq |\psi_{n+1}|$$

Therefore, if we consider the case of no aperture ($\Sigma = \mathbb{R}^2$), the sum can be considered as an infinite sum

$$|\psi_p| = \sum_{n=1}^{\infty}(-1)^{n+1}|\psi_n| = \frac{1}{2}|\psi_1| + \left(\frac{1}{2}|\psi_1| - |\psi_2| - \frac{1}{2}|\psi_3|\right) + \left(\frac{1}{2}|\psi_3| - |\psi_4| - \frac{1}{2}|\psi_5|\right) + \cdots \quad (10.49)$$

Therefore, considering that also $|\psi_n| \approx |\psi_{n+1}|$ for big values of n, the contributes inside the parentheses cancel out, therefore

$$|\psi_p| = \frac{1}{2}|\psi_1|$$

This indicates how a bright spot can be seen in the center. This is also true in general due to the relation of absolute values of adjacent zones, proving a conundrum posed by experimental evaluation, which have shown the existence of a bright central spot which wasn't explained by Fraunhofer diffraction. This spot is known as *Arago's spot*.
Another consideration can be made by evaluating the positioning of an object in front of the object. Using Babinet's principle again, we can evaluate the behavior of the shadow when the diffracting object is either offset or centered.
When it's offset from the center, ψ_p hardly changes, and higher contributions go quickly to zero, while if the object is centered, terms diminish at both ends and inside the shadow zone we get $I \approx 0$, which it's what we usually expect from a shadow.

§§§ 10.3.1.1 Zone Plates

This behavior can be "harvested" by using what's known as *zone plates*, Physical objects which block specific contributions from Fresnel zones. As an example we can build a Fresnel plate which blocks only even zones, giving therefore

$$|\psi_p| = \sum_{n=1}^{N} |\psi_{2n+1}| \quad (10.50)$$

Which makes the plate act as a lens. This lens, if we evaluate its focal length f as

$$f = \frac{R_1^2}{\lambda}$$

We see that it behaves as a *very* chromatic lens.

§§ 10.3.2 Rectangular Aperture

For evaluating the problem of diffraction from a rectangular aperture in close field regimes, we have in the coordinates x, y of the rectangular aperture we have

$$R^2 = x^2 + y^2$$

Therefore

$$r + r' = h + h' + \frac{1}{2f}\left(x^2 + y^2\right)$$

The approximations applied to the KF integral are the following:

1. The obliquity is constant and approximately 1 on the aperture
2. $(rr')^{-1}$ is approximately constant on the aperture

Thus, the integral becomes

$$\psi_P = C \iint_\Sigma e^{\frac{ik}{2f}(x^2+y^2)} \mathrm{d}^2 s = C \int_{x_1}^{x_2} e^{ik\frac{x^2}{2f}} \mathrm{d}x \int_{y_1}^{y_2} e^{ik\frac{y^2}{2f}} \mathrm{d}y \tag{10.51}$$

The integrals are better visualized with the following substitutions:

$$\begin{cases} u = x\sqrt{\dfrac{k}{f\pi}} \\ v = y\sqrt{\dfrac{k}{f\pi}} \\ \psi_0 = \dfrac{Cf\pi}{k} \end{cases}$$

The result is the following integral

$$\psi_P = \psi_0 \int_{u_1}^{u_2} e^{\frac{i\pi u^2}{2}} \mathrm{d}u \int_{v_1}^{v_2} e^{\frac{i\pi v^2}{2}} \mathrm{d}v \tag{10.52}$$

The previous two integrals are known as *Fresnel integrals*. Expanding the complex exponentials, we could write for each coordinate u, v

$$\int_0^s e^{\frac{i\pi z^2}{2}} \mathrm{d}z = \int_0^s \cos\left(\frac{\pi z^2}{2}\right) \mathrm{d}z + i \int_0^s \sin\left(\frac{\pi z^2}{2}\right) \mathrm{d}z = C(s) + iS(s) \tag{10.53}$$

These two integral functions define a spiral, known as the *Cornu spiral*. The solution to the square aperture problem will be a piece of the spiral corresponding to an appropriate interval Δs.
In the case of no aperture, i.e. $\Sigma = \mathbb{R}^2$, we have

$$\begin{aligned} \lim_{s\to\infty} C(s) = \lim_{s\to\infty} S(s) = \frac{1}{2} \\ \lim_{s\to-\infty} C(s) = \lim_{s\to-\infty} S(s) = -\frac{1}{2} \end{aligned} \tag{10.54}$$

We obtain the result for an unperturbed wave

$$\psi_P = \psi_0 (1+i)^2 = 2i\psi_0 \tag{10.55}$$

Therefore, indicating $\psi_1 = \psi_0 (1+i)^2$ for the general finite case, i.e. for a square aperture we get

$$\psi_P = \frac{\psi_1}{(1+i)^2} \left(C(u) + iS(u)\right)_{u_1}^{u_2} \left(C(v) + iS(v)\right)_{v_1}^{v_2} \tag{10.56}$$

§§§ 10.3.2.1 Single Slit and Straightedge

Fresnel diffraction from a long slit can be treated using the equations we found before for the 2D square aperture, constrained to a single dimension.
Taken $u \in \mathbb{R}$ and $v \in [v_1, v_2]$, integrating and using the limits we found before we get

$$\psi_P = \frac{\psi_1}{1+i}\left(C(v) + iS(v)\right)_{v_1}^{v_2} \tag{10.57}$$

The straightedge is the limiting case of the single slit for which $v_1 = -\infty$, therefore

$$\psi_P = \frac{\psi_1}{1+i}\left(C(v_2) + iS(v_2)\right) + \frac{1}{2}\psi_1 \tag{10.58}$$

§ 10.4 Fourier Theory of Diffraction

Consider the case of Fraunhofer (far field) again, in the most general case of a generic aperture Σ which might also have generic transmission properties (phase retardation etc…).
All rays leaving Σ in a generic direction which we will specify with the director cosines $\hat{\mathbf{r}} = (\alpha, \beta, \gamma)$, towards a measuring point in the screen which we will call P. Therefore, we can say approximately that

$$P = (X, Y) \simeq (L\alpha, L\beta)$$

Where L is the distance between the aperture and the screen and (X, Y) are the coordinates of the screen, note that we assume that the angles are small in according to the far field approximation, so $\alpha \approx \tan\alpha, \beta \approx \tan\beta$. Also we assume $\gamma \approx 1$. Said Q a point on the aperture, with coordinates $(x, y) = \mathbf{R}$. Said $\hat{\mathbf{n}}$ the associated versor of $\mathbf{R}$ we have $\hat{\mathbf{n}} = \hat{\mathbf{r}}$ we have that in a small movement of the point Q

$$\delta r = \mathbf{R} \cdot \hat{\mathbf{n}} = \alpha x + \beta y = x\frac{X}{L} + y\frac{Y}{L}$$

The KF integral of this configuration is then

$$\psi\left(X, Y\right) = \iint_\Sigma e^{ik\delta r}\mathrm{d}^2 s = \iint_\Sigma e^{\frac{ik}{L}(xX+yY)}\mathrm{d}^2 s \tag{10.59}$$

This works for a uniform aperture Σ.
For a non uniform aperture we must introduce an aperture function $g(x, y)$ defined in a way such that $g(x, y)\mathrm{d}x\mathrm{d}y$ is the amplitude of the diffracted wavelet by a surface element $\mathrm{d}^2 s$. Thus, introducing the following substitution $(\mu(X), \nu(Y))$ which we will call the "spatial frequency"

$$\begin{aligned} \mu &= \frac{kX}{L} \\ \nu &= \frac{kY}{L} \end{aligned} \tag{10.60}$$

The KF integral becomes a 2 dimensional Fourier transform of the aperture function

$$\psi(\mu, \nu) = \iint_\Sigma g(x, y) e^{i(\mu x + \nu y)}\,\mathrm{d}x\,\mathrm{d}y \tag{10.61}$$

Written differently, we can say that the image, aka the diffraction pattern ψ is the Fourier pair of the aperture function g.
Consider now a 1D infinite grating, for simplicity. The aperture function can be described as an infinite sum of step functions as follows

$$g_{dg}(y) = \sum_{n=0}^{\infty} g_n \cos(n\nu_0 y) \qquad \nu_0 = \frac{2\pi}{h} \tag{10.62}$$

Its Fourier transform is the diffraction pattern that we already know for the gratings, where the higher order maxima correspond to Fourier components with $n > 1$ of $g(y)$

§§ 10.4.1 Apodization

The process of *apodization* (a-pod-ization - to remove the feet) is the process of modification of the aperture function such that the energy is redistributed in the diffraction pattern. It's employed for reducing the irradiance of secondary maxima of the wave.
Consider a single slit, here the aperture function is the set indicator function, which we will indicate with χ. For a slit long b we have

$$g(y) = \chi_{[-b/2,b/2]}(y) \tag{10.63}$$

The Fourier transform of this function is obvious, and after a quick integration we get

$$\hat{\mathcal{F}}\left[g(y)\right](\nu) = b\,\mathrm{sinc}\left(\frac{1}{2}\nu b\right) = \psi(\nu) \tag{10.64}$$

Note that this is exactly what we found before in far field regimes.
Suppose that now we apodize g, making it a new function, as an example

$$g_A(y) = \cos\left(\frac{\pi y}{b}\right) \qquad y \in \left[-\frac{b}{2}, \frac{b}{2}\right]$$

Remembering that

$$\cos\left(\frac{\pi y}{b}\right) = \frac{1}{2}\left(e^{\frac{i\pi y}{b}} + e^{-\frac{i\pi y}{b}}\right)$$

And then integrating using the linearity of the integral, plus a couple power rules we get the apodized diffraction pattern

$$\psi_A(\nu) = \frac{1}{2}\int_{-\frac{b}{2}}^{\frac{b}{2}} e^{iy(\nu+\frac{\pi}{b})} + e^{iy(\nu-\frac{\pi}{b})}\,\mathrm{d}y = \frac{b}{b\nu-\pi}\sin\left[\frac{1}{2}(\nu b-\pi)\right] + \frac{b}{b\nu+\pi}\sin\left[\frac{1}{2}(\nu b+\pi)\right]$$

Or, using trigonometric identities, we could write

$$\psi_A(\nu) = \cos\left(\frac{1}{2}\nu b\right)\left(\frac{b}{b\nu-\pi} + \frac{b}{b\nu+\pi}\right) \tag{10.65}$$

The apodized $\psi_A(\nu)$ is clearly smaller than the original unapodized function for higher frequencies.
Apodization techniques are usually used on telescope apertures, permitting that dimmer objects can be seen next to brighter objects, which clearly helps in the observation of binary systems, or even in this day and age, exoplanets.

§§ 10.4.2 Spatial Filtering

Consider again the previous setup, suppose that the xy plane (Σ) is the location of a *coherently* illuminated object which is imaged by some optical system like a lens, and then imaged again on the focal plane of the lens. Said μ, ν the plane of the optical system and XY the plane of the image, we have that $g(x, y)$ is simply our object. The shadow casted on the XY plane, which we will call $\tilde{g}(x, y)$, is the Fourier transform of the optical disturbance ψ on the lens (μ, ν). Or

$$S \longrightarrow g(x,y) \longrightarrow \hat{\mathcal{F}}[g] = \psi(\mu,\nu) \longrightarrow \hat{\mathcal{F}}[\psi] = \tilde{g}(X,Y) \tag{10.66}$$

If all (μ, ν) are transmitted equally by the optical system, then $\tilde{g} \propto g$ and the image is a precise representation of the object. This is clearly only possible for a finite aperture, thus some spatial frequencies get limited.
Optical phenomena like aberrations, defects, etc…result in a modified ψ. This modification can be incorporated via a *transfer function* T, such that the modified ψ, which we will call ψ' is obtained by multiplication

$$\psi'(\mu,\nu) = T(\mu,\nu)\psi(\mu,\nu) \tag{10.67}$$

The image function obtained, g', will then be the Fourier transform of $T\psi$, and the integration limits are determined by the function T. The transfer function can be modified by placing screens and apertures on the (μ, ν) plane (the lens). This procedure is known as *spatial filtering* due to its parallels to electrical filters.

§§ 10.4.3 Phase Gratings

A phase grating is an object composed by 2 high-low refraction index strata which are perfectly transparent.
The aperture function of such object is

$$g(y) = e^{i\phi(u)} \tag{10.68}$$

Where $\phi(y)$ is a periodic step function with height $\Delta\phi = kz\Delta n$, where z is the thickness and $\Delta n = n_H - n_L$ is the difference of refraction index between the two strata.
If $\Delta\phi << 1$ we can approximate the exponential to

$$g(y) \approx 1 + i\phi(y)$$

Thus, the optical disturbance is

$$\psi(\nu) = \int_{-\frac{b}{2}}^{\frac{b}{2}} e^{i\nu y}\mathrm{d}y + i\int_{-\frac{b}{2}}^{\frac{b}{2}} \phi(y)e^{i\nu y}\mathrm{d}y = \mathfrak{Re}\psi + i\mathfrak{Im}\psi \tag{10.69}$$

Indicating the real part with ψ_1 and the imaginary part with ψ_2, we have that the first function is the whole object diffraction pattern, and the second function is the diffraction pattern of the function ϕ. By definition the two function are dephased by $\pi/2$.
Phase gratings are used in the phase contrast method together with phase plates which shift ψ_2 by an another $\psi/2$.
A *phase plate* is a transparent glass plate whose thickness is $z = z_{plate} + \frac{1}{4}\lambda$. The thicker section is on the optical object plane (μ, ν). The phase plate applies a parity transformation on the function, which

makes the image function a sum of the Fourier transform of both.
Phase contrast works then like a phase modulated signal getting transformed to an amplitude modulated signal through this parity transform

11 Optics of Solids

§ 11.1 The General Wave Equation

§§ 11.1.1 Macroscopic Fields

The electromagnetic field at any given point of space is described by 4 quantities

1. $\frac{\mathrm{d}Q}{\mathrm{d}V} = \rho$, the volumetric density of charge
2. $\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}V} = \mathbf{P}$ dielectric polarization, also known as shearing polarization
3. $\frac{\mathrm{d}\mathbf{m}}{\mathrm{d}V} = \mathbf{M}$ magnetization
4. $\frac{\mathrm{d}\mathbf{j}}{\mathrm{d}V} + \frac{1}{c}\frac{\mathrm{d}}{\mathrm{d}V}\frac{\partial \mathbf{E}}{\partial t} = \mathbf{J}$ total current density

All these quantities can be considered as smooth with respect to the microscopic variations that we have due to the discrete composition of matter.
Noting how the polarization is tied to the polarization charge density, and the magnetization is tied to the magnetization currents, we can modify Maxwell's equations as follows, where we will indicate the polarization charge density ρ_P and the magnetization currents as $\mathbf{J}_m$, the flux of the polarization charges is here indicated with Φ_P

$$\begin{cases} \nabla \cdot \mathbf{E} = \dfrac{1}{\epsilon_0}(\rho + \rho_P) \\ \nabla \times \mathbf{E} = -\mu_0 \dfrac{\partial \mathbf{B}}{\partial t} \\ \nabla \cdot \mathbf{B} = 0 \\ \nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \mathbf{J}_m + \epsilon_0 \mu_0 \dfrac{\partial \mathbf{E}}{\partial t} + \mu_0 \dfrac{\mathrm{d}\Phi_P}{\mathrm{d}t} \end{cases} \tag{11.1}$$

We then substitute the following two relations for $\rho_P, \mathbf{J}_m$

$$\begin{aligned} \rho_P &= -\nabla \cdot \mathbf{P} \\ \mathbf{J}_m &= \nabla \times \mathbf{M} \end{aligned} \tag{11.2}$$

Therefore, the equations become

$$\begin{cases} \nabla \cdot \left(\mathbf{E} + \frac{1}{\epsilon_0} \mathbf{P} \right) = \frac{\rho}{\epsilon_0} \\ \nabla \times \mathbf{E} = -\mu_0 \frac{\partial \mathbf{H}}{\partial t} - \frac{\partial \mathbf{M}}{\partial t} \\ \nabla \cdot (\mathbf{H} + \mathbf{M}) = 0 \\ \nabla \times (\mu_0 \mathbf{H} + \mathbf{M}) = \mathbf{J} + \frac{\partial}{\partial t} (\epsilon_0 \mathbf{E} + \mathbf{P}) \end{cases} \tag{11.3}$$

Note that we will treat, for now, only linear media, therefore we can define two new quantities, i.e. the scalars $\epsilon = \epsilon_0(1 + \chi_e)$ and $\mu = \mu_0(1 + \chi_m)$.
Remembering the relations between the vector fields $\mathbf{D}, \mathbf{H}$ and the vector fields $\mathbf{E}, \mathbf{B}$ we have, for linear media

$$\begin{aligned} \mathbf{D} &= \epsilon \mathbf{E} = \epsilon_0 \mathbf{E} + \mathbf{P} \\ \mathbf{B} &= \mu \mathbf{H} = \mu_0 \mathbf{H} + \mathbf{M} \end{aligned} \tag{11.4}$$

In general these relations are not scalar, but of tensorial nature, where the (electric/magnetic) permittivity χ is a rank 2 tensor, as we will see later with crystals.

§§ 11.1.2 The Wave Equation in Solids

For finding the general wave equation in solids we firstly proceed to understand how do solids work in general.
For a neuter solid, we can assume that it is electrically neutral, therefore we can assume $\rho = 0, \mathbf{M} = 0$.
Maxwell's equations are therefore modified to the following set of coupled PDEs

$$\begin{cases} \epsilon_0 \nabla \cdot \mathbf{E} = -\nabla \cdot \mathbf{P} \\ \nabla \times \mathbf{E} = -\mu_0 \frac{\partial \mathbf{H}}{\partial t} \\ \nabla \cdot \mathbf{H} = 0 \\ \nabla \times \mathbf{H} = \mathbf{J} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} + \frac{\partial \mathbf{P}}{\partial t} \end{cases} \tag{11.5}$$

Taking the curl of the second equation and the time derivative of the fourth, we get

$$\nabla \times \nabla \times \mathbf{E} + \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} = -\mu_0 \frac{\partial J}{\partial t} - \mu_0 \frac{\partial^2 \mathbf{P}}{\partial t^2} \tag{11.6}$$

Or, using the definition of the d'Alambertian, after applying the vector identity for the double curl

$$\nabla (\nabla \cdot \mathbf{E}) + \Box \mathbf{E} = -\mu_0 \frac{\partial \mathbf{J}}{\partial t} - \mu_0 \frac{\partial^2 \mathbf{P}}{\partial t^2}$$

We could insert the first equation inside the first addendum, but we'll limit ourselves to the generic equation (11.6).
In that equation we define the right hand side as the *source terms* for the electromagnetic wave $\mathbf{E}$.
We will approximate the equation further later, noting that matter can be in general divided in two types of media (for what interests us right now)

1. Dielectric media (nonconducting)
2. Conducting media

§ 11.2 Waves in Dielectrics

We now begin to analyze the behavior of electromagnetic waves in dielectrics. By definition, dielectrics are non-conducting, therefore microscopically they can be identified with electrons tightly bound to the atoms, and therefore, with no freely moving electrons, we can say $\mathbf{J} = 0$.
The generic wave equation then becomes

$$\nabla \times \nabla \times \mathbf{E} + \frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} = -\mu_0 \frac{\partial^2 \mathbf{P}}{\partial t^2} \tag{11.7}$$

By definition of $\mathbf{P}$, we have that applying an electromagnetic field to the body, the electrons and the cores of the atom get displaced in opposite direction due to having opposite charge signs, generating microscopic dipole moment and therefore the polarization $\mathbf{P}$.
Due to the boundedness of the electrons generating this dipole moment field, we can think that the force exerted on them is elastic. Said K the elastic force constant for this rebound force and used Newton's second law we have that the general force applied is

$$\mathbf{F} = -e\mathbf{E} = K\mathbf{r}$$

Where $\mathbf{r}$ is the displacement of the electron from the equilibrium state. We can also say, by definition of the polarization $\mathbf{P}$ and by the discreteness of the system, that if there are N electrons with charges $-e$, the macroscopic polarization is

$$\mathbf{P}_s = Q_T \mathbf{r} = -Ne\mathbf{r}$$

Note: we're not yet considering dynamic fields, therefore $\mathbf{P}_s$ is the *static* polarization.
Putting the two equations together and solving for $\mathbf{r}$ in the first equation, we get the following result

$$\mathbf{r} = -\frac{e}{K}\mathbf{E} \implies \mathbf{P}_s = -\frac{Ne^2}{K}\mathbf{E} \tag{11.8}$$

We now "turn on" the time dependence of our wave. Due to the oscillation of $\mathbf{E}$ we know that $\mathbf{r}$ will also oscillate, with a damping given from the boundedness of the electrons.
The force equation therefore becomes a *damped harmonic oscillator*

$$\mathbf{F} = m\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} + m\gamma\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} + K\mathbf{r} = -e\mathbf{E} \tag{11.9}$$

We solve this equation using the method of similarity, since we know already that $\mathbf{E} \propto e^{-i\omega t}$, thus, said $\mathbf{r} \propto e^{-i\omega t}$ we have

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = -i\omega\mathbf{r}e^{-i\omega t} \qquad \frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} = -\omega^2\mathbf{r}e^{-i\omega t}$$

Using the wholeness of the exponential in the complex plane, we get the following solution

$$\left(m\omega^2 + im\gamma\omega - K\right)\mathbf{r} = e\mathbf{E} \tag{11.10}$$

Solving for $\mathbf{r}$ and imposing that $\mathbf{P} = -Ne\mathbf{r}$ we have that the polarization is therefore the following

$$\mathbf{P} = \frac{Ne^2}{K - m\omega^2 - im\gamma\omega}\mathbf{E}$$

Or, bringing outside Ne^2/m and defining the *effective resonance frequency* ω_0 as follows

$$\omega_0 = \sqrt{\frac{K}{m}} \tag{11.11}$$

We have the final result for the dynamic polarization field

$$\mathbf{P} = \frac{Ne^2}{m}\left(\frac{1}{\omega_0^2 - \omega^2 - i\gamma\omega}\right)\mathbf{E} \tag{11.12}$$

The effective resonance frequency ω_0 we defined before clearly depends on the medium studied, since K clearly varies with material.
The name is not random, since for frequencies around ω_0 we expect (and see) optical resonance, giving this frequency also the nickname of the *natural frequency*.
We can now plug the results on the general wave equation (11.7), getting the following PDE for our field (note that we indicate $c^2\epsilon_0 = \mu_0^{-1}$

$$\nabla \times \nabla \times \mathbf{E} + \frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} + \frac{Ne^2}{c^2\epsilon_0 m}\left(\frac{1}{\omega_0^2 - \omega^2 - i\gamma\omega}\right)\frac{\partial^2 \mathbf{E}}{\partial t^2} = 0 \tag{11.13}$$

Using the fact that $\mathbf{E} \propto \mathbf{P}$ and $\nabla \cdot \mathbf{E} = 0$ we can expand the double curl as $-\nabla^2\mathbf{E}$, getting the general wave equation in a dielectric

$$\nabla^2\mathbf{E} = \left(1 + \frac{Ne^2}{m\epsilon_0}\frac{1}{\omega_0^2 - \omega^2 - i\gamma\omega}\right)\frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{11.14}$$

§§ 11.2.1 Resonant Frequency and Dispersion

From the previous equation, we can try to find a general solution in terms of plane waves

$$\mathbf{E} = \mathbf{E}_0 e^{i\kappa z - i\omega t} \qquad \kappa \in \mathbb{C} \tag{11.15}$$

Noting that $\nabla^2\mathbf{E} = -\kappa^2\mathbf{E}$ and $\partial_t^2\mathbf{E} = -\omega^2\mathbf{E}$ we have that this plane wave is a solution if and only if

$$\kappa^2 = \frac{\omega^2}{c^2}\left(1 + \frac{Ne^2}{m\epsilon_0}\frac{1}{\omega_0^2 - \omega^2 - i\gamma\omega}\right) \tag{11.16}$$

Noting that $z^{-1} = \overline{z}|z|^{-2}$ for complex numbers, we could rewrite

$$\frac{1}{\omega_0^2 - \omega^2 - i\gamma\omega} = \frac{\omega_0^2 - \omega^2 + i\gamma\omega}{\left(\omega_0^2 - \omega^2\right)^2 + \gamma^2\omega^2}$$

Thus we can rewrite κ^2 as follows

$$\kappa^2 = \frac{\omega^2}{c^2} + \frac{Ne^2\omega^2}{m\epsilon_0 c^2}\frac{\omega_0^2 - \omega^2 + i\gamma\omega}{\left(\omega_0^2 - \omega^2\right)^2 + \gamma^2\omega^2} \tag{11.17}$$

From κ we define the *complex refraction index* $\mathcal{N}$ as follows

$$\kappa = \frac{\omega}{c}\mathcal{N} \tag{11.18}$$

Using the explicitly complex form of $\kappa = k + i\alpha$, we can rewrite the wave solution as follows

$$\mathbf{E} = \mathbf{E}_0 e^{i(k+i\alpha)z - i\omega t} = \mathbf{E}_0 e^{ikz - i\omega t} e^{-\alpha z} \tag{11.19}$$

The exponential damping of the wave indicates the presence of *absorption* in the medium. In fact, noting that $I \propto \|\mathbf{E}\|^2$ we have

$$I \propto e^{-2\alpha z}$$

The coefficient $2\alpha = a$ is known as the *absorption coefficient* of the medium.
Going back to the definition of κ (11.17) we have that the square of the complex refraction index is

$$\mathcal{N}^2 = 1 + \frac{Ne^2}{m\epsilon_0}\frac{\omega_0^2 - \omega^2 + i\gamma\omega}{\left(\omega_0^2 - \omega^2\right)^2 + \gamma^2\omega^2} \tag{11.20}$$

Indicating $\mathcal{N} = n + i\eta$ and reusing the relationship between κ and $\mathcal{N}$ (11.18) we have

$$\mathcal{N} = \frac{c}{\omega}\kappa \implies \frac{c}{\omega}(k + i\alpha) = n + i\eta \tag{11.21}$$

Separating real and imaginary parts and equating them we have

$$\begin{aligned} \mathfrak{Re}\{\mathcal{N}\} &= \frac{c}{\omega}\mathfrak{Re}\{\kappa\} = \frac{c}{\omega}k = n \\ \mathfrak{Im}\{\mathcal{N}\} &= \frac{c}{\omega}\mathfrak{Im}\{\kappa\} = \frac{c}{\omega}\alpha = \eta \end{aligned} \tag{11.22}$$

1. From the real part equality we get the usual relation between the *real* refraction index and what we will call the wavenumber
$$k = \frac{\omega}{c}n$$

2. From the imaginary equality we get the relation between the absorption coefficient and the imaginary part of the complex refraction index
$$\alpha = \frac{\omega}{c}\eta$$

These relationships we found are not yet explicit. We can begin to try to find their explicit form by starting again from the square of the complex refraction index. Noting that

$$\mathcal{N}^2 = n^2 - \eta^2 + 2in\eta = 1 + \frac{Ne^2}{m\epsilon_0}\left(\frac{\omega_1^2 - \omega^2}{\left(\omega_0^2 - \omega^2\right)^2 + \gamma^2\omega^2} + i\frac{\gamma\omega}{\left(\omega_0^2 - \omega^2\right)^2 + \gamma^2\omega^2}\right)$$

Equating real and imaginary parts again, we get a rather complex system of two equations which can be solved numerically

$$\begin{aligned} n^2 - \eta^2 &= 1 + \frac{Ne^2}{m\epsilon_0}\frac{\omega_0^2 - \omega^2}{\left(\omega_0^2 - \omega^2\right)^2 + \gamma^2\omega^2} \\ 2n\eta &= \frac{Ne^2}{m\epsilon_0}\frac{\gamma\omega}{\left(\omega_0^2 - \omega^2\right)^2 + \gamma^2\omega^2} \end{aligned} \tag{11.23}$$

It's clear that both n and η depend on frequency. This frequency dependence, especially for n, is known as *dispersion*.
Plotting the numerical solutions to the previous system we get two particular functional shapes for both n and η.
Centering ourselves around ω_0 we can see that the absorption η has a maximum for $\omega \approx \omega_0$, and $n > 1$ for $\omega < \omega_0$.
The peak of absorption of the dielectric indicates how transparent dielectrics have their resonant frequencies in the ultraviolet region of the electromagnetic spectrum, therefore shifting the absorption peak.
For n, we can define two kinds of dispersion with respect to the natural frequency ω_0.

1. Normal dispersion, when $\omega \lesssim \omega_0$ and n increases with increasing frequency
2. Anomalous dispersion, when $\omega \gtrsim \omega_0$ and n decreases with increasing frequency.

In general tho, we have that $\lim_{\omega\to 0} n(\omega) = 1$ and $\lim_{\omega\to\infty} n(\omega) = 1$.

§§ 11.2.2 Sellmeier Equation

All previous calculations are valid if and only if the electrons are bound *equally* to their respective atom, which is impossible considering quantum mechanics.
In general we might assume that each fraction of electrons f_j has its resonant frequency ω_j and damping constant γ_j. We then modify the complex refraction index as follows

$$\mathcal{N}^2 = 1 + \frac{Ne^2}{m\epsilon_0}\sum_{j=0}^{N}\frac{f_j}{\omega_j^2 - \omega^2 - i\gamma_j\omega} \tag{11.24}$$

The fractions f_j are known as *oscillator strengths*, due to their quantum harmonic origin.
Note that by definition, we have that at limit frequencies

$$\begin{cases} \lim\limits_{\omega\to 0}\mathcal{N}^2(\omega) = 1 + \dfrac{Ne^2}{m\epsilon_0}\displaystyle\sum_{j=0}^{N}\frac{f_j}{\omega_j} = 1 + \chi_e \\ \lim\limits_{\omega\to\omega}\mathcal{N}^2(\omega) = \infty \end{cases} \tag{11.25}$$

And if we consider only the real part in the approximation $\gamma_j << 1$, we can derive a dispersion formula known as *Sellmeier's equation*

$$n^2 = 1 + \frac{Ne^2}{m\epsilon_0}\sum_{j=0}^{N}\frac{f_j}{\omega_j^2 - \omega^2} \tag{11.26}$$

This equation is related to Cauchy's dispersion relation

$$n(\lambda) = A + \frac{B}{\lambda^2} \tag{11.27}$$

Note how Sellmeier's formula is more general in scope, while Cauchy's equation is directly retrieved via experimental considerations in the optical part of the spectrum

§ 11.3 Waves in Conductors

We begin by reconsidering the general wave equation (11.6). As we considered the electrons bound to the atoms when analyzing dielectrics, we consider conductors as practically the opposite. Here all electrons are not bound to the atoms but instead they're freely moving on the surface of the medium. Due to this, we can say that $\mathbf{P} \approx 0$, and the wave equation we are going to consider is the following

$$\nabla \times \nabla \times \mathbf{E} + \frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} = -\mu_0 \frac{\partial \mathbf{J}}{\partial t} \tag{11.28}$$

We begin to find a solution by again considering a static field.
We begin by noting that the velocity of the electron, is deeply tied to the current. For N electrons we have

$$\mathbf{J} = -Ne\mathbf{v} \tag{11.29}$$

The force equation that these electron will follow is not anymore a damped harmonic oscillator, but can be instead considered as the equation of free motion minus a "drag" term given by the attraction by the atoms. Thus, we can write

$$m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} - \frac{m}{\tau}\mathbf{v} = -e\mathbf{E} \tag{11.30}$$

Rewriting the velocity in terms of the current from equation (11.29) we have, after multiplying both sides by Ne/m

$$\frac{\mathrm{d}\mathbf{J}}{\mathrm{d}t} + \tau^{-1}\mathbf{J} = \frac{Ne^2}{m}\mathbf{E} \tag{11.31}$$

This equation has the associate homogeneous solution

$$\mathbf{J}(t) = \mathbf{J}_0 e^{-\frac{t}{\tau}} \tag{11.32}$$

Here, the constant τ is known as the *relaxation time* of the medium. The homogeneous solution is known technically as *transient current*.
We now consider the two outstanding cases: static and harmonic electric fields.
In the first case the time derivative of the current is zero, and the force equation reduces to

$$\tau^{-1}\mathbf{J} = \frac{Ne^2}{m}\mathbf{E} \tag{11.33}$$

From Ohm's law we can also say that $\mathbf{J} = \sigma\mathbf{E}$, where σ is the conductivity of the metal. Inserting into the previous equation, we have that in the case of a static field it is

$$\sigma = \frac{Ne^2\tau}{m} \tag{11.34}$$

We now try to apply Ohm's law to the harmonic case. We begin by using the similarity method, therefore, we have

$$\frac{\partial \mathbf{J}}{\partial t} = -i\omega\mathbf{J}$$

Therefore, the force equation implies, after substituting σ in the right hand side

$$\left(\tau^{-1} - i\omega\right)\mathbf{J} = \frac{\sigma}{\tau}\mathbf{E}$$

Solving for $\mathbf{J}$, we get the harmonic version of Ohm's law

$$\mathbf{J} = \frac{\sigma}{\tau\left(\tau^{-1} - i\omega\right)}\mathbf{E} = \frac{\sigma}{1 - i\omega\tau}\mathbf{E} \tag{11.35}$$

Note how the limit of zero frequency of this solution is exactly Ohm's law in its original form.

§§ 11.3.1 Skin Depth

We now begin to attack the equation (11.28). Noting that $\mathbf{P} = 0$, we have that $\nabla\nabla\cdot\mathbf{E} = 0$, and writing in terms of the d'Alambertian, the equation becomes way easier to solve

$$\Box\mathbf{E} = -\mu_0\frac{\partial \mathbf{J}}{\partial t} \tag{11.36}$$

Applying the d'Alambertian to the generic plane wave solution we get

$$\Box\mathbf{E} = -\left(\frac{\omega^2}{c^2} + \kappa^2\right)\mathbf{E}$$

And inserting (11.35) on the right hand side after multiplying by $-\mu_0$ and deriving once with respect to time we have

$$\Box\mathbf{E} = -\frac{\mu_0\sigma}{1 - i\omega\tau}\frac{\partial \mathbf{E}}{\partial t} \tag{11.37}$$

And, therefore

$$\left(\kappa^2 + \frac{\omega^2}{c^2}\right)\mathbf{E} = \frac{i\omega\mu_0\sigma}{1 - i\omega\tau}\mathbf{E}$$

Rearranging the equation we have the definition of κ^2 in conductors

$$\kappa^2 = \left(\frac{\omega^2}{c^2} + \frac{i\omega\mu_0\sigma}{1 + i\omega\tau}\right) \tag{11.38}$$

As before, noting that $\kappa = k + i\alpha$, we begin analyzing this complex wavenumber in limit cases. Firstly, considering the case of $\omega \to 0$, we have that Ohm's law applies again, and

$$\kappa_{LF}^2 \approx i\omega\mu_0\sigma \implies \kappa_{LF} = \sqrt{i\omega\mu_0\sigma} = \sqrt{\frac{\omega\mu_0\sigma}{2}}\,(1 + i)$$

Clearly here we have $k = \alpha$

$$k = \alpha = \sqrt{\frac{\omega\mu_0\sigma}{2}} \tag{11.39}$$

Also, considering the complex refraction index

$$\mathcal{N}_{LF} = \frac{c}{\omega}\sqrt{\frac{\omega\mu_0\sigma}{2}}(1+i) = \sqrt{\frac{\sigma}{2\omega\epsilon_0}}(1+i)$$

Which also gives

$$n = \eta = \sqrt{\frac{\sigma}{2\omega\epsilon_0}} \tag{11.40}$$

These results can be used to better analyze the behavior of waves in conductors. Considering the e-folding value of the wave amplitude in the medium, we have that it's deeply tied to α. We define the depth of penetration of the wave in the conductor the *skin depth* of the medium δ, defined as follows

$$\delta = \frac{1}{\alpha} = \sqrt{\frac{2}{\omega\mu_0\sigma}} \tag{11.41}$$

Rewriting this in terms of vacuum wavelength $\lambda_0 = 2\pi c/\omega$ we have

$$\delta = \sqrt{\frac{\lambda_0}{\pi c\mu_0\sigma}} \tag{11.42}$$

Note that this depth is inversely proportional to the square root of the conductivity, and explains why good conductors are highly opaque.

§§ 11.3.2 Plasma Frequency and Dispersion

We now analyze κ at all frequencies. As before we have

$$\kappa^2 = \frac{\omega^2}{c^2} + \frac{i\omega\mu_0\sigma}{1+i\omega\tau}$$

And

$$\mathcal{N}^2 = \frac{c^2}{\omega^2}\kappa^2 = 1 + \frac{ic^2\mu_0\sigma}{\omega}\frac{1}{1+i\omega\tau}$$

As for dielectrics, here it's possible to (hardly) see a particular resonant frequency, known as the *plasma frequency* ω_p defined as follows

$$\omega_p = \sqrt{\frac{\mu_0\sigma c^2}{\tau}} = \sqrt{\frac{Ne^2\tau}{m}}\sqrt{\frac{\mu_0 c^2}{\tau}} = \sqrt{\frac{Ne^2}{m\epsilon_0}} \tag{11.43}$$

Which lets us rewrite the complex refraction index as

$$\mathcal{N} = 1 - \frac{\omega_p^2}{\omega^2 + i\omega\tau^{-1}}$$

Again, as with dielectrics, we rewrite the right hand side in a more explicitly complex way by rationalizing the fraction

$$\mathcal{N}^2 = 1 - \frac{\omega_p^2}{\omega^2 + \tau^{-2}} + \frac{i}{\omega\tau}\frac{\omega_p^2}{\omega^2 + \tau^{-2}} = n^2 - \eta^2 + 2in\eta \tag{11.44}$$

Again, by equating imaginary and real parts we get the system we have to solve numerically for finding the dispersion relation, precisely we have

$$\begin{cases} \mathfrak{Re}\left\{\mathcal{N}^2\right\} = n^2 - \eta^2 = 1 - \dfrac{\omega_p^2}{\omega^2 + \tau^{-2}} \\ \mathfrak{Im}\left\{\mathcal{N}^2\right\} = 2n\eta = \dfrac{1}{\omega\tau}\dfrac{\omega_p^2}{\omega^2 + \tau^{-2}} \end{cases} \tag{11.45}$$

Both optical parameters are clearly determined by three parameters: frequency ω, plasma frequency ω_p and relaxation time τ.
In general we can get an idea of these parameters. For metals $\tau \propto 10^{-13}$ s which corresponds to a resonance in infrared frequencies, and $\omega_p \propto 10^{15}$ Hz, corresponding to visible and near ultraviolet resonances.
These are quite important results, since (after solution and plotting of the two indexes) it shows how metals become transparent at high frequencies, since $\lim_{\omega\to\infty}\eta = 0$.
More generally we see how n decreases up to the plasma frequency, while the absorption keeps decreasing indefinitely. At ω_p we have that $n = \eta$.

§§§ 11.3.2.1 Waves in Semiconductors

For semiconductors and poor conductors, we can find a complex refraction index as the sum of the refraction index we found for conductors and dielectrics, which, precisely is

$$\mathcal{N}_{SC}^2 = \mathcal{N}_C^2 + \mathcal{N}_D^2 = 1 - \frac{\omega_p^2}{\omega^2 + i\omega\tau^{-1}} + \frac{Ne^2}{m\epsilon_0}\sum_{j=0}^{N}\frac{f_j}{\omega_j^2 - \omega^2 - i\gamma_j\omega} \tag{11.46}$$

All solution methods we found before are still valid, and we can derive, albeit with numerical methods, the values of n and η.

§ 11.4 Reflection and Refraction in Absorbing Media

Let ψ be some electromagnetic wave which happens to be incident to some absorbing medium, with complex refraction index $\mathcal{N} = n + i\eta$ and absorption coefficient α. The two things we must keep an eye on will then be

$$\begin{cases} \mathcal{N} = n + i\eta \\ \boldsymbol{\kappa} = \mathbf{k} + i\boldsymbol{\alpha} \end{cases}$$

Suppose that the wave is coming from a non absorbing region from the left, with index n_1, and goes inside the absorbing region on the right with complex refraction index $\mathcal{N}$ on the right.
We will have, as usual, three waves: the incident wave, the reflected wave and the transmitted wave,

with the following exponential dependencies at the boundary:

$$\begin{cases} \psi_1 \propto e^{i\mathbf{k}_1 \cdot \mathbf{r}} \\ \psi_R \propto e^{i\mathbf{k}_R \cdot \mathbf{r}} \\ \psi_T \propto e^{i\boldsymbol{\kappa} \cdot \mathbf{r}} \end{cases} \tag{11.47}$$

As usual, in the first region we will have $\mathbf{k}_1 = \mathbf{k}_R$ and we get the usual law of reflection, while in the second region we must consider that $\boldsymbol{\kappa} = \mathbf{k} + i\boldsymbol{\alpha}$, therefore

$$\mathbf{k}_1 \cdot \mathbf{r} = (\mathbf{k} + i\boldsymbol{\alpha}) \cdot \mathbf{r} \implies \begin{cases} \mathbf{k}_1 \cdot \mathbf{r} = \mathbf{k} \cdot \mathbf{r} \\ \boldsymbol{\alpha} \cdot \mathbf{r} = 0 \end{cases} \tag{11.48}$$

Where we equated the real and imaginary part of the left hand side and right hand side.
It's clear that $\mathbf{k}$ and $\boldsymbol{\alpha}$ do not have the same direction. This kind of wave is known as non homogeneous wave. Note that $\boldsymbol{\alpha} \cdot \mathbf{r} = 0$ implies that $\boldsymbol{\alpha}$ is perpendicular to the boundary.
For non homogeneous waves we can define *two* planes in the absorbing region:

1. Phase planes, where $\mathbf{k}$ is constant
2. Amplitude planes, where $\boldsymbol{\alpha}$ is constant

Denoting θ as our incidence angle and ϕ as the transmission angle we can get via phase matching at the boundary a "Snell" law. This law is clearly not exactly Snell's, since the absorbing nature of the medium makes k actually depend on the transmission angle ϕ. Precisely we have

$$\mathbf{k}_1 \cdot \mathbf{r} = k_1 \sin\theta = \mathbf{k}(\phi) \sin\phi \tag{11.49}$$

The functional relation $k(\phi)$ can be derived from the modified wave equation

$$\nabla^2 \mathbf{E} = \frac{\mathcal{N}^2}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{11.50}$$

Using $\nabla^2 \mathbf{E} = (\boldsymbol{\kappa} \cdot \boldsymbol{\kappa})\mathbf{E}$ and $\partial_t \mathbf{E} = -i\omega \mathbf{E}$ we have

$$(\mathbf{k} + i\boldsymbol{\alpha})^2 \, \mathbf{E} = \frac{\omega^2}{c^2} \mathcal{N}^2 \mathbf{E} = k_0^2 \mathcal{N}^2 \mathbf{E}$$

Equating right hand side and left hand side after writing $\mathcal{N}^2 = (n + i\eta)^2$ and writing the explicit squares, we have

$$k^2 - \alpha^2 + 2i\mathbf{k} \cdot \boldsymbol{\alpha} = k_0^2 \left(n^2 - \eta^2 + 2in\eta\right) \implies \begin{cases} k^2 - \alpha^2 = k_0^2 \left(n^2 - \eta^2\right) \\ k\alpha \cos\phi = k_0^2 n\eta \end{cases} \tag{11.51}$$

The solution to this system of equation is not immediate, but the result can be shown to be the following

$$k\cos\phi + i\alpha = k_0 \sqrt{\mathcal{N}^2 - \sin^2\theta} \tag{11.52}$$

For normal incidence $\sin\theta = 0$, thus

$$k\cos\phi + i\alpha = k_0 \mathcal{N}$$

In a purely formal way we could also imagine to write "Snell's law" as the following relation

$$\mathcal{N} = \frac{\sin\theta}{\sin(z)} \qquad \begin{cases} \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}] \\ z \in \mathbb{C} \end{cases} \tag{11.53}$$

From the previous definition, and with the help of this complex "angle", we can define the following

$$\cos(z) = \sqrt{1 - \frac{\sin^2\theta}{\mathcal{N}^2}} \tag{11.54}$$

Therefore, inserting this into (11.52) we get

$$\begin{aligned} k\cos\phi + i\alpha &= k_0\mathcal{N}\sqrt{1 - \frac{\sin^2\theta}{\mathcal{N}^2}} \\ k\cos\phi + i\alpha &= k_0\mathcal{N}\cos(z) \end{aligned} \tag{11.55}$$

Thus

$$\mathcal{N} = \frac{k\cos\phi + i\alpha}{k_0\cos(z)} \tag{11.56}$$

From the equations (11.54) and (11.56), it's possible to find the reflectance of the absorbent medium. We begin by noting that Maxwell's equations need to be appropriately modified to account for the previous results, and this can be done with the complex value z we defined before.
As usual, in the first region we have

$$\begin{aligned} \mathbf{H} &= \frac{1}{\mu_0\omega}\mathbf{k}_1 \times \mathbf{E} \\ \mathbf{H}_R &= \frac{1}{\mu_0\omega}\mathbf{k}_1 \times \mathbf{E}_R \end{aligned} \tag{11.57}$$

While, in the second region we have

$$\mathbf{H}_t = \frac{1}{\mu_0\omega}\boldsymbol{\kappa} \times \mathbf{E}_T = \frac{1}{\mu_0\omega}(\mathbf{k} \times \mathbf{E}_T + i\boldsymbol{\alpha} \times \mathbf{E}_T) \tag{11.58}$$

Evaluating the modulus of these equations and inserting them for the Senkrecht (s) and Parallel (p) polarization states, we get the two following systems of equations

$$\begin{cases} E + E_R = E_T \\ (H - H_R)\cos\theta = H_T \end{cases} \qquad \begin{cases} H - H_R = H_T \\ (E + E_R)\cos\theta = E_T\cos(z) \end{cases} \tag{11.59}$$

Which, after dividing by E and inserting the relationship between E and H, can be rewritten as follows

$$\begin{cases} 1 + r_s = t_s \\ k_0(1 - r_s)\cos\theta = (k\cos\phi + i\alpha)\,t_s \end{cases} \qquad \begin{cases} k_0(1 - r_p) = k_0\mathcal{N}t_p \\ (1 + r_p)\cos\theta = t_p\cos(z) \end{cases} \tag{11.60}$$

Starting from the equations for s polarization, we have that

$$(1 - r_s)k_0\cos\theta = (1 + r_s)(k\cos\phi + i\alpha)$$

Noting that on the right hand side we can substitute $k\cos\phi + i\alpha = k_0\mathcal{N}\cos(z)$, we have

$$(1 - r_s)k_0\cos\theta = (1 + r_s)k_0\mathcal{N}\cos(z)$$

Which, after distributing the product and solving for r_s by simple division, we get

$$r_s = \frac{\cos\theta - \mathcal{N}\cos(z)}{\cos\theta + \mathcal{N}\cos(z)} \tag{11.61}$$

From the equation $t_s = 1 + r_s$ we get the second result

$$t_s = 1 + \frac{\cos\theta - \mathcal{N}\cos(z)}{\cos\theta + \mathcal{N}\cos(z)} = \frac{2\cos\theta}{\cos\theta + \mathcal{N}\cos(z)} \tag{11.62}$$

For p polarization the process for finding r_p and t_p is completely analogous to the non absorbing case.

$$\begin{cases} \dfrac{1}{\mathcal{N}}(1 - r_p) = t_p \\ (1 + r_p)\cos\theta = \dfrac{1}{\mathcal{N}}(1 - r_p)\cos(z) \end{cases}$$

Thus, again

$$r_p\left(\cos\theta + \frac{1}{\mathcal{N}\cos(z)}\right) = \frac{1}{\mathcal{N}}\cos(z) - \cos\theta$$

Which gives

$$r_p = \frac{\cos(z) - \mathcal{N}\cos\theta}{\cos(z) + \mathcal{N}\cos\theta} \tag{11.63}$$

Analogously, we have

$$t_p = \frac{1}{\mathcal{N}}(1 - r_p) = \frac{1}{\mathcal{N}}\left(1 - \frac{\cos(z) - \mathcal{N}\cos\theta}{\cos(z) + \mathcal{N}\cos\theta}\right)$$

I.e.

$$t_p = \frac{2\cos\theta}{\cos(z) + \mathcal{N}\cos\theta} \tag{11.64}$$

All reunited in one place, we have

$$\begin{cases} r_s = \dfrac{\cos\theta - \mathcal{N}\cos(z)}{\cos\theta + \mathcal{N}\cos(z)} \\ t_s = \dfrac{2\cos\theta}{\cos\theta + \mathcal{N}\cos(z)} \end{cases} \qquad \begin{cases} r_p = \dfrac{\cos(z) - \mathcal{N}\cos\theta}{\cos(z) + \mathcal{N}\cos\theta} \\ t_p = \dfrac{2\cos\theta}{\cos(z) + \mathcal{N}\cos\theta} \end{cases} \tag{11.65}$$

As usual, it's possible to find the coefficients R and T with the usual evaluations.
Note that for p polarization, if $\mathfrak{Im}\{\mathcal{N}\} \neq 0$, also $r_p \neq 0$, i.e. there is *no Brewster angle in absorbing media*. Instead, we can define the *principal angle of incidence* θ_1, for which $r_p(\theta_1) = \min\{r_p\}$.
As can be imagined, non polarized light and light which is neither s nor p polarized gets transmitted in elliptical polarization. The complex index of refraction can then be evaluated by measuring the transmitted irradiance I_T using *ellipsometry*

§§ 11.4.1 Normal Incidence

In the case of normal incidence, we have as usual $\theta = \phi = 0$, thus $r_p = r_s$, where

$$r = \frac{1-\mathcal{N}}{1+\mathcal{N}} = \frac{1-n-i\eta}{1+n+i\eta} \tag{11.66}$$

Which implies

$$R = \frac{(1-n)^2+\eta^2}{(1+n)^2+\eta^2} \tag{11.67}$$

Note that it's the same equation that we get in the non-absorbing case if $\eta = 0$.
Remembering that $R = R(\mathcal{N})$ and that $\mathcal{N} = \mathcal{N}(\omega)$, we have that in the low frequency limit

$$\mathfrak{Re}\{\mathcal{N}\} = \mathfrak{Im}\{\mathcal{N}\} = \sqrt{\frac{\sigma}{2\omega\epsilon_0}} \tag{11.68}$$

Thus, the reflectance at low frequencies becomes

$$R = 1 - \frac{2}{n} = 1 - \sqrt{\frac{8\omega\epsilon_0}{\sigma}} \tag{11.69}$$

This formula is commonly known as the *Huygens-Rubens formula*. Therefore, for frequency in the red part of the spectrum (big wavelengths), good conductors become better and better reflectors, as is the case for metals. As an example, we can see that for Cu, Ag, Au we have

$$\begin{cases} R(\lambda) \approx 1 & \lambda \in \text{NIR},\ (\lambda \approx 1\ \mu\text{m} - 2\ \mu\text{m}) \\ R(\lambda) \approxeq 1 & \lambda \in \text{FIR},\ (\lambda \geq 20\ \mu\text{m} \end{cases}$$

12 Optics of Crystals

§ 12.1 The Electric Susceptibility Tensor

The main property of crystalline matter, in the context of optics, is *electrical anisotropy*. This indicates that the polarization produced by the application of an electric field to such matter is direction dependent. In crystals, there are usually two possible values of propagation velocity in a given direction, tied to mutually orthogonal polarization states. This property is better known as *birefringence*, i.e. crystals are (usually) doubly refracting.
Note tho how some crystals do not exhibit birefringence, as it's deeply tied to their symmetry. Cubic crystals like NaCl (table salt) *do not* exhibit birefringence, while other kinds of crystals do exhibit it.
A practical way to understand the physics behind this phenomenon is thinking that the lattice atoms bond like springs between each other, but with different strengths K, thus electron displacement when an electric field is applied is different in each bond direction, thus the dependence $\mathbf{P}(\mathbf{E})$ is not a relation of direct proportionality, but it's instead a *tensorial relation*[1].

$$P^i = \epsilon_0 \chi^i_j E_j \tag{12.1}$$

The tensor χ^i_j is the *electric susceptibility tensor*. As usual we can again define the electric displacement field by substituting the constant 1 with the Kronecker delta δ^i_j

$$D^i = \epsilon_0 \left(\delta^i_j + \chi^i_j\right) E^j = \epsilon^i_j E_j \tag{12.2}$$

We defined here ϵ^i_j as the *dielectric tensor*, which is the tensorial equivalent of the constant ϵ.
For ordinary, non-absorbing crystals, the tensor χ^i_j is diagonalizable, and the eigenvalues define the principal susceptibilities of the crystal, each corresponding to a dielectric constant

$$(\epsilon_r)_i = 1 + \chi_i \tag{12.3}$$

We thus modify the general wave equation in non-conducting dielectrics (11.7) as follows

$$\epsilon^i{}_{jk} \partial^j \epsilon^k{}_{lm} \partial^l E^m + \frac{1}{c^2} \frac{\partial^2 E^i}{\partial t^2} = -\frac{\chi^i_j}{c^2} \frac{\partial^2 E^i}{\partial t^2} \tag{12.4}$$

[1] here, we will use again the tensor notation we all love

§§ 12.1.1 K-Surfaces

For solving the previous partial differential equation we impose the usual plane wave solution and shove it inside the partial differential equation. Therefore, as usual

$$\begin{cases} \dfrac{\partial}{\partial x^j} = ik_j \\ \dfrac{\partial}{\partial t} = -i\omega \end{cases}$$

And the partial differential equation becomes

$$\epsilon^i{}_{jk} k^j \epsilon^k{}_{lm} k^l E^m + \frac{\omega^2}{c^2} E^i = -\frac{\omega^2}{c^2} \chi^i_j E^j \tag{12.5}$$

If the susceptibility tensor is diagonal, these become three coupled equations

$$\begin{cases} \left(\dfrac{\omega^2}{c^2} - k_y^2 - k_z^2\right) E_x + k_x k_y E_y + k_x k_z E_z = -\dfrac{\omega^2}{c^2} \chi_1^1 E_x \\ \left(\dfrac{\omega^2}{c^2} - k_x^2 - k_z^2\right) E_y + k_y k_z E_z + k_y k_x E_x = -\dfrac{\omega^2}{c^2} \chi_2^2 E_y \\ \left(\dfrac{\omega^2}{c^2} - k_y^2 - k_x^2\right) E_z + k_z k_y E_y + k_z k_x E_x = -\dfrac{\omega^2}{c^2} \chi_3^3 E_z \end{cases} \tag{12.6}$$

In this configuration, the system is quite complex to solve, so for now, suppose that the wave is propagating inside the crystal in what we choose as the x direction, thus $k_y = k_z = 0$. The system becomes

$$\begin{cases} \dfrac{\omega^2}{c^2} E_x = -\dfrac{\omega^2}{c^2} \chi_1^1 \\ \left(\dfrac{\omega^2}{c^2} - k^2\right) E_y = -\dfrac{\omega^2}{c^2} \chi_2^2 E_y \\ \left(\dfrac{\omega^2}{c^2} - k^2\right) E_z = -\dfrac{\omega^2}{c^2} \chi_3^3 E_z \end{cases}$$

The possible solutions are 2. Clearly $E^i \perp k^i$, but:

1. $E_y \neq 0$, then $k_1 = \sqrt{1 + \chi_2^2}$
2. $E_z \neq 0$, then $k_2 = \sqrt{1 + \chi_3^3}$

Since $\frac{\omega}{k}$ is the phase velocity of the wave, we can define *two distinct phase velocities*

$$\begin{cases} u_1 = \dfrac{c}{\sqrt{1 + \chi_2^2}} \\ u_2 = \dfrac{c}{\sqrt{1 + \chi_3^3}} \end{cases} \tag{12.7}$$

More generally, we could write everything in terms of refraction indexes $n = \sqrt{1 + \chi}$. For a diagonalizable susceptibility tensor we have at most 3 refraction indexes n_i, known as the *principal refraction*

indexes

$$\begin{cases} n_1 = \sqrt{1+\chi_1^1} \\ n_2 = \sqrt{1+\chi_2^2} \\ n_3 = \sqrt{1+\chi_3^3} \end{cases} \tag{12.8}$$

These indexes come in handy to simplify equation (12.6).
Taken back that system of equations, we have that in order to have non banal solutions, we must have that

$$\det \begin{pmatrix} \frac{n_1^2\omega^2}{c^2} - k_y^2 - k_z^2 & k_y k_z & k_z k_z \\ k_y k_x & \frac{n_2^2\omega^2}{c^2} - k_x^2 - k_z^2 & k_y k_z \\ k_z k_y & k_z k_y & \frac{n_3^2\omega^2}{c^2} - k_x^2 - k_y^2 \end{pmatrix} \neq 0 \tag{12.9}$$

The result of this equation is a surface in $\mathbb{R}^3$, or to be precise, a surface in $\mathbf{k}$-space.
We begin to solve for xy, therefore imposing $k_z = 0$.
Evaluating the determinant we get the following algebraic equation

$$\left(\frac{n_3^2\omega^2}{c^2} - k_x^2 - k_y^2\right)\left[\left(\frac{n_1^2\omega^2}{c^2} - k_y^2\right)\left(\frac{n_2^2\omega^2}{c^2} - k_x^2\right) - k_x^2 k_y^2\right] = 0 \tag{12.10}$$

The solutions are two, which clearly are

$$\begin{cases} k_x^2 + k_y^2 = n_3^2 \dfrac{\omega^2}{c^2} \\ \left(\dfrac{n_1\omega}{c}\right)^2 \left(\dfrac{n_2\omega}{c}\right)^2 - k_y^2 \left(\dfrac{n_2\omega}{c}\right)^2 - k_x \left(\dfrac{n_1\omega}{c}\right)^2 = 0 \end{cases}$$

The first solution is already clear, and it's a circle with radius $n_3\omega/c$. The second can be made clearer dividing by $\frac{n_1^2\omega^2}{c^2}\frac{n_2^2\omega^2}{c^2}$. The resulting equation is that of an ellipse

$$\frac{k_x^2}{\frac{n_2^2\omega^2}{c^2}} + \frac{k_y^2}{\frac{n_1^2\omega^2}{c^2}} = 1$$

Analogous equations can be derived for the planes xz and yz.
The two surfaces, which we will call C_1 and Ce_2, are then simply the following sets in $\mathbf{k}-$space

$$\begin{cases} C_1 = \left\{k_x^2 + k_y^2 = \left(\dfrac{n_3\omega}{c}\right)^2\right\} \\ Ce_2 = \left\{\left(\dfrac{c}{n_2\omega}\right)^2 k_x^2 + \left(\dfrac{c}{n_1\omega}\right)^2 k_y^2 = 1\right\} \end{cases} \tag{12.11}$$

Solving also for the other planes we finally get the complete solution as a sphere S_1 and an ellipsoid E_2.
The $\mathbf{k}$ then lays on both surfaces, which together form an *inner spherical sheet* and an *outer ellipsoidal sheet*. The two different possible values that k can take then define the two different orthogonal polarizations of $\mathbf{E}$, with two different phase velocities.

Said z the propagation direction of the electromagnetic wave and said $\hat{\boldsymbol{\chi}}_1, \hat{\boldsymbol{\chi}}_2$ the two principal directions of the crystal in the xy plane, we have then that a wave with generic polarization gets decomposed in the following two components

$$\mathbf{E}(\mathbf{r},t) = (\mathbf{E}\cdot\hat{\boldsymbol{\chi}}_1)\, e^{i\mathbf{k}_1\cdot\mathbf{r}-i\omega t} + (\mathbf{E}\cdot\hat{\boldsymbol{\chi}}_2)\, e^{i\mathbf{k}_2\cdot\mathbf{r}-i\omega t} \tag{12.12}$$

Each $\mathbf{k}_i$ corresponds to one of the two possible refraction indexes, defined as before

$$n_1 = \sqrt{1+\chi_1^1} \qquad n_2 = \sqrt{1+\chi_2^2}$$

Where χ_1^1 and χ_2^2 are the eigenvalues of χ_j^i.
We can also see that the two phase surfaces S_1, E_2 have *nonzero intersection*. Said $OP = S_1 \cap E_2$ the set of these intersections, if we define axes passing through the origin and these points, we see that $\mathbf{k}_1 = \mathbf{k}_2$ in these points. These axes are known as the *optical axes of the crystal*. Here, we also have $u_1 = u_2$ and therefore we have *no* birefringence.
In general, we can determine wether a crystal is birefringent or not by looking at the eigenvalues of the χ_j^i tensor.
With this categorization, we can define three kinds of crystals

1. Isotropic crystals, $\chi_1^1 = \chi_2^2 = \chi_3^3 = \chi$ and the crystal *is not* birefringent

$$\chi_j^i = \begin{pmatrix} \chi & 0 & 0 \\ 0 & \chi & 0 \\ 0 & 0 & \chi \end{pmatrix} \qquad n = \sqrt{1+\chi} \tag{12.13}$$

   Due to the non birefringent nature of isotropic crystals we have that there is only one optical axis. Cubic crystals fall into this category

2. Uniaxial crystals, here only two of the three eigenvalues are equal $\chi_1 = \chi_1^1 = \chi_2^2 \neq \chi_3^3 = \chi_2$ and the crystal exhibits birefringence.

$$\chi_j^i = \begin{pmatrix} \chi_1 & 0 & 0 \\ 0 & \chi_1 & 0 \\ 0 & 0 & \chi_2 \end{pmatrix} \qquad n_O = \sqrt{1+\chi_1}, \quad n_E = \sqrt{1+\chi_2} \tag{12.14}$$

   This kind of crystal has only one optical axis. The two possible refraction indexes are known as the *ordinary* refraction index n_O and the *extraordinary* refraction index n_E, due to the presence of only two refraction indexes, two subcategories of uniaxial crystals can be defined

   - Positive uniaxial crystals, with $n_E > n_O$
   - Negative uniaxial crystals, with $n_O > n_E$

   Trigonal, tetragonal and hexagonal crystals fall in this category.

3. Biaxial crystals, here all eigenvalues are different, and the crystal exhibits birefringence

$$\chi_j^i = \begin{pmatrix} \chi_1^1 & 0 & 0 \\ 0 & \chi_2^2 & 0 \\ 0 & 0 & \chi_3^3 \end{pmatrix} \qquad n_1 = \sqrt{1+\chi_1^1}, \quad n_2 = \sqrt{1+\chi_2^2}, \quad n_3 = \sqrt{1+\chi_3^3} \tag{12.15}$$

   This kind of crystal has two optical axes. Triclinic, monoclinic and orthorombic crystals fall into this category

In terms of phase surfaces, we have that isotropic crystals only have the sphere, biaxial crystals have the sphere and the ellipsoid together and the uniaxial crystal has the sphere and a *revolution ellipsoid*. in general, we have that for positive uniaxial crystals the sphere is contained inside the ellipsoid, while if for negative uniaxial crystals the ellipsoid is instead contained in the sphere.

§§ 12.1.2 Phase Velocity Surface

We can rephrase what we found before in terms of phase velocities. We know that by definition $k = \omega/u$, and $\mathbf{k} = \mathbf{u}\omega/u^2$, where the second is the vectorial counterpart to the previous statement. From this, it's possible to write the determinant (12.11) in terms of phase velocities

$$\det \begin{pmatrix} n_1^2 \frac{u^4}{c^2} - u_y^2 - u_z^2 & u_x u_y & u_x u_z \\ u_y u_x & n_2^2 \frac{u^4}{c^2} - u_x^2 - u_z^2 & u_y u_z \\ u_z u_x & u_z u_y & n_3^2 \frac{u^4}{c^2} - u_y^2 - u_x^2 \end{pmatrix} \tag{12.16}$$

Solving in a way completely analogous to what we found for the k-surfaces, we find as a solution circles and fourth degree ovals. For the xy planes the two equations define the two following surfaces

$$\begin{Bmatrix} u_x^2 + u_y^2 = \dfrac{c^2}{n_3^2} \end{Bmatrix} \\ \begin{Bmatrix} \dfrac{u_x^2}{n_3^2} + \dfrac{u_y^2}{n_1^2} = \dfrac{u^4}{c^2} \end{Bmatrix} \tag{12.17}$$

These two surfaces are *reciprocal* of the two k-surfaces and are known as the *phase velocity surfaces*

§§ 12.1.3 Ray Velocity Surface

Due to the anisotropic nature of crystals it's clear that $\mathbf{k}$ is not parallel to the Poynting vector $\mathbf{S} = \mathbf{H} \times \mathbf{E}$, since in general $\mathbf{k}$ is not parallel to $\mathbf{E}$.
In the case of a beam of light in a generic crystal we can still use the vector $\mathbf{k}$ in order to define the planes of constant phase, but its direction is not anymore parallel to the direction of propagation of the wave.
Calling θ the angle between $\mathbf{k}$ and $\mathbf{S}$, i.e.

$$\theta = \arccos\left(\frac{\mathbf{S} \cdot \mathbf{k}}{\|Sk\|}\right) \tag{12.18}$$

We can define the *ray velocity* $\mathbf{v}$ which will define the real direction of the ray.

$$v = \frac{u}{\cos\theta} \tag{12.19}$$

From this definition, we can see how it's possible to define a *ray velocity surface*.
We begin to rewrite the equations in terms of the electric displacement vector

$$D^i = \epsilon_0 \left(\delta_j^i + \chi_j^i\right) E^j$$

Substituting the definition into the wave equation we have

$$\epsilon^i{}_{jk} k^j \epsilon^k{}_{lm} k^l E^m = -\frac{\omega^2}{c^2 \epsilon_0} D^i$$

Using the properties of the Levi-Civita symbol we have

$$k^i \left(k^j E_j\right) - k^2 E^i = -\frac{\omega^2}{c^2 \epsilon_0} D^i$$

Since $k^i D_i = 0$ by definition of D^i, we can multiply both sides by D^i, getting[2]

$$k^2 \mathbf{D} \cdot \mathbf{E} = \frac{\omega^2}{c^2 \epsilon_0} D^2 \tag{12.20}$$

Using $u = \omega/k$ and that the vector $\mathbf{D}$ and $\mathbf{E}$ are separated by θ as we defined before, due to their tensorial relationship through χ^i_j. We can write then

$$\mathbf{D} \cdot \mathbf{E} = ED \cos\theta = \frac{u^2}{c^2 \epsilon_0} D^2 \tag{12.21}$$

Putting ourselves in the coordinate system of the crystal, i.e. we rotate into the eigenbasis of χ^i_j, we have that

1. $$\epsilon^i_j = 1 + \chi^i_j = \begin{pmatrix} n_1^2 & 0 & 0 \\ 0 & n_2^2 & 0 \\ 0 & 0 & n_3^2 \end{pmatrix}$$

2. $$D^i = \epsilon^i_j E^j \implies D^i = \epsilon_0 n^2_{(j)} E^i$$

Thus, the ray velocity surface can then be written in terms of $\mathbf{D}$ as the solution of the following set of equations

$$\begin{cases} \left(\dfrac{c^2}{n_1^2} - v_x^2 - v_z^2\right) D_x + v_x v_y D_y + v_x v_z D_z = 0 \\ v_y v_x D_x + \left(\dfrac{c^2}{n_2^2} - v_x^2 - v_z^2\right) D_y + v_y v_z D_z = 0 \\ v_z v_x D_z + v_z v_y D_y + \left(\dfrac{c^2}{n_3^2} - v_x^2 - v_y^2\right) D_z = 0 \end{cases} \tag{12.22}$$

As usual, the solution can be found as the roots of the determinant of the associated matrix. Taken the determinant for only one of the three possible planes thanks to symmetry relations, we find again the double surfaces

$$\begin{gathered} \left\{ v_x^2 + v_y^2 = \frac{c^2}{n_3^2} \right\} \\ \left\{ n_1^2 v_x^2 + n_2^2 v_y^2 = c^2 \right\} \end{gathered} \tag{12.23}$$

These surfaces are again a sphere and an ellipse, but these define the *ray axes* of the crystal, using again origin and intersection of the surfaces.

[2] For clarity I switched back again to the boldface vector notation

§ 12.2 Birefringence

We now treat mathematically the concept of birefringence. Consider a wave inside a birefringent crystal transmitting through one of the boundaries of the crystal. Due to having two possible values for k we must consider two transmitted waves, this ones are the waves that will create the double image that we see in the process of birefringence.
Called the incoming wave $(\mathbf{k}, \mathbf{E}, \mathbf{H})$ and the two transmitted waves $(\mathbf{k}_1, \mathbf{E}_1, \mathbf{H}_1)$ and $(\mathbf{k}_2, \mathbf{E}_2, \mathbf{H}_2)$ we have at the boundary

$$\begin{cases} k_2 \sin\phi_2 = k \sin\theta \\ k_1 \sin\phi_1 = k \sin\theta \end{cases}$$

This, as we saw already before, might look like a version of Snell's law for crystals. This is not the case, remember, since k_i depends directly on ϕ_i.
This changes if and only if we're treating an uniaxial crystal. Said ϕ_O and ϕ_E the ordinary and extraordinary transmission angles, noting that the ordinary wavevector $\mathbf{k}_O$ is by definition on the spherical phase surface, i.e.

$$\|\mathbf{k}_O\| = n_O^2 \frac{\omega^2}{c^2}$$

We have, just for the ordinary wave, again Snell's law

$$\sin\theta = n_O \sin\phi_O \tag{12.24}$$

This doesn't hold for $\mathbf{k}_E$, since it lays on the phase ellipsoid.
In general though, we can say that

- For positive uniaxial crystals $n_O < n_E$
$$\phi_E \leq \phi_O$$
- For negative uniaxial crystals $n_E < n_O$
$$\phi_E \geq \phi_O$$

This feature of uniaxial crystals can be used in order to create *polarizing prisms*. Consider the same wave as before, coming from inside an uniaxial crystal, and hitting the boundary. Suppose that the optical axis is perpendicular to the plane of incidence.
Here, if $\mathbf{E}_E$ is the extraordinary wave and $\mathbf{E}_O$ is the ordinary wave we have that they are respectively parallel and perpendicular to the optical axis.
We can distinguish even more these two crystals if we have either *positive* or *negative* uniaxial crystals. For negative crystals, where $\phi_E \geq \phi_O$ and $n_E \geq n_O$, we have that Snell holds for the ordinary angle, and

$$n_E < \frac{1}{\sin\theta} < n_O \tag{12.25}$$

This is the condition for total internal reflection for the ordinary wave! Therefore what happens, is that the ordinary wave gets totally reflected, while the extraordinary wave gets transmitted. Since the extraordinary wave is *parallel* to the optical axis, we get a *completely* polarized wave on the direction of the optical axis.

§ 12.3 Optical Activity

In general, birefringent material that acts on polarization via the rotation of the polarization plane, is known as *optically active*. If a polarized wave travels a path long l inside such media, its polarization plane will get rotated by an angle $\theta \propto l$.
It's possible to define a *specific rotatory power* δ for this kind of medium, which indicates the amount of rotation per unit length.
These objects are divided in two categories depending on the handedness of the rotation applied to the polarization plane:

- Levorotatory
- Dextrorotatory

This phenomena can be explained supposing that this media is anisotropic, and has a "right" and a "left" refraction indexes n_R and n_L. Using Jones' vectors, specifically in the circular polarization basis, we can say that a wave in such medium will can be written in terms of the following basis

$$\begin{aligned} |RCP\rangle &= \begin{pmatrix} 1 \\ -i \end{pmatrix} e^{ik_R z - i\omega t} \\ |LCP\rangle &= \begin{pmatrix} 1 \\ i \end{pmatrix} e^{ik_L z - i\omega t} \end{aligned} \tag{12.26}$$

Suppose that we shine into this material a linearly polarized wave $|k\rangle$, where

$$|k\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

In the previous basis this wave can be represented as follows

$$|k\rangle = \frac{1}{2}\left(|RCP\rangle + |LCP\rangle\right)$$

After traveling a distance l, a dephasement is introduced to both left and right components, precisely, the wave will be described as follows

$$|k(l)\rangle = \frac{1}{2}e^{ik_R l}|RCP\rangle + \frac{1}{2}e^{ik_L l}|LCP\rangle = \frac{1}{2}e^{\frac{1}{2}i(k_R+k_L)l}\left(e^{\frac{1}{2}i(k_R-k_L)l}|RCP\rangle + e^{-\frac{1}{2}i(k_R-k_L)l}|LCP\rangle\right) \tag{12.27}$$

Introducing the two following angles

$$\begin{cases} \psi = \dfrac{l}{2}\left(k_R + k_L\right) \\ \theta = \dfrac{l}{2}\left(k_R - k_L\right) \end{cases}$$

We have that

$$|k(l)\rangle = \frac{1}{2}e^{i\psi}\left(e^{i\theta}|RCP\rangle + e^{-i\theta}|LCP\rangle\right) = e^{i\psi}\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \tag{12.28}$$

Dropping the general phase at the beginning, we see that the end result is a linearly polarized wave that has been rotated by an angle θ.
Writing θ in terms of n_R, n_L we have

$$\theta = \frac{l\omega}{2c}(n_R - n_L) = \frac{\pi l}{\lambda_0}(n_R - n_L) \implies \delta = \frac{\pi}{\lambda_0}(n_R - n_L) \tag{12.29}$$

Where δ is again our specific rotatory power. Note that this will also depend implicitly on wavelength due to dispersion in the material.

§§ 12.3.1 Susceptibility Tensor of Optically Active Media

Given an optically active medium, it's clear that if E_x and E_y get rotated, the susceptibility tensor will be similar to a rotation matrix, i.e. with off-diagonal imaginary values. Precisely, such tensor will take the following matricial form

$$\chi_j^i = \begin{pmatrix} \chi_1^1 & i\chi_{12} & 0 \\ -i\chi_{12} & \chi_2^2 & 0 \\ 0 & 0 & \chi_3^3 \end{pmatrix} \tag{12.30}$$

For a wave traveling through the medium along the z axis, we have

$$\begin{cases} \left(\frac{\omega^2}{c^2} - k^2\right) E_x = -\frac{\omega^2}{c^2}\left(\chi_1^1 + i\chi_1^2 E_y\right) \\ \left(\frac{\omega^2}{c^2} - k^2\right) E_y = -\frac{\omega^2}{c^2}\left(\chi_2^2 E_y - i\chi_{12} E_x\right) \\ \frac{\omega^2}{c^2} E_z = -\frac{\omega^2}{c^2} E_z \end{cases} \tag{12.31}$$

The last equation gives the banal solution $E_z = 0$, while the other two can be solved by finding the roots of the following polynomial

$$\det \begin{pmatrix} \frac{\omega^2}{c^2}\left(1 + \chi_1^1\right) - k^2 & i\frac{\omega^2}{c^2}\chi_{12} \\ -i\frac{\omega^2}{c^2}\chi_{12} & \frac{\omega^2}{c^2}\left(1 + \chi_2^2\right) - k^2 \end{pmatrix} = 0 \tag{12.32}$$

The equation to solve is a biquadratic equation with k as a parameter

$$k^4 - k^2\frac{\omega^2}{c^2}\left(1 + \chi_2^2 + \chi_1^1\right) - \frac{\omega^4}{c^4}\left(\chi_{12}^2 - \left(1 + \chi_1^1\right)\left(1 + \chi_2^2\right)\right) = 0 \tag{12.33}$$

Which, has solutions for

$$k = \frac{\omega}{c}\sqrt{1 + \chi_1^1 \pm \chi_{12}} \tag{12.34}$$

Solving again for $\mathbf{E}$ we get

$$E_x = \pm i E_y \tag{12.35}$$

Where the sign depends on the chirality (handedness) of the polarization of the wave.
Remembering also that $\frac{\omega}{c}k = n$ we also have that the right and left refraction indexes are

$$\begin{cases} n_R = \sqrt{1 + \chi_1^1 + \chi_{12}} \\ n_L = \sqrt{1 + \chi_1^1 - \chi_{12}} \end{cases} \tag{12.36}$$

Thus, the specific rotatory power δ is

$$\delta = \frac{\pi}{\lambda_0}(n_R - n_L) \approx \frac{\chi_{12}\pi}{n_O \lambda_o} \tag{12.37}$$

§§§ 12.3.1.1 The Special Case of Quartz

A cool example of an object which is both birefringent and optically active is *quartz*. Solving the algebraic equation for finding the k-surfaces of this material, we find that the ellipsoid and the sphere have a null intersection. The separation between the surfaces depends directly on χ_{12}, therefore it becomes also a measure of the specific rotatory power

§ 12.4 Magneto-optic and Electro-optic Effects

§§ 12.4.1 Faraday Rotations

If an isotropic dielectric is immersed in a magnetic field, and a beam of *nearly* polarized light is sent through the material in the direction of the field, we can measure a rotation of the polarization plane of the wave, i.e. the magnetic field activates the dielectric.
This was first discovered in 1845 by Faraday, which saw that the amount of rotation of the polarization plane is proportional to the magnetic field intensity and the distance traveled in the medium.
Said V a proportionality constant, then we have

$$\theta = VBl \tag{12.38}$$

The constant V is commonly known as the *Verdet constant*.
The physical explanation of this effect comes from the application of the force equation to bound electrons.
Said B the magnetic field intensity, then we have that the force equation for the bound electrons is as follows

$$m\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} + K\mathbf{r} = -e\mathbf{E} - e\mathbf{v}\times\mathbf{B} \tag{12.39}$$

Again, using the similarity method and supposing an harmonic solution $\mathbf{r} \propto e^{i\omega t}$ we have

$$K\mathbf{r} - m\omega^2\mathbf{r} = i\omega e\mathbf{r}\times\mathbf{B} - e\mathbf{E} \tag{12.40}$$

Using again $\mathbf{P} = -Ne\mathbf{r}$ we have

$$\left(K - m\omega^2\right)\mathbf{P} = Ne^2\mathbf{E} - i\omega e\mathbf{P}\times\mathbf{B} \tag{12.41}$$

Rewriting $P^i = \epsilon_0\chi^i_j E^j$, we have that the susceptibility tensor must be similar to the tensor for optically active media. Writing for ease of notation $\omega_0 = \sqrt{K/m}$ the natural resonance frequency of the

dielectric and $\omega_c = eB/m$ the cyclotron frequency, the tensor will have the following components

$$\begin{cases} \chi_1^1 = \dfrac{Ne^2}{m\epsilon_0}\left(\dfrac{\omega_0^2-\omega^2}{\left(\omega_0^2-\omega^2\right)^2-\omega^2\omega_c^2}\right) \\ \chi_3^3 = \dfrac{NeP2}{m\epsilon_0}\left(\dfrac{1}{\omega_0-\omega^2}\right) \\ \chi_{12} = \dfrac{Ne^2}{m\epsilon_0}\left(\dfrac{\omega\omega_c}{\left(\omega_0^2-\omega^2\right)^2-\omega^2\omega_c^2}\right) \end{cases} \tag{12.42}$$

Although a dielectric becomes optically active when a $\mathbf{B}$ field is applied, the birefringent effects are minimal unless $\omega \approx \omega_0$, in what's known as the *Voigt effect*.

§§ 12.4.2 Kerr Effect

As we said before, when a magnetic field is applied to a dielectric, the medium becomes optically active, but not birefringent (unless the frequency is in the proximity of the optical resonance frequency).
Consider now an optically isotropic substance placed in a strong $\mathbf{E}$ field, it has been observed in 1875 by Kerr that the substance becomes birefringent. This effect is observed in both gases and liquids.
The main idea of explanation of this effect comes from the alignment of molecules along the direction of the electric field, making the material behave like an uniaxial crystal, where the optical axis is determined by the direction of the field.
The strength of the effect is proportional to E^2, where if we indicate $n_\parallel, n_\perp$ as respectively the refraction index in the direction parallel and perpendicular to the electric field, we have that

$$n_\parallel - n_\perp = KE^2\lambda_0 \tag{12.43}$$

The constant K is known as the *Kerr constant*.

§§ 12.4.3 Other Magneto-optic and Electro-optic Effects

Other effects of the same branch as these are

- Cotton Moutton effect: It's the magnetic analog of the Kerr effect in liquids. The alignment of the molecules is here given by the magnetic field $\mathbf{B}$, and we also have that the strength of the effect is proportional to the square of the field B^2.
- Pockels effect: For some birefringent crystals the indexes of refraction n_i are affected by the strengths of the fields E, B.

<++>

13 Lasers

§ 13.1 General Description

The laser, an abbreviation of *light amplification by means of stimulated emission of radiation*, are objects that are capable to emit coherent electromagnetic radiation up to frequencies of 10^9 Hz. Their behavior comes from the laws of quantum mechanics themselves and they were known already before the 1950s, these lasers emitted microwave radiation and therefore are known as *masers*.
The first laser was a He-Ne laser built at Bell laboratories. At around the same time ruby lasers were developed. Both emit coherent light in the red, while the He-Ne laser also emits light in the infrared spectrum.
Since these two lasers we managed to develop various kinds of lasers which emit at different frequencies.
In general lasers work as optical oscillators which have an amplifying medium inside the resonator, also known as laser cavity.
The amplification of light is achieved via means of external excitation. The oscillation in the cavity can be seen as a standing wave.
The final output is an intense beam of highly monochromatic light, which can be used in various fields.

§ 13.2 Stimulated Emission

The physical explanation of the inner workings of a laser starts from quantum mechanics itself.
Consider a quantum system with energy levels $n = 1, 2, 3, \cdots$ with associated energies E_n and level populations N_n. If the system is at thermal equilibrium at some temperature T we can evaluate the ratio of population of two levels using Boltzmann statistics.
Said $\beta = (k_B T)^{-1}$ and considered the first two levels we have

$$\frac{N_2}{N_1} = \frac{e^{-\beta E_2}}{e^{-\beta E_1}} = e^{-\beta(E_2 - E_1)} \tag{13.1}$$

Since we're working with photons we have $E_2 - E_1 = \hbar\omega$, therefore

$$\frac{N_2}{N_1} = e^{-\beta\hbar\omega} \tag{13.2}$$

Where ω_{12} is the difference of frequency between the second and the first level.
Said $\rho(\omega)$ the radiation density at a given frequency and B_{12}, B_{21} the stimulated emission and stimulated

absorption coefficients, we have that the rate of change per unit time of population of the first or the second level is

$$\begin{cases} \dfrac{\mathrm{d}N_1}{\mathrm{d}t} = N_1 B_{12} \rho(\omega) \\ \dfrac{\mathrm{d}N_2}{\mathrm{d}t} = N_2 B_{21} \rho(\omega) \end{cases} \tag{13.3}$$

Including also the spontaneous emissions from the levels we also add

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = N_2 A_{21} \tag{13.4}$$

At thermal equilibrium the total amount of population changes must equalize, i.e.

$$N_2 A_{21} + N_2 B_{21} \rho(\omega) = N_1 B_{12} \rho(\omega) \tag{13.5}$$

Solving for the radiation density we have

$$\rho(\omega) = \frac{A_{21}}{B_{21}} \frac{1}{\frac{N_1 B_{12}}{N_2 B_{21}} - 1} \tag{13.6}$$

Using (13.2) and noting that $\rho(\omega)$ must follow Planck's blackbody formula, thus

$$\rho(\omega) = \frac{\hbar \omega^3}{4\pi^3 c^2} \frac{1}{e^{\beta \hbar \omega} - 1} = \frac{A_{21}}{B_{21}} \frac{1}{\frac{B_{12}}{B_{21}} e^{\beta \hbar \omega} - 1} \implies \begin{cases} \dfrac{A_{21}}{B_{21}} = \dfrac{\hbar \omega^3}{4\pi^3 c^2} \\ B_{12} = B_{21} \end{cases} \tag{13.7}$$

Note that we can also define the ratio between spontaneous and stimulated emission is the following

$$\frac{SE}{E} = \frac{B_{21}}{A_{21}} \rho(\omega) = \frac{1}{e^{\beta \hbar \omega} - 1}$$

What it indicates is that the actual spontaneous emission is really small for visible light at ordinary temperatures for the sources $10^2 \mathrm{K} < T < 10^3 \mathrm{K}$.
This also indicates how actually the majority of radiation is actually spontaneous random emission, thus also incoherent. We will use light amplification in order to amplify the amount of coherent stimulated radiation.

§§ 13.2.1 Light Amplification

Consider a medium in which radiation passes through. Suppose that its atoms have random energy levels E_n. We will consider a two level system only in this case, where $E_1 < E_2$. The equations which will regulate spontaneous absorption and emissions are the usual (13.3).
Since $B_{21} = B_{12}$ we will have more stimulated emission if and only if the higher level E_2 is more populated than the lower level E_1. This process is known as *population inversion*.
If our radiation beam passes multiple times inside this medium we will get a steady gain of power of the beam thanks to the addition of stimulated emission of radiation.
The stimulated radiation will have the same phase dependency of the initial beam, and therefore the final result will be highly coherent.

§§§ 13.2.1.1 Gain

In order to quantify this energy gain, consider a beam which propagates in a medium where population inversion is possible. For a collimated beam the spectral radiation density $\rho(\omega)$ is tied to the irradiance $I(\omega)$ in the interval $[\omega, \omega + \Delta\omega]$ with the following relation

$$\rho(\omega)\Delta\omega = I_\omega \frac{\Delta\omega}{c} \tag{13.8}$$

Including in this the results we found for stimulated absorption and emission we get

$$B_{12}\rho(\omega)\Delta N_1 = \frac{B_{12}}{c} I(\omega)\Delta N_1 \tag{13.9a}$$

$$B_{21}\rho(\omega)\Delta N_2 = \frac{B_{21}}{c} I(\omega)\Delta\omega \tag{13.9b}$$

For each transition (13.9a) we get one quanta less of energy ($\Delta E = -\hbar\omega$), but we get plus one quanta for each (13.9b) transition.
Thus, for unit time we have

$$\frac{\partial\rho}{\partial\omega}\Delta\omega = \hbar\omega \left(B_{21}\Delta N_2 - B_{12}\Delta N_1\right)\rho(\omega) \tag{13.10}$$

But, since $c^{-1}\,\mathrm{d}t = \mathrm{d}x$, we have, using (13.8)

$$\frac{\partial I}{\partial x} = \frac{\hbar\omega}{c}\left(\frac{\Delta N_2}{\Delta\omega} - \frac{\Delta N_1}{\Delta\omega}\right) B_{21} I(\omega, x) \tag{13.11}$$

This result indicates that per unit length of travel inside the amplifying medium, we will get an increase of irradiance equal to

$$\frac{\partial \log(I)}{\partial x} = \frac{\hbar\omega}{c\Delta\omega} B_{21}\left(\Delta N_2 - \Delta N_1\right) \tag{13.12}$$

Via integration of (13.12) we have

$$I(\omega, x) = I_0 e^{\alpha_\omega x} \tag{13.13}$$

The constant α_ω is known as the *gain constant*, which is specific to the system. By definition

$$\alpha_\omega = \frac{\hbar\omega}{c\Delta\omega} B_{21}\left(\Delta N_2 - \Delta N_1\right) \tag{13.14}$$

Due to line broadening, this value is actually taken at the center of the spectral line. Said the spectral thickness $\Gamma \approx \Delta\omega$, this result is correct up to a numerical constant $c = \mathcal{O}(10^0)$. Approximating also $\Delta N_i \approx N_i$ we have that the maximum possible value the gain constant is

$$\alpha_{\omega,max} = \frac{\hbar\omega}{c\Delta\omega} B_{12}\left(N_2 - N_1\right) \tag{13.15}$$

Using the fact also that

$$\frac{A_{21}}{B_{12}} = \frac{\hbar\omega^3}{4\pi^3 c^2}$$

We have that

$$\alpha_{\omega,max} \approx \frac{\hbar\omega}{c\Delta\omega}\frac{4\pi^3c^2}{\hbar\omega^3}A_{21} = \frac{2\pi^3c}{\omega^2\Delta\omega}A_{21}\left(N_2 - N_1\right) \tag{13.16}$$

Or, in terms of wavelength

$$\alpha_{\omega,max} \approx \frac{2\pi^3c\lambda^2}{\Delta\omega}A_{21}\left(N_2 - N_1\right) \tag{13.17}$$

§§§ 13.2.1.2 Gain Curve

In order to determine the dependency between gain and the frequency of the wave we need to consider the effect of *line broadening*. Considering a nonzero temperature we have that atoms in thermal motion will have a Gaussian dependency of their velocities.
Considering the isotropy of the system we can consider only the motion on the x direction

$$P(v_x)\Delta v_x = \sqrt{\frac{m\beta}{2\pi}}e^{-\frac{1}{2}m\beta v_x^2}\Delta v_x \tag{13.18}$$

Due to Doppler effect, the atoms will absorb and emit photon radiation in this direction at a slightly different frequency ω than the natural resonant frequency ω_0, thus

$$\frac{\omega - \omega_0}{\omega_0} = \frac{v_x}{c} \tag{13.19}$$

Substituting it into the Boltzmann equation for velocity we have that atoms in a given energy state $|i\rangle$ will absorb and emit radiation with the following probability density function

$$P(\omega)\Delta\omega = \frac{c}{\omega_0}e^{-\frac{\beta mc^2}{2\omega_0^2}(\omega-\omega_0)^2}\Delta\omega \tag{13.20}$$

This implies then that

$$\Delta N_i = \sqrt{\frac{\beta mc^2}{2\pi}}N_i e^{-\frac{\beta mc^2}{2\omega_0^2}(\omega-\omega_0)^2}\frac{\Delta\omega}{\omega_0} \tag{13.21}$$

Considering again the previous simpler two level approximation, we have that after substituting the values for $\Delta N_1, \Delta N_2$

$$\alpha_\omega = \frac{\hbar\omega_0}{c\delta\omega}\sqrt{\frac{\beta mc^2}{2\pi}}(N_2 - N_1)e^{-\frac{\beta mc^2}{2\omega_0^2}(\omega-\omega_0)^2}\frac{B_{21}}{\omega_0}\Delta\omega$$

Simplifying we have

$$\alpha_\omega = \frac{\hbar B_{21}}{c}\sqrt{\frac{\beta mc^2}{2\pi}}\left(N_2 - N_1\right)e^{-\frac{\beta mc^2}{2\omega_0^2}(\omega-\omega_0)^2} \tag{13.22}$$

Clearly, the variation of the gain constant follows a Gaussian curve centered on the resonant frequency ω_0. Setting the exponential equal to $\frac{1}{2}$ we get the half width half maximum line width $\Gamma_{\frac{1}{2}}$, thus

$$e^{-\frac{\beta mc^2}{2\omega_0^2}(\omega-\omega_0)^2} = \frac{1}{2} \implies (\omega-\omega_0)^2 = \frac{2\omega_0^2\log(2)}{\beta mc^2}$$

Which, taken the square root gives $\Gamma_{\frac{1}{2}}$

$$\Gamma_{\frac{1}{2}} = \frac{1}{2}\Gamma = \omega_0 \sqrt{\frac{2\log(2)}{\beta m c^2}} \tag{13.23}$$

This implies that the maximum value is

$$\alpha_{\omega,max} = \frac{1}{2}\frac{\lambda_0}{\Delta\omega}\sqrt{\frac{\log(2)}{\pi}}\left(N_2 - N_1\right)A_{21} \tag{13.24}$$

§§ 13.2.2 Experimental Production of Population Inversions

- Photon excitation, also known as *optical pumping*
- Electron excitation
- Inelastic atomic collisions
- Chemical reactions

For optical pumping, an external light source is employed for producing a high population in non-ground state levels in the laser medium via selective absorption. This method is used in solid state lasers, like ruby lasers.
Direct electron excitation are used instead in gas lasers, like argon lasers. Here the laser medium carries itself the discharge current.
In inelastic atomic collisions the electric discharge is used in a way such that two different elements A, B transfer the excitation via collision. The general reaction of excitation transfer is as follows

$$A^\star + B \to A + B^\star$$

If $B^\star$ is metastable, it will have a population inversion and subsequent laser transition. The He-Ne laser uses this process, where the excited helium transfers the excitation to the neon atoms, which will undergo laser transition.
Chemical lasers instead use chemical reactions which leave a molecule or an atom in an excited state, like HF lasers, which use the following reaction

$$\mathrm{H_2} + \mathrm{F_2} \to 2\mathrm{HF}^\star \to 2\mathrm{HF} + \gamma \tag{13.25}$$

§ 13.3 Optical Resonators

§§ 13.3.1 Laser Oscillations

The optical cavity of a laser, known as the resonator, is composed by two mirrors that can be either plane or curved, where inside the amplifying the medium. If a sufficient population inversion is reached, the radiation gets amplified and establishes itself as a standing wave between the mirrors. The energy

is coupled to the resonator via partially transmitting mirrors.
The plane resonator works in a similar way to Fabry-Perot interferometers.
The emitted result is equally spread bands with a free spectral range of

$$\omega_{n+1} - \omega_n = \frac{\pi c}{d} \tag{13.26}$$

Where c is the speed of light and d is the distance between the mirrors.
The frequencies we get are known as *longitudinal and transverse modes* of the resonator. Oscillations can occur at one or more frequencies depending on the spacing of modes. Usually the standing waves oscillate at multiple modes. If high spectral purity is needed, it's possible to fine tune the laser parameters in order to get single mode oscillation. The inherent line width Γ is then determined by the so called *quality factor* Q of the resonator.
In general $\Gamma \propto 1$ Hz, but it's in practice possible to obtain $\Gamma \propto 10^3$ Hz, depending on thermal and mechanical stability of the system.

§§§ 13.3.1.1 Oscillation Threshold

In the amplifying medium, as we said before, we have

$$I(\omega, x) = I_{0,\omega} e^{\alpha_\omega x}$$

Suppose that we have a standing wave, as for a cavity. At every passage there will be a loss of energy due to scattering, reflections etc. In order to have laser oscillations, we need that the gain is higher than the loss. Said δ the amount of loss, we have

$$I(\omega, x) - I_{0,\omega} \geq \delta I(\omega)$$

If the cavity is long l, then, after a passage we have

$$I_{0,\omega} \left(e^{2\alpha_\omega l} - 1 \right) \geq \delta I(\omega, x)$$

If $2\alpha_\omega l << 1$ we can approximate by power series, which implies

$$2\alpha_\omega l \geq \delta \tag{13.27}$$

When the condition is satisfied the oscillation grows in irradiance till it gets into equilibrium with the loss. Note that δ is independent of the irradiation itself.
When the gain is at equilibrium with the oscillation, the system is said to incur in *hole burning*. Here the line profile looks like an inverted harmonic oscillator, known as the *Lorentz profile*. The line width of this profile is inversely proportional to the radiative lifetime of the element.
If $\Gamma_L \geq \Gamma$ all atoms are said to be in *communication with the oscillating laser mode*. This process is known as *homogeneous broadening*. If $\Gamma_L < \Gamma$ only a fraction of atoms participate in a given laser mode, and this process is known as *inhomogeneous broadening*

§§ 13.3.2 Resonator Stability

In order to talk about the stability of the oscillation, we have to firstly discuss about optical resonators themselves.

In general, the spatial modes of electromagnetic radiation in a closed cavity can be described by 3 integers, related to the standing wave pattern.
For lasers, the cavity is not closed. Here the resonator still supports a standing wave known as a *quasi-mode*. Part of the energy of the mode will spill around the mirror and is then loss via diffraction, the *diffraction loss* of the resonator is really important in low frequency lasers like He-Ne lasers, with a gain per pass of a few percent.
Consider now the wave itself which we will call ψ, and call the coordinates on the two mirrors (x, y) and (x', y'). If $\psi(x, y), \psi(x', y')$ are the complex amplitudes, then using KF theory we have

$$\psi'(x', y') = -\frac{ik}{4\pi} \iint_L \psi(x, y) \frac{e^{ikr}}{r} (1 + \cos\theta) \mathrm{d}^2 x \tag{13.28}$$

Where we indicated the cavity with L. Taken a point on the first mirror with respect to the second mirror, we have using trigonometry

$$\begin{cases} r = \sqrt{d^2 + (x' - x)^2 + (y' - y)^2} \\ \cos\theta = \dfrac{d}{r} \end{cases}$$

Where we indicated with d the distance between the centers of the two mirrors.
Considering a pair of identical mirrors, at equilibrium we must have $\psi' \propto \psi$, therefore

$$\gamma\psi(x', y') = \iint_L \psi(x, y) K(x, y, x', y') \mathrm{d}^2 x \tag{13.29}$$

This is what's known as an *integral equation* with eigenvalue γ and kernel K. With simple comparison, we have that

$$K(x, y, x', y') = -\frac{ik}{4\pi r} e^{ikr} (1 + \cos\theta)$$

Noting that $\gamma \in \mathbb{C}$, for each mode ψ_n, we will have an associated phase shift of $\arg(\gamma_n)$.
Since the irradiation of a mode can be evaluated as $\|\psi_n\|^2$, we can write the general diffraction energy loss formula

$$\delta_D = \frac{I_n - I_{loss}}{I_n} = 1 - \|\gamma_n\|^2 \tag{13.30}$$

Going back to the definition of the kernel, by comparison we can also find the Fraunhofer diffraction kernel as

$$K_F(x, y, x', y') = Ce^{-ik(xx' + yy')} \tag{13.31}$$

This is as usual, the kernel of a 2D Fourier transform. If we evaluate the integrals, in the most general case we will get a composition of Hermite polynomials and a Gaussian. The generic state will then be defined as

$$|TEM\rangle_{npq} = H_p\left(\frac{\sqrt{2}x}{w}\right) H_q\left(\frac{\sqrt{2}y}{w} e^{-\frac{x^2+y^2}{w}}\right) \tag{13.32}$$

Where w is a weight, p, q are the transverse mode numbers and n is the longitudinal mode number.
Generally, resonators are made by the following typologies of mirrors:

- Plane parallel mirrors

- Plane concave mirrors
- Confocal mirrors

The most common are the latter. Confocal resonators are the by far the easier to collimate, needing only 0.25 degrees of precision, while the others need a collimation precision of around 1 arcsec.
The diffractive loss of the system is evaluated in terms of the Fresnel number N_F, defined in terms of mirror curvature r and separation d as follows

$$N_F = \frac{r^2}{\lambda d} \tag{13.33}$$

A resonator is said to be *stable* only if $N_F > 1$, which means that after one reflection the beam stays collimated with the optic axis of the mirrors.
Note that the diffractive losses are negligible in a confocal resonator when $N_F > 1$, making this configuration the most efficient of the three.

§§§ 13.3.2.1 Spot Size

Let's go back a bit and check the weight parameter w we introduced before when talking about transverse modes.
We have that the e-folding distance of the radiation is exactly $\sqrt{x^2+y^2} = w$, therefore

$$\sqrt{x^2+y^2} = r \implies I = e^{-2}$$

The parameter w is then known as the *spot size* of the mode $|TEM\rangle_{0,0}$, which is the dominant oscillation mode. In general, it depends on the longitudinal distance from the midpoint z and the wavelength λ as follows

$$w^2(z) = w_0^2 + \left(\frac{\lambda z}{\pi w_0}\right)^2 \tag{13.34}$$

Here, w_0 is a parameter indicating the spot size at the center, and it depends on the distance of the two mirrors and their curvature R

$$w_0^2 = \frac{\lambda}{\pi}\sqrt{\frac{d}{2}\left(R - \frac{d}{2}\right)} \tag{13.35}$$

Also, we can define the radius of curvature of the wave in terms of the parameters of the resonator

$$r_c = z + \frac{d}{4z}(2R - d) \tag{13.36}$$

With these formulas, for a confocal resonator, where $R = d$, we have

$$\begin{cases} w_{0l} = \sqrt{\dfrac{\lambda d}{2\pi}} \\ w = \sqrt{\dfrac{\lambda d}{\pi}} \end{cases} \tag{13.37}$$

In this special configuration we have that, at the mirrors $r_c = R$ and w has its maximum, and at the center $r_c = 0$ and w reaches its minimum. Clearly, if we substitute symmetric wavefronts with mirrors with $r_c = R$ we get again a confocal resonator.

§ 13.4 Laser Typologies

§§ 13.4.1 Gas Lasers

In gas lasers, the optical cavity is created by two external mirrors, coated by multilayer films needed for getting high reflectivity R at the desired λ. These mirrors are in confocal configuration.
The gas chamber is fitted with two Brewster windows for obtaining the maximum transparency for highly p-polarized light. The external excitation can be provided by

- AC/DC electrodes
- Electrodeless HF discharges
- High voltage pulses

The simplest and most efficient way is through electrodes, with AC being the simplest, and DC the most advantageous for continuous lasers

§§§ 13.4.1.1 He-Ne Lasers

In He-Ne lasers, in order to get the laser transition, helium atoms are excited via electron collisions. The populations of the $\left|^3S\right\rangle, \left|^1S\right\rangle$ states then build up thanks to dipole selection rules, and, we have also that

$$E^{\mathrm{He}}_{^1S} = E^{\mathrm{Ne}}_{3s}$$
$$E^{\mathrm{He}}_{^1S} = E^{\mathrm{Ne}}_{2s}$$

Giving a high probability of energy transfer. Since also $E^{\mathrm{Ne}}_{3s} > E^{\mathrm{Ne}}_{3p}$ and $E^{\mathrm{Ne}}_{2s} > E^{\mathrm{Ne}}_{2p}$ with a sufficient buildup of population, an inversion is possible.
The optimal pressure for obtaining the laser transition is $p \approx 1$ atm, with a concentration ratio of helium and neon is of 7 moles of helium per mole of neon.
The laser transitions are

$$\begin{cases} |3s\rangle \to |2p\rangle & \lambda \approxeq 632.8 \text{ mm} \\ |3s\rangle \to |3p\rangle & \lambda \approxeq 339 \ \mu\text{m} \\ |2s\rangle \to |2p\rangle & \lambda \approxeq 1.1523 \ \mu\text{m} \end{cases} \tag{13.38}$$

The first one is the prevalent transition, and emits photons in red, while the other two emit IR photons.

§§§ 13.4.1.2 Other Gas Lasers

Electric discharges in pure gases and mixtures produce laser transitions at wavelengths between the far infrared and the ultraviolet parts of the spectrum.
All noble gases exhibit laser behavior, zinc, cadmium, mercury, lead and other metals exhibit laser behavior with pulsed discharges, together with the halogens.
It's also possible to build gaseous molecular lasers with N_2.

§§ 13.4.2 Optically Pumped Solid State Lasers

In solid state lasers, the optically active atoms are embedded in crystals or glasses. The solid is usually made as a cylindrical rod, polished and coated in order to behave like a resonator, or either provided with external mirrors.
The pumping is achieved with external light sources like high intensity lamps.

§§§ 13.4.2.1 Ruby Lasers

In ruby lasers specifically, the rod is made of synthetic sapphire doped with 0.05% dichromium trioxide. Here the chromium substitutes the aluminum in the lattice.
During pumping the chromium excites from the ground state A^4 to the states ${}^3T_1$ and ${}^4T_2$ which decay via *rapid radiationless transition RRLT* to the excited level ${}^2E$ from which a population inversion can happen with $\lambda = 6934$ Å.

§§ 13.4.3 Dye Lasers

Stimulated emission in liquids was first observed in 1966 at IBM labs using dye solutions pumped with a ruby laser first, and then with a fast flash lamp.
The organic compounds that were used are *fluorescine* and *rhodamine*, and the population inversion was obtained exciting the molecules to the fluorescence states. Due to the broad band of possible states, the frequency of dye lasers can be tuned with prisms, gratings and interferometers placed inside the cavity, permitting the construction of lasers at various frequencies.

§§ 13.4.4 Semiconductor Diode Lasers

The most compact laser possible is the *diode laser*, also known by the name of *injection laser*.
In the simplest configuration possible, the diode laser is composed by a P-N junction provided by a doped crystal like gallium arsenide. When a forward bias is applied to the diode, electrons are injected into the P side of the junction, and holes are formed in the N section.
The electron-hole interaction results in *recombination radiation*. This radiation, if the current is high enough, can produce a population inversion in the system. The laser action is then produced at the junction boundary. The action layer is quite small ($d \sim 1\ \mu m$), and the gallium arsenide crystals behave as partially reflective mirrors, creating a resonator cavity.
The emitted wavelength is between 830 and 850 nm

§ 13.5 Q-Switching and Mode Locking

In high power pumped lasers, the laser action begins when the population inversion transition density reaches the threshold where the gain exceeds the loss in the amplifying medium.
The inverse of the loss per cycle is known as the *resonant Q* of the cavity, hence by definition, a higher Q implies a lower density of population inversions.
It's possible to delay laser oscillations via *Q switching* in the cavity, using shutters known as *Q switches*. The Q switch is closed at the beginning of the pump cycle when the population inversion density is lower, and opened again when it's at the maximum.
Q switches can be created using the rotation of the mirrors around an axis perpendicular to the optical axis, electro-optic shutters created with Pockel cells and using saturable absorbents, dyes which become

transparent when strongly irradiated, creating a passive Q switch.
With Q switching it's possible to create gigawatt and terawatt laser pulses, which can then be used for high energy projects, such as laser fusion.
Another method of laser manipulation is *mode locking*. Given a nonlinear absorbent like a bleachable dye, placed inside the resonator cavity, it's possible to modify the laser output to short regular pulses. The time interval between pulses is a whole fraction of the cycle time, where for each pulse the radiation is bunched up into narrow packets that bounce back and forth in the cavity.
A Fourier analysis shows that the spectrum consists of discrete frequencies separated by exactly the pulse frequency.
In a round trip, the travel time is

$$\tau = 2\frac{d}{u} = 2\frac{nd}{c}$$

With d the mirror spacing. The frequency is then separated by

$$\nu_B = \frac{1}{\tau} = \frac{u}{2d} = \frac{c}{2nd}$$

Here, the resulting modes are coherent, and the laser is said to be *mode or phase locked*. The pulse line width is $\Gamma \propto \Delta\nu^{-1}$, and the total bandwidth is occupied by coherent modes. For N modes, we have that the pulse width is $\frac{1}{N}$ the pulse interval, creating very short multimode pulses.
In gas lasers these mode locked pulses are limited to a few ns pulses, whereas it's possible to have 3 ps pulses in dye or neodymium-glass lasers. These short pulses are thick around 1 mm, and can be considered as energy sheets traveling at the speed of light.

14 Geometric Optics

§ 14.1 Reflection and Refraction at a Spherical Surface

§§ 14.1.1 Mirrors

Virtually, every surface of an optical instrument can be bunched up in two categories

- Plane surfaces
- Curved surfaces

The path taken by light when interacting can be conveniently treated using ray theory, or geometric optics.
Consider two main cases now:

1. Reflection by a spherical surface
2. Refraction through a spherical surface separating 2 different mediums

Consider in both cases the origin of the ray as a point source P. Said O the intercept of the line that connects the mirror to the source and called Q the point where the beam intercepts the line OP as in figure.

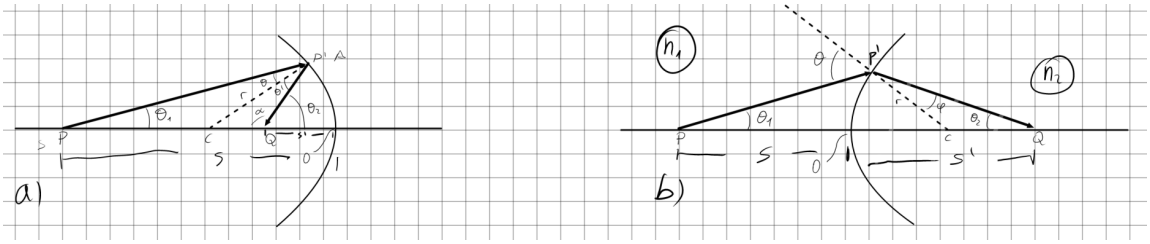


Figure 14.1: The two configurations described above

Call the distance $OP = s$ the *object distance*, $OQ = s'$ the *image distance* and the radius of curvature $OC = r$. We will have for the case of the spherical mirror that

$$PC = s - r \implies \frac{\sin\theta}{s-r} = \frac{\sin\theta_1}{r}$$

And

$$CQ = r - s' \implies \frac{\sin\theta'}{r - s'} = \frac{\sin\theta_2}{r}$$

Since we must have $\theta = \theta'$, we get the law of reflection for spherical mirrors

$$\frac{r - s'}{s - r} = \frac{\sin\theta_1}{\sin\theta_2} \tag{14.1}$$

Analogously for the refracting surface we have

$$r\sin\theta = (r + s)\sin\theta_1 \qquad r\sin\phi = (s' - r)\sin\theta_2$$

Applying $n_1 \sin\theta = n_2 \sin\phi$ we get

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{n_2}{n_1}\frac{s' - r}{r + s} \tag{14.2}$$

Where $n = \frac{n_2}{n_1}$.
Note how a bundle of rays originating from an axial point doesn't have the same focus. In fact the focus is function of θ_1. This characteristic is typical of spherical optical surfaces and leads to what's known as *spherical aberration*.

§§§ 14.1.1.1 General Mirror Equation and Focal Length

Consider now the case of the source point placed at a height h from the optical axis of the mirror. The image at the point Q will then be at a height h' from the same axis at the point Q as in the next figure.

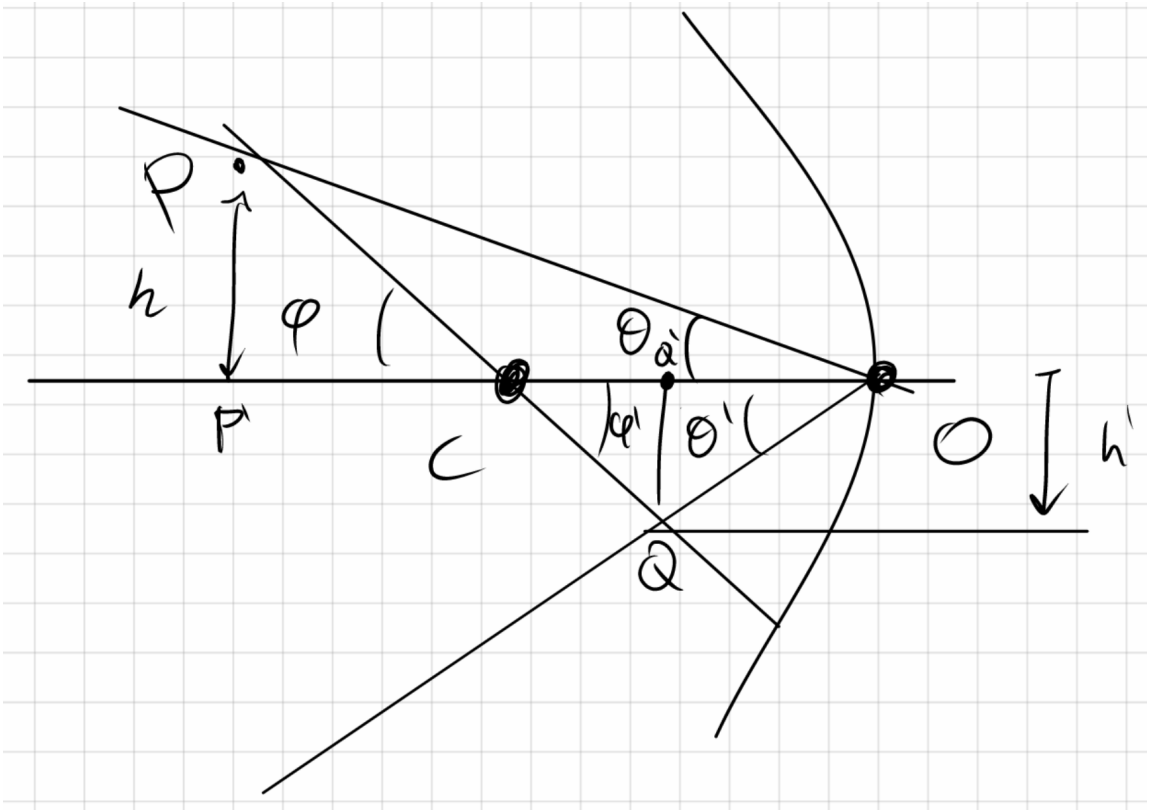


Figure 14.2: Source offset by h, spherical mirror

From the previous figure, we can see two triangles which we can use for finding the relationship between the position of the source and the position of the image. Precisely we choose the triangles $PP'C$, $PP'O$ and the triangles CQQ', $QQ'O$. Since $PP' = h$, $QQ' = h'$, $OP' = s$, $OQ' = s'$, $P'C = s - r$, $Q'C = r - s'$ we can say without doubts that

$$\begin{cases} \tan\theta = \dfrac{h}{s} \\ \tan\theta' = \dfrac{h'}{s'} \\ \tan\varphi = \dfrac{h}{s-r} \\ \tan\varphi' = \dfrac{h'}{r-s'} \end{cases} \tag{14.3}$$

Since it's also true that

$$\begin{cases} \tan(\theta) = -\tan(\theta') \\ \tan(\varphi) = -\tan(\varphi') \end{cases}$$

We get that

$$\begin{cases} \dfrac{h}{s} = -\dfrac{h'}{s'} \implies \dfrac{h}{h'} = -\dfrac{s}{s'} \\ \dfrac{h}{s-r} = -\dfrac{h'}{r-s'} \implies \dfrac{h}{h'} = -\dfrac{s-r}{r-s'} \end{cases}$$

Using the previous equalities we get

$$\frac{s}{s'} = \frac{s-r}{r-s'} \implies \frac{r-s'}{s'} = \frac{s-r}{r} = 1 - \frac{r}{s}$$

Fixing the relationship with algebra we get finally

$$\frac{r}{s} + \frac{r}{s'} = 2 \implies \frac{1}{s} + \frac{1}{s'} = \frac{2}{r} \tag{14.4}$$

This equation is *exact*, and is commonly known as the *general mirror equation*. The chosen signs are *merely a convention*.
We define the *focus* f of a mirror the convergence point of rays coming from a source placed at infinity. Imposing $s = \infty$ we then have

$$\frac{1}{s'} = \frac{2}{r} = \frac{1}{f} \tag{14.5}$$

In general, in geometric optics, for calculating the path of the rays in the method of *ray tracing of the first order*, the so called *paraxial approximation* is employed. In this approximation, the source is far enough to consider the approximation $\theta \approx \sin\theta \approx \tan\theta$. In this approximation, the general mirror equation (14.4), with the definition of focal distance becomes

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f} \tag{14.6}$$

For a spherical mirror, the focus is at the half radius point, as we have proven now in paraxial approximation.
With the previous formula, it's also possible to evaluate the magnification and orientation of the reflected image.

$$M = \frac{h'}{h} \tag{14.7}$$

Notice the sign here: if $M > 0$ both h, h' share the same sign and therefore the image is not flipped.
Another convention is given to the mirror radii. If the mirror is convex, as we defined before $r < 0$ and the reflected image is called *real*, while it's called virtual in the remaining case.
It's possible to redefine the magnification starting from the previous parameters and using the general mirror equation, getting

$$M = -\frac{f - s'}{f}$$

Which, after inserting the general formula becomes

$$M = -\frac{s'}{s} \tag{14.8}$$

In the second case of transmission through a convex lens, the solution is completely analogous, giving us the following mirror equation

$$\frac{n_1}{s} + \frac{n_2}{s'} = \frac{n_2 - n_1}{r} \tag{14.9}$$

A thing which is not really clear from the previous evaluations, is that in general $s' \in \mathbb{R}$, therefore there might be configurations where $s' < 0$ or also that $f < 0$. In general, the convention used is that $r > 0$ when the mirror is concave with respect to the source, and vice versa. Note how having a negative image distance in this framework simply means that the rays diverge.

§ 14.2 Lenses

§§ 14.2.1 Thin Lenses

Consider a simple lens with radii r_1, r_2 and negligible thickness d. If the lens is immersed in air ($n_1 = 1$) and made of glass (n_g), with the help of the previous equations that when the ray passes through the first and second curved surfaces we have that

$$\begin{cases} \frac{1}{s_1} + \frac{1}{s_1'} = \frac{n_g - 1}{r_1} \\ \frac{1}{s_2} + \frac{1}{s_2'} = \frac{1 - n_g}{r_2} \end{cases} \tag{14.10}$$

For the full lens the straightforward result is then

$$\frac{1}{s} + \frac{1}{s'} = (n_g - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) = \frac{1}{f} \tag{14.11}$$

The last formula is valid only in paraxial approximation, and it's known commonly as the *Lensmaker formula*.
With the derivation that we used it's clear that with multiple lenses in contact, the *total* focal distance can be calculated via inverse addition. Therefore, called f_e the effective focal lens of an optical system, we can say

$$\frac{1}{f_e} = \sum_{i=1}^{N} \frac{1}{f_i} \tag{14.12}$$

Note that if we introduce a separation l between two lenses, the equation should be modified accordingly, as follows

$$\frac{1}{f_e} = \frac{1}{f_1} + \frac{1}{f_2} + \frac{l}{f_1 f_2} \tag{14.13}$$

§§ 14.2.2 Thick Lenses

Consider now the case of lenses which have a non-negligible thickness d. Here, the object and image distance are measured from the position of two *principal planes* placed at distances

$$\begin{aligned} d_1 &= fd\left(\frac{1-n}{r_2}\right) \\ d_2 &= fd\left(\frac{1-n}{r_1}\right) \end{aligned} \tag{14.14}$$

The focal length of the lens is evaluated from the following equation

$$\frac{1}{f} = (n-1)\left[\frac{1}{r_1} + \frac{1}{r_2} - d\frac{(n-1)^2}{nr_1r_2}\right] \tag{14.15}$$

§§ 14.2.3 Chromatic and Spherical Aberration

Due to the presence of dispersion all previous formulas are *color dependent*. This kind of aberration is known as *chromatic aberration* and can be corrected by using multiple lenses with different dispersion relations, such combination is known as an *achromatic combination* and is obtained if and only if

$$\begin{cases} \delta_1 = \dfrac{1}{n_1-1}\dfrac{\mathrm{d}n_1}{\mathrm{d}\lambda} & \dfrac{1}{n_2-1}\dfrac{\mathrm{d}n_2}{\mathrm{d}\lambda} \\ f_1 = f\left(1-\dfrac{\delta_1}{\delta_2}\right) & f_2 = f\left(1-\dfrac{\delta_2}{\delta_1}\right) \end{cases} \tag{14.16}$$

Note that the process of achromatization of a lens is possible only in a finite interval of wavelengths. One other typical aberration for lenses is the *spherical aberration*. Here, we have that the focal distance also varies with the height of the rays with respect to the optical axis of the lens. Considering a paraxial ray hitting the lens at a height h from the optical axis, the variation of focal length Δf is

$$\Delta f = \frac{1}{2}h^2f^2\frac{n-1}{n^2}\left[\frac{1}{r_1^3} + \left(\frac{1}{f}+\frac{1}{r_2}\right)\left(\frac{n+1}{f}+\frac{1}{r_2}\right)\right] = \frac{1}{2}Kh^2 \tag{14.17}$$

It's also possible to find that the minimum possible variation is

$$\Delta f_{min} \quad \text{for} \quad \frac{r_1}{r_2} = \frac{n+4-2n^2}{n+2n^2} \tag{14.18}$$

§ 14.3 Ray Equations

Considep a set of paraxial rays traveling in the general direction of the optical axis of an object, which we will indicate with z. The position and direction of any ray can be determined with only two parameters

1. The distance from the optical axis ρ

2. The angle between the ray and the optical axis θ

In the particular case of a free traveling ray, starting from a position z_1 to a position $z_1 + d$, if at z_1 the ray is described by the parameters (ρ_1, θ_1), then, remembering that $\sin\theta \approx \tan\theta \approx \theta$

$$\begin{pmatrix}\rho_2\\ \theta_2\end{pmatrix} = \begin{pmatrix}\rho_1 + \theta_1 d\\ \theta_1\end{pmatrix} \tag{14.19}$$

If the ray passes through two different dielectrics, then at the boundary

$$\begin{pmatrix}\rho_2\\ \theta_2\end{pmatrix} = \begin{pmatrix}\rho_1\\ \frac{n_1}{n_2}\theta_1\end{pmatrix} \tag{14.20}$$

While, if the boundary is curved with curvature R

$$\begin{pmatrix}\rho_2\\ \theta_2\end{pmatrix} = \begin{pmatrix}\rho_1\\ \frac{n_1}{n_2}\theta_1 - \frac{\rho_1}{r}\left(1 - \frac{n_1}{n_2}\right)\end{pmatrix} \tag{14.21}$$

Note that for lenses and curved mirrors, using the general mirror equation, it reduces to the simpler equation

$$\begin{pmatrix}\rho_2\\ \theta_2\end{pmatrix} = \begin{pmatrix}\rho_1\\ \theta_1 - \frac{\rho_1}{f}\end{pmatrix} \tag{14.22}$$

These vectors should not be confused with Jones vectors, but as for those, every interaction can be evaluated with 2×2 matrices known as *ray matrices*. In general, we can define the following ray matrices for the previous interactions

$$\begin{aligned}
&\begin{pmatrix}1 & d\\ 0 & 1\end{pmatrix} && \text{free travel}\\
&\begin{pmatrix}1 & 0\\ 0 & \frac{n_1}{n_2}\end{pmatrix} && \text{plane dielectric}\\
&\begin{pmatrix}1 & 0\\ \frac{1}{R}\left(\frac{n_1}{n_2} - 1\right) & \frac{n_1}{n_2}\end{pmatrix} && \text{curved dielectric}\\
&\begin{pmatrix}1 & 0\\ -\frac{1}{f} & 1\end{pmatrix} && \text{thin lens / mirror}
\end{aligned} \tag{14.23}$$

Note that, as usual, if the system is composed by multiple smaller elements, like in the case of two lenses with focal lengths f_1, f_2, the total ray matrix of the system is

$$\begin{pmatrix}a & b\\ c & d\end{pmatrix} = \begin{pmatrix}1 & 0\\ -\frac{1}{f_1} & 1\end{pmatrix}\begin{pmatrix}1 & 0\\ -\frac{1}{f_2} & 1\end{pmatrix} = \begin{pmatrix}1 & 0\\ -\frac{1}{f_1} - \frac{1}{f_2} & 1\end{pmatrix} = \begin{pmatrix}1 & 0\\ -\frac{1}{f_e} & 1\end{pmatrix}$$

Note how we gained again the result we found before

§§ 14.3.1 Periodic Lenses and Resonators

Ray matrices are fundamental in the study of multiple lens systems and resonators. We'll begin considering a system of two lenses with equal focal length f, uniformly distant d between one another.

Considering the refraction of one lens and the subsequent free travel, we have that at the surface of the second lens the ray will be

$$\begin{pmatrix}\rho_2\\ \theta_2\end{pmatrix} = \begin{pmatrix}1 & 0\\ -\frac{1}{f} & 1\end{pmatrix}\begin{pmatrix}1 & d\\ 0 & 1\end{pmatrix}\begin{pmatrix}\rho_1\\ \theta_1\end{pmatrix} = \begin{pmatrix}1 & d\\ -\frac{1}{f} & 1-\frac{d}{f}\end{pmatrix}\begin{pmatrix}\rho_1\\ \theta_1\end{pmatrix} \tag{14.24}$$

Now consider a resonator, where the ray will pass multiple times through the system. One question we might ask is: *is there a ray which is invariant for the system described?*
The answer is yes, and it can be determined via the usage of the spectral theorem. The secular equation for the last matrix is

$$\det\begin{pmatrix}1-\lambda & d\\ -\frac{1}{f} & 1-\frac{d}{f}-\lambda\end{pmatrix} = (1-\lambda)\left(1-\frac{d}{f}-\lambda\right) = \lambda^2 + \lambda\left(2-\frac{\lambda}{f}\right)+1 = 0 \tag{14.25}$$

Said $\alpha = 1-\frac{d}{2f}$ we have that the equation simplifies notably, and gives the following result

$$\lambda_{1,2} = \begin{cases}-\alpha \pm \sqrt{\alpha^2-1} & |\alpha| > 1\\ -\alpha \pm i\sqrt{1-\alpha^2} = e^{\pm i\phi} & |\alpha| < 1\end{cases} \tag{14.26}$$

Note that $\alpha \in \mathbb{R}$ by definition, and therefore $\lambda \in \mathbb{C}$. For a general eigenray of the system, after N lenses, we have that

$$\begin{pmatrix}\rho_N\\ \theta_N\end{pmatrix} = \lambda^N \begin{pmatrix}\rho_1\\ \theta_1\end{pmatrix} \tag{14.27}$$

This system is the exact idea of system that we have in lasers. The idea of *ray stability* is then clearly the need to have the beam as close as possible to the optical axis after N passages. We thus prefer the case where $|\alpha| \leq 1$, such that $\lambda^N = e^{\pm iN\phi}$. This result is then

$$0 \leq 1-\frac{d}{2f} \leq 1 \implies \begin{cases}0 \leq d \leq 4f & \text{lenses}\\ 0 \leq d \leq 2r & \text{mirrors}\end{cases} \tag{14.28}$$

Note that we *always* need that $f \geq 0$, therefore we are stuck to using converging or flat lenses or mirrors.
In the confocal configuration of the resonator, as for laser chambers, we need to have the focuses of the lenses/mirrors to coincide, i.e. $d = 2f = R$, and the stability criterion is satisfied.
If the two lenses are different, then we will define two values of α, as

$$\alpha_1 = 1-\frac{s}{2f_1} \qquad \alpha_2 = 1-\frac{d}{2f_2}$$

And the stability condition becomes

$$0 < \alpha_1\alpha_2 < 1 \tag{14.29}$$

Part IV

Circuit Analysis

15 Direct Current Circuits

§ 15.1 Basic Components and Laws

§§ 15.1.1 Resistors, Capacitors and Inductors

An electric circuit is an interconnection of components in a closed loop of conductors, where current flows.
The simplest possible circuit is composed by a simple dcvsource in a closed loop of cables. Circuits are described graphically via technical diagram as the following one for a simple dcvsource in a closed loop

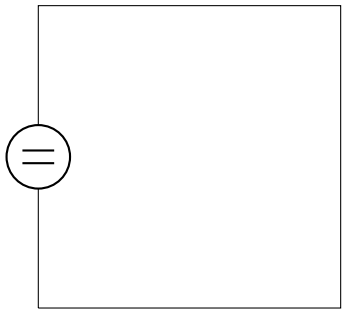


Figure 15.1: Simple closed loop dcvsource

Other components which we will treat in this chapter are resistors R, capacitors C and inductors (aka coils) L. They are described by the following drawings

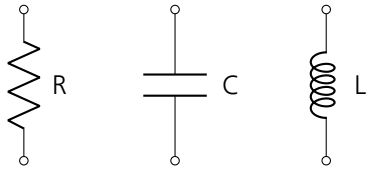


Figure 15.2: From right to left: a resistance, a capacitor and an inductor

These components can be inserted in circuits in various ways, creating different effects which will

be studied using the basic laws that we can obtain from classical electromagnetism

§§ 15.1.2 Ohm's Law

Ohm's law is one of the basic laws of circuit analysis. The derivation is simple enough, and has already been treated before, but just as a refresher it will be re-derived here and then rewritten in more familiar terms.
A metal can be summarized in electromagnetism as a sea of free electrons roaming freely on its surface, moving at speed v. These electrons will collide with each other, and the average distance of free travel before a collision (aka *mean free path*) is defined as follows

$$\lambda = \tau v$$

Where the quantity τ has dimensions of time, and is known as the *relaxation time* of the metal.
Using Coulomb's force law together with Newton's second law we get that the acceleration a is

$$a = -\frac{e}{m}E$$

And therefore, the average velocity is

$$\langle v \rangle = a\tau = -\frac{e\tau}{m}E$$

Also, by definition we have that the total current density is $-Nev$, i.e.

$$J = \frac{Ne^2\tau}{m}E$$

Which gives us the static version of Ohm's law

$$\mathbf{J} = \sigma\mathbf{E}, \qquad \sigma = \frac{Ne^2\tau}{m} \tag{15.1}$$

With σ being the conductivity of the metal.
For an homogeneous linear metallic cable with section S we can say without much doubt that the conductivity σ doesn't depend on the electric field, and with a simple calculation, we can say that, if the cable is long l

$$V = \int_l \mathbf{E}\cdot\hat{\mathbf{t}}\mathrm{d}l = El, \qquad I = \iint_S \mathbf{J}\cdot\hat{\mathbf{n}}\mathrm{d}^2s = JS$$

Substituting we get

$$I = \frac{\sigma S}{l}V$$

Definition 15.1.1 (Resistance)**.** We define the resistance R of a conductor the amount of resistivity per surface times its length, i.e.

$$R = \frac{l}{\sigma S} = \frac{\rho}{S}l$$

The resistance is measured in *ohms* Ω

The previous definition gives us the well known Ohm's law for circuit analysis

$$V = RI \tag{15.2}$$

The electronic component which is tied to this quantity is the *resistor*. Resistors with a resistance values comprised between $0.01\ \Omega$ and $10^{12}\ \Omega$ can be found commercially, with power ratings between $1/8$ W and 250 W with a precision in between 0.005% and 20%.
The cheapest resistors are made using a carbon mix, and they guarantee a resistance $\pm 5\%$ from the nominal value.
Note that in real world applications, external factors can influence the resistance of components and circuits, changing the final, actual, resistance value. Here's a short list:

- Welding induces a permanent change in resistance values of around 2% of the nominal value
- Vibrations (2G) and shocks (100G) induce permanent changes of around 2% of the nominal value
- Humidity (95% at 40 degrees C) induce a temporary change of the value between 6 and 10%
- Temperature changes from:
  - 25 to -15 degrees C change real resistance values from 2.5% deviation to 4.5% deviation
  - 25 to 85 degrees C induce changes between 3.3% to 5.9%

Other defects in resistors which change the resistance value are parasitic inductances at high frequencies, noise at low frequencies and defects in the mix.
A higher priced but more precise kind of resistors are metallic film resistors, which guarantee tolerances of around 1% and oscillations around the nominal value of less than 0.1%.

§§ 15.1.3 Ideal Generators

An ideal generator is an object which provides a voltage V or a current I to the circuit. Ideal generators are generators for which their resistance is zero.
There are two types of generator, as it was hinted before: voltage and current generators.
A voltage generator is an energy source which keeps the voltage V constant between its poles, independently from the load.
A current generator is slightly more complex. It varies the voltage between its poles in order to keep the current I constant. Diagrams for both generators follow

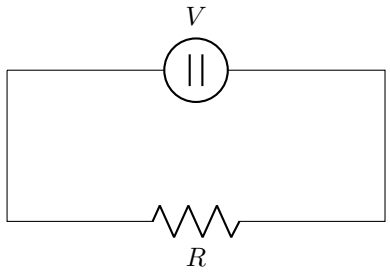


Figure 15.3: Ideal voltage generator connected to a resistance

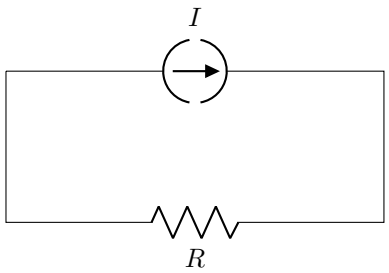


Figure 15.4: Ideal current generator connected to a resistance

§ 15.2 Simple Circuits

We are now ready to attack the first and simplest examples of circuits in the realm of circuit analysis. The simplest possible circuit is one composed from resistances connected in series to an ideal generator. Consider this simple circuit

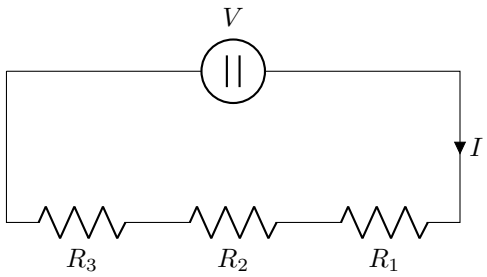


Figure 15.5: Simple circuit composed by an ideal voltage generator connected in series to three resistors R_1, R_2, R_3

We have using Ohm's law that the voltage drop at each resistor is exactly

$$V_i = R_i I$$

And since the voltage at each side of the generator is equal to V, we must have

$$V = \sum_i V_i = I \sum_i R_i = R_s$$

The value R_s is the total resistance of the circuit and it's therefore known as the *equivalent resistance* of the circuit, for a series of resistors.
We now see the same problem, but for a circuit composed by a voltage generator and three resistances connected in parallel, as in the following scheme

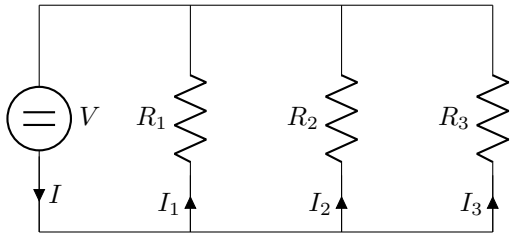


Figure 15.6: Circuit with three resistors connected in parallel to a voltage source

From Ohm's law we deduce that the voltage must be the same and equal to V for each resistor, thus, using again Ohm's law we can find the unknown current drops on each resistor. We have

$$I_i = \frac{V}{R_i}$$

And, since the total current is I and must remain constant, we get

$$I = \sum_i I_i = V \sum_i \frac{1}{R_i}$$

We define the equivalent resistance for parallel resistors then as follows

$$\frac{1}{R_p} = \sum_i \frac{1}{R_i} = \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \right) = \frac{R_1 + R_2 + R_3}{R_1 R_2 R_3}$$

Thus, finally

$$V = R_p I = \frac{R_1 R_2 R_3}{R_1 + R_2 + R_3} I \tag{15.3}$$

We consider now a more complicate case of a circuit with resistors connected both in parallel and in series, like the one in the next figure

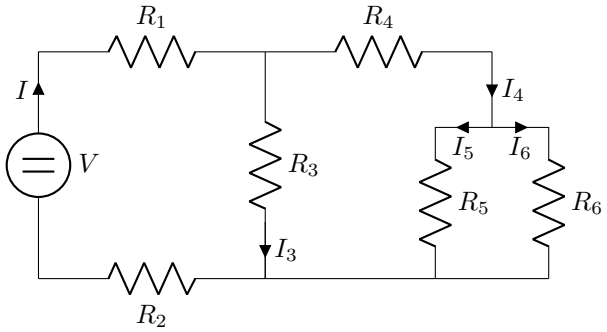


Figure 15.7: Complex circuit with multiple resistors connected both in parallel and in series

This circuit is also known as a *network*, and in order to simplify it we use the formulas we defined before.

- The current passing through the resistor R_1 and R_2 must be the same
- The voltage between R_5 and R_6 is the same

We define then the two following equivalent resistance, reducing the problem

$$R_{s1} = R_1 + R_2 \qquad \frac{1}{R_{p1}} = \frac{1}{R_5} + \frac{1}{R_6}$$

The reduced intermediate circuit is then

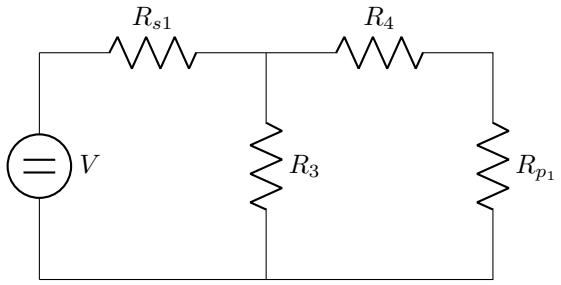


Figure 15.8: The previous circuit after the first iteration of simplification

We proceed the analysis by simplifying further this circuit, noting that

- R_{p1} and R_4 are connected in series
- The series between R_4 and R_{p1} is connected in parallel to R_3

Thus

$$R_{s2} = R_4 + R_{p1} = R_4 + \frac{R_5 R_6}{R_5 + R_6}$$

And, therefore

$$R_{p2} = \left(\frac{1}{R_{s2}} + \frac{1}{R_3}\right)^{-1} = \frac{R_{s2} R_3}{R_{s2} + R_3} = R_3 \frac{R_4 + \frac{R_5 R_6}{R_5 + R_6}}{R_3 + R_4 + \frac{R_5 R_6}{R_5 + R_6}}$$

The circuit now can then be simplified to this one

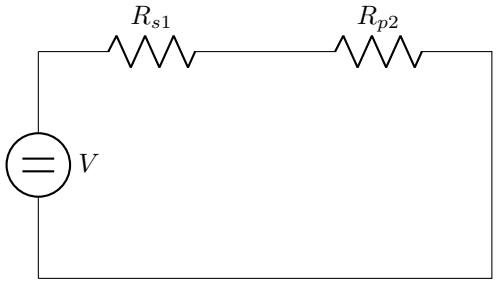


Figure 15.9: Second iteration of the simplification of the circuit

Finally, the total resistance R_T is simply the series sum of the previous two, giving the following result

$$R_T = R_{s1} + R_{p2} = R_1 + R_2 + R_3 \frac{R_4 + \frac{R_5 R_6}{R_5 + R_6}}{R_3 + R_4 + \frac{R_5 R_6}{R_5 + R_6}}$$

This is the total or equivalent resistance of the circuit, i.e. if we reconstruct the circuit as only the generator and a single resistance with value R_T, we will get a completely equivalent circuit.

§§ 15.2.1 Kirchhoff Laws

The previous considerations work almost every time, but they tend to end up in tedious and long calculations. Here the geometry and topology of the circuit come in handy. We give two main definitions

Definition 15.2.1 (Node). A circuit node is a point in which multiple conductors converge

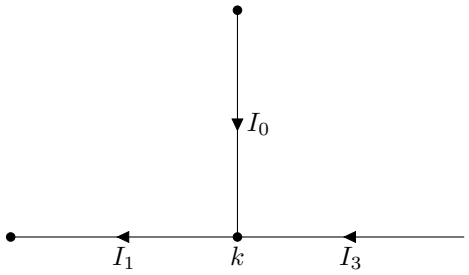
Definition 15.2.2 (Mesh). A mesh is a closed path starting and ending in one single node

With the previous two definitions, every circuit, independently from its complexity, can be considered as the union of multiple nodes and meshes. Two theorems come in help for our calculations

Theorem 15.1 (First Kirchhoff Law). *The sum of currents entering and exiting each node is zero, in formulas, for a node k, we have if we indicate positive currents the one entering the node and with negative currents the ones exiting the node, we have*

$$\sum_i I_i^{(k)} = 0 \tag{15.4}$$

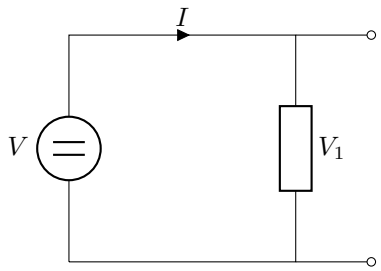
Consider the following part of a random circuit



Here, from Kirchhoff's first law, we get that

$$I_0 + I_3 - I_1 = 0$$

Theorem 15.2 (Second Kirchhoff Law). *The sum of voltages in every mesh is equal to zero. This is the direct consequence of Ohm's law applied to the first theorem. Consider the following simple circuit:*



From Ohm's law we have that $V_1 = RI_i$*, thus*

$$\sum V_i = 0$$

Consider now a real example, like the following circuit

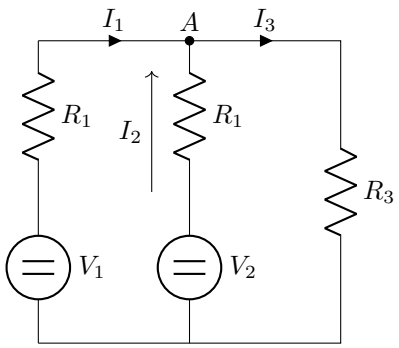


Figure 15.10: A slightly more complex circuit with multiple voltage sources

This circuit is quite complex to solve with only Ohm's law, but it's directly solvable using Kirchhoff laws. Firstly we choose what's the "positive" direction of current, which we choose it as the clockwise direction.
On the mesh on the left we must have, using the second Kirchhoff law

$$V_1 - V_2 = R_1 I_1 - R_2 I_2$$

And at the node A, we have

$$I_1 + I_2 - I_3 = 0$$

At the second mesh on the right, instead we have

$$V_2 = I_2 R_2 + I_3 R_3$$

We now have a system we can solve, which is the following

$$\begin{cases} V_1 - V_2 = R_1 I_1 - R_2 I_2 \\ I_1 + I_2 - I_3 = 0 \\ V_2 = I_2 R_2 + I_3 R_3 \end{cases}$$

Solving the second equation for I_3 and inserting it in the first equation we have

$$\begin{cases} V_2 = I_2 R_2 + (I_1 + I_2) R_3 = (R_3 + R_2) I_2 + R_3 I_1 \\ V_1 - V_2 = R_1 I_1 - R_2 I_2 \end{cases}$$

Substituting the first equation into the second one we get then

$$V_1 - V_2 = R_1 I_1 - \frac{R_2}{R_2 + R_3} (V_2 - I_1 R_3)$$

After some juggling and rearrangement we can get to the final result, which we will omit

§ 15.3 Special Configurations

§§ 15.3.1 Voltage Dividers

Consider the following circuit, known as a *voltage divider*

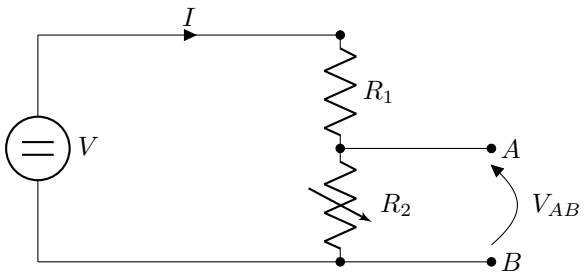


Figure 15.11: A voltage divider with a *variable resistor* R_2

In this circuit we want to know the voltage at the ends of R_2, i.e. V_{AB}. We can calculate it in two ways. Using Ohm's law we must have

$$V = V_1 + V_2 = R_1 I + R_2 I$$

Thus

$$I = \frac{V}{R_1 + R_2} \implies V_{AB} = R_2 I_2 = \frac{R_2}{R_1 + R_2} V$$

Or using Kirchhoff's laws, we have

$$\begin{cases} I_1 - I_2 = 0 \\ V = I_1 R_1 + I_2 R_2 \end{cases}$$

Solving we get the exact previous result.
Note how with the previous result, the value of V_{AB} is deeply tied to the value of the resistance R_2, thus we can reduce the outgoing voltage by changing the resistor. Note that V_{AB} is $50\%\ V$ only when $r_1 = R_2$

§§ 15.3.2 Wheatstone Bridge

A particular case of the voltage divider is known as the *Wheatstone bridge*

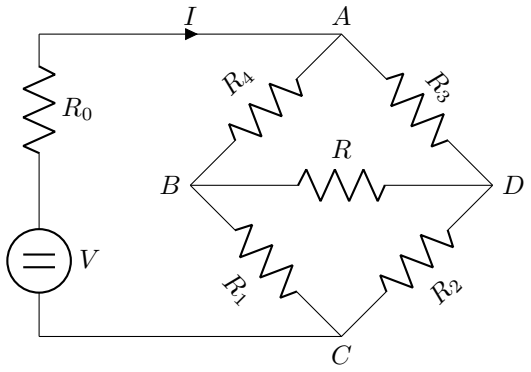


Figure 15.12: Diagram of a Wheatstone bridge

The idea of this circuit is to regulate the current flow between the nodes B and D. The shape of the circuit makes it seem more complex than it actually is, but it's possible to redraw it as two voltages dividers as follows

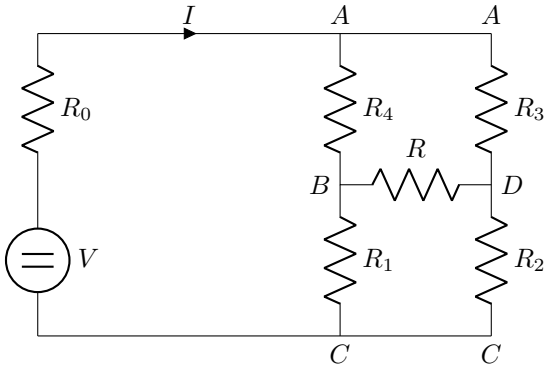


Figure 15.13: Redrawing of the Wheatstone bridge as two voltage dividers connected in parallel

The final objective of this bridge is to balance resistances in a way such that $I_{BD} = 0$. From the previous diagram, where the circuit has been redrawn as a voltage dividers gives us the answer to this problem, without making systems using Kirchhoff's law.

We have

$$\begin{cases} V_{BC} = R_1 I_1 = \dfrac{R_1}{R_1 + R_4} V \\ V_{DC} = R_2 I_2 = \dfrac{R_2}{R_2 + R_3} V \\ I_{BD} = \dfrac{V_{BD}}{R} \\ V_{BD} = V_{BC} - V_{DC} \end{cases}$$

Thus

$$I_{BD} = \frac{V}{R} \left(\frac{R_1}{R_1 + R_4} - \frac{R_2}{R_2 + R_3} \right)$$

Solving for $I_{BD} = 0$ we have

$$R_1 R_3 = R_2 R_4 \tag{15.5}$$

When this condition is satisfied, the bridge is said to be *balanced*.
The Wheatstone bridge can be used to measure the resistance value of a resistor using a variable resistor ($R(x)$) and three equal resistances on the bridge, the setup is described with the following diagram

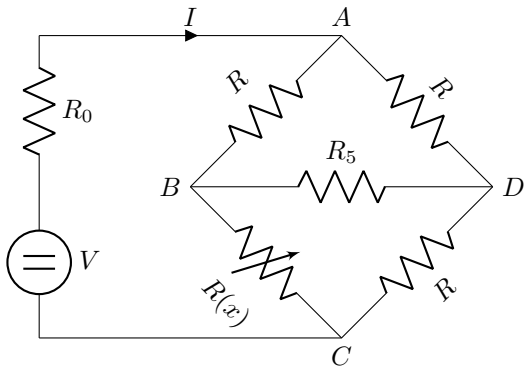


Figure 15.14: Diagram of a Wheatstone bridge in a setup which we can use for evaluating the resistance value

We have to make sure that the resistance $R_5 >> R$ in order to be sure that there is basically no current flowing through the poles B, D, so that

$$V_{BD}(x) = V \left(\frac{R(x)}{R(x) + R} - \frac{1}{2} \right) \approx \frac{V}{4R} \left(R(x) - R \right)$$

We will easily find the resistance value by variating $R(x)$ until $R(x_0) = R$ and thus the bridge is balanced and $V_{BD} = 0$. Clearly this evaluation is really sensible, and permits a very precise evaluation of the value of the unknown resistance R

§§ 15.3.3 Superposition Principle of Circuits

Another very useful principle for evaluating circuits is the *superposition principle*. It states simply that for circuits with more than one source, the total current flow I is equal to the sum of the currents of

the circuits with only one source. This is a direct derivation from the Kirchhoff theorems. Take as an example the circuit (15.10), it can be divided in these two circuits

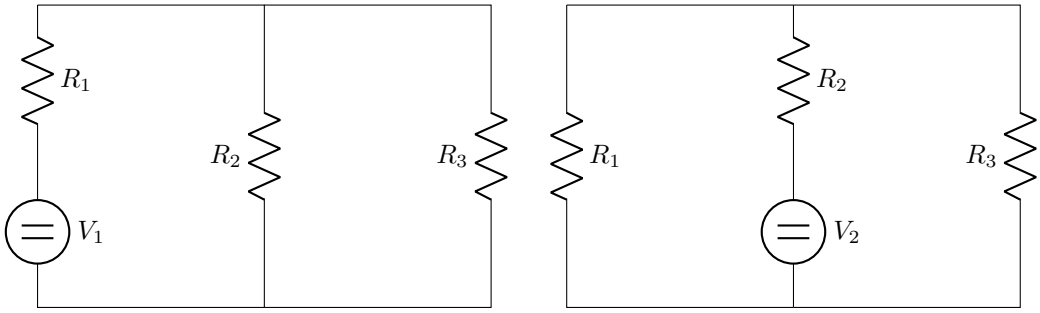


Figure 15.15: The two circuits which summed give back the circuit (15.10)

If we call the current passing through R_3 I_3, we must have

$$I_3 = I_{3a} + I_{3b}$$

Calling the first circuit a and the second b, we can use Kirchhoff to find

$$\begin{cases} V_2 = V_3 \qquad I_2 R_2 = I_{3a} R_3 \\ I_1 = \dfrac{V_1}{R_{eq}} \\ I_1 - I_2 - I_{3a} = 0 \end{cases}$$

The equivalent resistance R_{eq} is simply the one obtained by R_2, R_3 connected in parallel and the resistor R_1 connected in series to the two parallel resistors, thus

$$R_{eq,a} = R_1 + \frac{R_2 R_3}{R_2 + R_3}$$

Noting that the second circuit is equal to the first but R_1 is switched to R_2 we have

$$R_{eq,b} = R_2 + \frac{R_1 R_3}{R_1 + R_3}$$

Solving for I_3 for either of the two gives

$$\begin{cases} I_{3a} = \dfrac{R_2}{R_3} I_2 \\ I_1 = \dfrac{V_1}{R_{eq,a}} \\ \dfrac{V_1}{R_{eq,a}} = \left(1 + \dfrac{R_2}{R_3}\right) I_2 \end{cases}$$

$$I_{3a} = \frac{R_2 V_1}{R_2 + R_3} \frac{1}{R_1 + \frac{R_2 R_3}{R_2 + R_3}}$$

And therefore, with the previous trick, we have

$$I_3 = I_{3a} + I_{3b} = \frac{R_2 V_1}{R_2 + R_3} \frac{R_2 + R_3}{R_1 (R_2 + R_3) + R_3 R_2} + \frac{R_1 V_2}{R_2 + R_3} \frac{R_1 + R_3}{R_2 (R_1 + R_3) + R_1 R_3}$$

§ 15.4 Thevenin and Norton Theorems

§§ 15.4.1 Thevenin Theorem

All previous considerations can be synthesized into two theorems that can be used to reduce complex resistor-source networks to simple single source - single resistor circuits. The first theorem is the *Thevenin theorem*

Theorem 15.3 (Thevenin theorem). *Given an electrical network containing only voltage sources, current sources and resistors can be replaced at its terminals by an equivalent ideal voltage source V_{Th} connected in series to a resistance R_{Th}.*

- *The equivalent voltage V_{Th} is the voltage between the short circuited terminals*
- *The equivalent resistance R_{Th} is obtained evaluating the equivalent resistance when*
  - *All ideal voltage sources are substituted by a short circuit (cable)*
  - *All ideal current sources are replaced by an open circuit*
- *The current flowing between the two chosen terminals will be simply calculated by using Ohm's law*

Let's take as an example this circuit with two sources and evaluate the Thevenin equivalent circuit at the terminals A, B

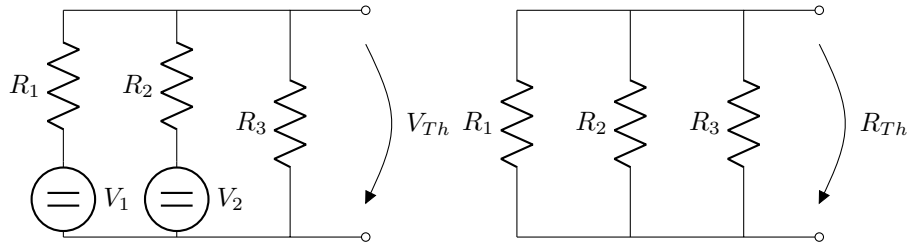


Figure 15.16: Example circuit, on the right with two voltage sources (for finding V_{Th}) and on the right with the two sources shorted in order to find R_{Th}

The solution is trivial and has already been discussed, thus we gloss over to the general application of the theorem. Consider an unknown network with two open dipoles, thanks to this theorem we can *always* say that its behavior at the dipoles can always be approximated to a single voltage source V_{Th}

connected in series with a resistor R_{Th}, and defined the impedance $G = R^{-1}$ we can use Kirchhoff for finding

$$\begin{cases} I_{Th} = \sum_i G_i V_i - GV = \dfrac{V_{Th}}{R_{Th}} \\ V_{Th} = \sum_i \dfrac{G_i}{G} V_i - \dfrac{I_{Th}}{G} \end{cases} \tag{15.6}$$

§§§ 15.4.1.1 Voltage Divider with Thevenin

We can see the power of this theorem directly by re-evaluating the voltage divider in the following configuration

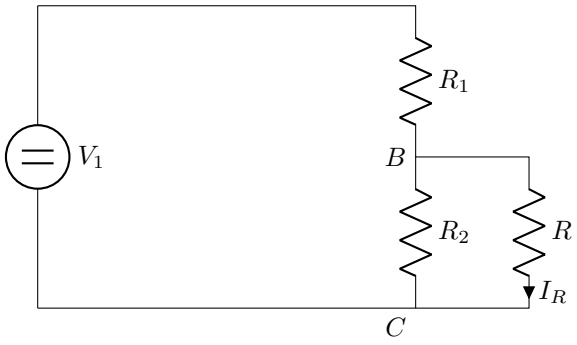


Figure 15.17: Voltage divider with a connected resistor R

We want to find I_R. How do we proceed?
We begin by shorting V_1 and opening the circuit at the connection with R, thus obtaining the following circuit

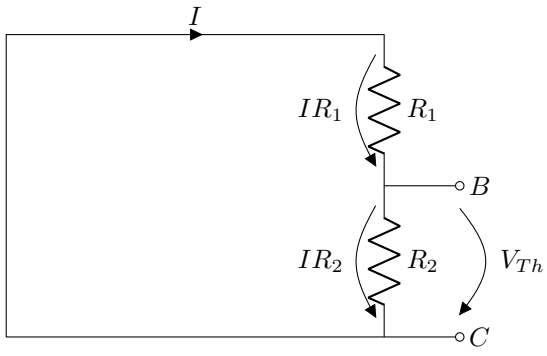


Figure 15.18: First application of the Thevenin theorem to the previous circuit

We have

$$I = \frac{V_1}{R_1 + R_2} \implies V_{BC} = IR_2 = \frac{R_2}{R_1 + R_2} V_1 = V_{Th}$$

From the same circuit, we have

$$R_{Th} = \left(\frac{1}{R_1} + \frac{1}{R_2} \right)^{-1} = \frac{R_1 R_2}{R_1 + R_2}$$

Thus, having found the Thevenin equivalent voltage and the Thevenin equivalent resistance that

$$I_R = \frac{V_{Th}}{R + R_{Th}} = \frac{R_2 V_1}{RR_1 + RR_2 + R_1 R_2}$$

§§§ 15.4.1.2 Wheatstone Bridge with Thevenin

Another good example for showing the power of the Thevenin theorem is solving the Wheatstone bridge using it. The circuit we are looking at is the following

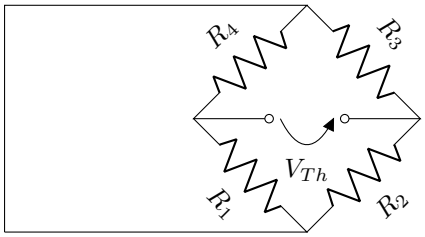


Figure 15.19: Thevenin theorem application to the Wheatstone bridge

As before, this circuit if divided in two is simply two voltage dividers, where

$$\begin{cases} V_{th,1} = \dfrac{R_1 V}{R_1 + R_4} \\ V_{th,2} = \dfrac{R_2 V}{R_2 + R_3} \\ R_{th,1} = \dfrac{R_3 R_1}{R_1 + R_3} \\ R_{th,2} = \dfrac{R_4 R_2}{R_2 + R_4} \end{cases}$$

Considering the current flow we have

$$V_{Th} = V_{th,2} - V_{th,1} = V \left(\frac{R_1}{R_1 + R_4} - \frac{R_2}{R_2 + R_3} \right)$$

Solving for R_{Th} we have

$$R_{Th} = R_{th,1} + R_{th,2} = \frac{R_3 R_1}{R_1 + R_3} + \frac{R_4 R_2}{R_2 + R_4}$$

Which implies

$$I_{Th} = \frac{V_{Th}}{R + R_{Th}} \implies V_{Th} \neq 0 \iff R_1 R_3 \neq R_2 R_4$$

Which is the usual condition for having a non-balanced Wheatstone bridge.

§§ 15.4.2 Norton Theorem

The second theorem which lets us simplify circuits into simple ideal source-resistor networks is the *Norton theorem*, which states

Theorem 15.4 (Norton). *Every linear circuit containing only resistors and sources can be simplified to a single circuit composed by a single current source connected in parallel to an equivalent resistance* R_{No}.
If we consider a random circuit with two open poles we have that

- I_{No} *is the current obtained by shorting the poles*
- R_{No} *is the equivalent resistance obtained by shorting voltage sources and opening current sources*

As stated this theorem is directly dual to the Thevenin theorem, noting that, if we consider the following diagram

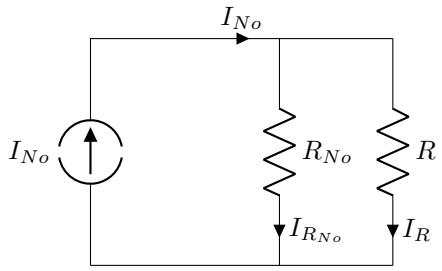


Figure 15.20: Norton equivalent circuit with a load R

We have then

$$\begin{cases} I_{R_{No}} R_{No} = R I_R \implies I_{R_{No}} = \dfrac{R}{R_{No}} I_R \\ I_{No} = I_{R_{No}} + I_R = \left(\dfrac{R}{R_{No}} + 1 \right) I_R \end{cases}$$

But

$$V_{Th} = I_{No} R_{No} = (R + R_{No})\, I_R$$

Therefore

$$I_R = \frac{V_{Th}}{R + R_{No}}$$

This finally implies

$$\begin{cases} R_{Th} = R_{No} \\ V_{Th} = I_{No} R_{No} \\ I_{No} = \dfrac{V_{Th}}{R_{Th}} \end{cases} \tag{15.7}$$

§ 15.5 Real Generators

A *real* generator is a current or voltage source with two poles, one positive and one negative, and can be described, in the case of a voltage generator, using the Thevenin theorem as in the following diagram

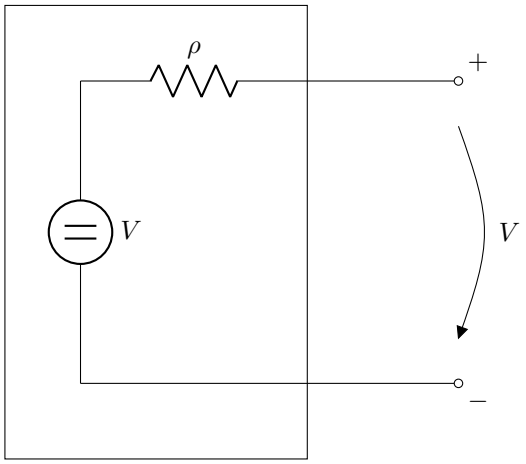


Figure 15.21: Real voltage generator, here $R_{th} = \rho$ and $V = V_{th}$

The equivalent Thevenin resistance is known as the *internal resistance*. A particularity of real voltage generator is that, if we connect a load R to the two poles we have

$$\frac{\partial V}{\partial I} \neq 0$$

For Thevenin, also

$$\begin{cases} I = \dfrac{V}{R+\rho} \\ V_R = RI = \dfrac{R}{R+\rho} V \end{cases}$$

For a current generator instead we have the same setup we'd get from Norton's theorem

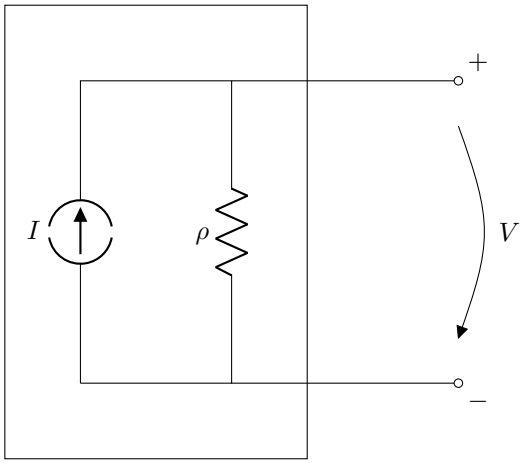


Figure 15.22: Real current generator, here $R_{No} = \rho$ and $I = I_{No}$

Both generators can be approximated to ideal generators when $\rho << R$, with R being the resistance of the connected load

§§ 15.5.1 Joule Effect

The main effect of real generators is *Joule effect*. Considering a simple battery, there's a energy conversion from chemical energy to electric energy, and part of it gets dissipated both in the internal resistance of the source and of the load as thermal energy.
The work done by the source is

$$\mathrm{d}W = V\mathrm{d}Q = VI\mathrm{d}t$$

Where we used $\mathrm{d}Q = I\mathrm{d}t$
The power P is by definition the time derivative of work, thus

$$P = \frac{\mathrm{d}W}{\mathrm{d}t} = VI \tag{15.8}$$

Thanks to this definition, we can define the power dissipated in the load R as

$$P_R = V_R I_R$$

Where

$$\begin{cases} V_R = \dfrac{R}{R+\rho}V \\ I_R = \dfrac{V}{R+\rho} \end{cases}$$

Thus

$$P_R = \frac{RV^2}{(\rho+R)^2}$$

J The dissipated power is clearly dependent on the resistance of the load, thus we can optimize the impedance through optimization methods. We have

$$\frac{\partial P_R}{\partial R} = V^2 \frac{\rho - R}{(\rho + R)^2}$$

Finding the maximal extreme of the first derivative we get that it's true only if and only if $R = \rho$. When the load resistance has this particular value, it's said to be *impedance matched*. Here, we have

$$\max P_R = \frac{V^2}{4\rho}$$

Note that an impedance matched load is doesn't correspond to the maximum impedance. We can imagine to define a power transfer efficiency as η, where

$$\eta = \frac{P_R}{P_{source} + P_R} = \frac{R}{R + \rho} \tag{15.9}$$

Note that $\eta \approx 1$ if and only if $R >> \rho$.

16 Alternate Current Circuits

§ 16.1 Variable Current Circuits

§§ 16.1.1 RC Circuit in DC

Let's begin to consider what happens when currents aren't anymore constant in time but instead are variable. The simplest possible circuit that we can imagine with variable currents (but with a continuous source) is the following one

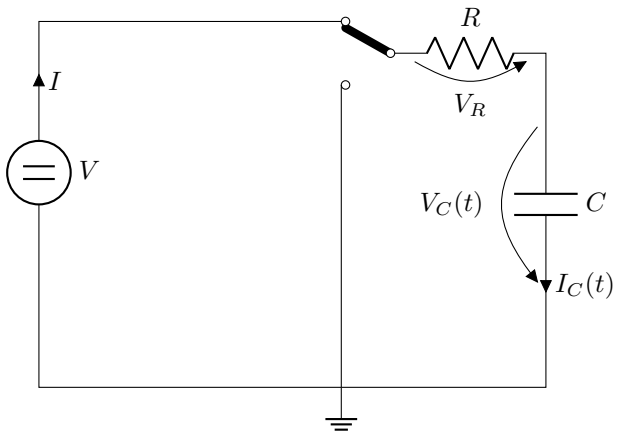


Figure 16.1: Diagram of a circuit which charges a capacitor with DC current and then lets it discharge freely to ground

If we consider two time frames, one when the switch is on the other configuration and charges the capacitor, till a time t_0, when the capacitor is fully charged, and then gets switched and the charged capacitor discharges on the ground, we have, during the charge of the capacitor

$$V = V_R + V_C = RI + \frac{Q}{C} = R\frac{\mathrm{d}Q}{\mathrm{d}t} + \frac{C}{C}$$

Where we used the following identities

$$\begin{cases} C = \dfrac{Q}{V_C} \\ I = \dfrac{\mathrm{d}Q}{\mathrm{d}t} \end{cases}$$

Solving for the current we have

$$RC\frac{\mathrm{d}Q}{\mathrm{d}t} = CV - Q$$

And solving this ordinary differential equation between $t = 0$ and $t = t_0$ we get

$$log\left(\frac{CV - Q}{CV - Q(0)}\right) = -\frac{t}{RC}$$

Which after solving for Q becomes, noting that at $t = 0$ we have $Q(0) = 0$

$$Q(t) = CV\left(1 - e^{-\frac{t}{RC}}\right) \tag{16.1}$$

We define the *relaxation time of the circuit* $\tau = RC$, and after substituting the solution into the definition of the capacitance and solving for V_C we have

$$V_C(t) = V\left(1 - e^{-\frac{t}{\tau}}\right) \tag{16.2}$$

If we solve for V_R, noting that $V_R = RI = R\dot{Q}$ we get that the voltage at the poles of the resistor at a time t is

$$V_R(t) = \frac{RCV}{\tau}e^{-\frac{t}{\tau}} = Ve^{-\frac{t}{\tau}} \tag{16.3}$$

The circuit is said to be stable when $t \gtrsim 4\tau$, so that we can assume the capacitor as fully charged. We can now analyze the case when the capacitor is discharging, i.e. when the switch is flipped and the resistance and capacitor are isolated, and the circuit is practically the following

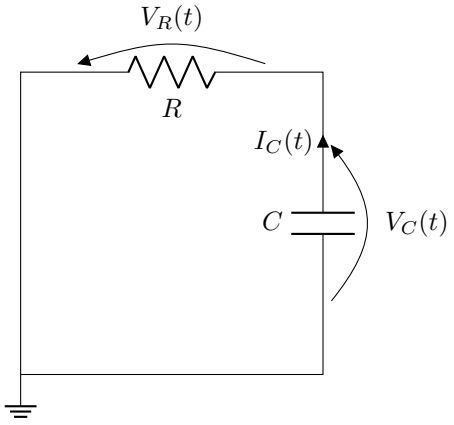


Figure 16.2: Diagram of the discharge of the capacitor C

Here, since we're directly connected to ground, we must have $V = 0$, thus

$$I(t)R + V_C(t) = 0$$

Thus, through substitution we have

$$R\frac{\mathrm{d}Q}{\mathrm{d}t} = -\frac{Q(t)}{C} \implies Q(t) = Q(0)e^{-\frac{t}{RC}}$$

But $Q(0) = CV$m and thus

$$\begin{cases} Q(t) = CVe^{-\frac{t}{\tau}} \\ V_C(t) = \dfrac{Q(t)}{C}Ve^{-\frac{t}{\tau}} \end{cases} \tag{16.4}$$

§§ 16.1.2 RC Circuit in AC

Consider now the same circuit using an alternate voltage source $V(t)$. Consider the special case where this source is an harmonic source, as

$$V(t) = V_0 \cos(\omega t)$$

The circuit diagram is the following

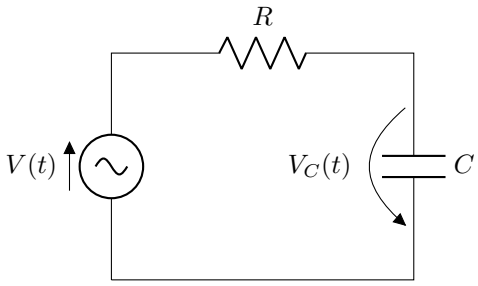


Figure 16.3: Alternate current version of the resistor-capacitor circuit treated in the previous section

As we have already saw before, we have

$$\begin{cases} V_C = \dfrac{Q}{C} \\ V_R = R\dfrac{\mathrm{d}Q}{\mathrm{d}t} \end{cases}$$

Thus, the *characteristic differential equation of the circuit* is

$$V(t) - R\frac{\mathrm{d}Q}{\mathrm{d}t} - \frac{Q}{C} = 0$$

Or, noting that

$$\frac{Q}{C} = V_C \implies Q = CV_C \implies R\frac{\mathrm{d}Q}{\mathrm{d}t} = RC\frac{\mathrm{d}V_C}{\mathrm{d}t}$$

We have

$$V(t) - RC\frac{\mathrm{d}V_C(t)}{\mathrm{d}t} - V_C(t) = 0 \tag{16.5}$$

Substituting the functional expression of $V(t)$, we get a not-so-difficult first order differential equation

$$RC\frac{\mathrm{d}V_C}{\mathrm{d}t} + V_C(t) = V_0\cos(\omega t) \tag{16.6}$$

We solve it by searching a solution via the similarity method, i.e. supposing that $V_C(T) \propto \cos(\omega t)$. Said V_{0C} the proportionality constant we get, adding a generic phase

$$\frac{\mathrm{d}V_C}{\mathrm{d}t} = -\omega V_{0C}\sin(\omega t + \varphi)$$

This is better expressed using phasor notation, which is built using Euler's identity. Keeping in mind that we actually measure the real part of this complex function, i.e. $V_C^{(M)}(t) = \mathfrak{Re}\{V_C(t)\}$, we get by deriving and plugging back into the differential equation

$$V_C(t) + RC\frac{\mathrm{d}V_C}{\mathrm{d}t} = V_{0C}e^{i\omega t+i\varphi} + i\omega RCV_{0C}e^{i\omega t+i\varphi} = V_0e^{i\omega t}$$

Solving, we have

$$(1 + i\omega RC)\,V_{0C}e^{i\varphi} = V_0 \implies V_{0C}e^{i\varphi} = \frac{1 - i\omega RC}{1 + (RC\omega)^2}V_0$$

And therefore, substituting $\tau = RC$

$$V_{0C}(t)e^{i\varphi} = \left(\frac{1}{1 + (\tau\omega)^2} - \frac{i\omega\tau}{1 + (\tau\omega)^2}\right)V_0 \tag{16.7}$$

Or, in terms of measured voltages

$$V_C^{(M)} = \frac{V_0}{1 + \omega^2\tau^2}\cos(\omega r + \arctan(-\omega\tau)) \tag{16.8}$$

Note that this solution is strongly dependent on the frequency ω of the input voltage, in fact, we specifically have, ignoring the phase $\varphi = \arctan(-\omega\tau)$

$$\begin{cases} \lim\limits_{\omega\to 0} V_{0C}(\omega) = V_0 \\ \lim\limits_{\omega\to\infty} V_{0C}(\omega) = 0 \end{cases} \tag{16.9}$$

This behavior clearly shows that the voltage is nonzero only when $\omega < \omega_R$, with ω_R being the resonance frequency. This type of circuit is widely known as a *low pass filter*, i.e. a circuit which permits the passage of voltage only to a limit *resonant* frequency <++>

Appendix

A Special Relativity

§ A.1 Principle of Relativity

The principle of relativity states a quite simple but deep affirmation: *All interaction propagate at a constant speed independent from the chosen frame of reference.* This speed is usually denoted as c and it's informally known as the speed of light, which has the following value (in SI units)

$$c = 2.998 \times 10^8 \ \mathrm{m/s} \tag{A.1}$$

In the part on classical mechanics we always intended between the lines that all interactions are instantaneous and therefore we'd have $c \to \infty$ formally. This can be interpreted as taking classical mechanics as an approximation of Einstein's relativity for which $v/c << 1$, which is the case for our really slow classical particles.
Note that this constant speed of propagation precludes that time isn't universal, and it is frame dependent. In order to understand this it's useful to get two coordinate frames K and $\tilde{K}$, where one is moving with respect to the other with a constant speed V.
Suppose now that a point A emits a signal towards two other points B and C

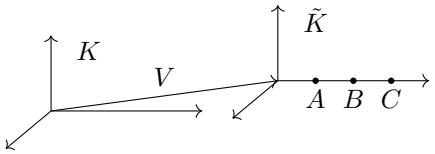


Figure A.1: The two frames K and $\tilde{K}$

In the frame $\tilde{K}$, where A is at rest, we see that the signal reaches both points at the same time, but the same CANNOT be true for the other system, since the relativity principle would be violated. Thinking in a different way, suppose that you're standing at the origin of the K system. If the velocity of the signal is constant in all reference frames we can for sure say that it's so where we're standing, therefore we end up seeing B moving towards the signal and C moving away from it, both with speed V. In this system we therefore must see a delay in when the two points receive such signal.
Although counterintuitive we're experimentally more than sure that this is actually a better approximation of nature than our beloved Newtonian mechanics.

§ A.2 Spacetime

Since time it's not anymore an universal thing and behaves itself as a coordinate, we can now think of our universe as a $4D$ manifold with time as a new coordinate. This is known as *Minkowsky Spacetime* or in short as *Spacetime*. This new definition follows:

Definition A.2.1 (Event). Given a spacetime with coordinates (ct, x, y, z) with c the speed of light, we define a point in spacetime as an *event* in such.
Since time only "flows" one way, we have that for every particle corresponds a wordline which connects all the events pertaining to such. Note that events are also known as *universe points*

Given the principle of relativity one might also ask rightfully how to formulate mathematically all of this, bringing out some invariants that might help with further derivations. Take again the previous system and call l the distance traveled by the signal after being emitted from A. Calling t_1 and t_2 the emission time and the arrival time respectively, we have that for obvious reasons

$$l = c(t_1 - t_2) \tag{A.2}$$

But, we can also write as follows

$$l = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \tag{A.3}$$

With (x_1, y_1, z_1) being the departure coordinates and (x_2, y_2, z_2) the arrival coordinates in K In $\tilde{K}$, analogously we have

$$\begin{aligned} \tilde{l} &= c(\tilde{t}_2 - \tilde{t}_1) \\ \tilde{l} &= \sqrt{(\tilde{x}_2 - \tilde{x}_1)^2 + (\tilde{y}_2 - \tilde{y}_1)^2 + (\tilde{z}_2 - \tilde{z}_1)^2} \end{aligned} \tag{A.4}$$

Tying up both equations we end with the following result

$$\begin{cases} c^2(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2 = 0 \\ c^2(\tilde{t}_2 - \tilde{t}_1)^2 - (\tilde{x}_2 - \tilde{x}_1)^2 + (\tilde{y}_2 - \tilde{y}_1)^2 + (\tilde{z}_2 - \tilde{z}_1)^2 = 0 \end{cases} \tag{A.5}$$

In "layman" words this basically means, that the following quantity

$$s_{12}^2 = c^2(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2 \tag{A.6}$$

Called, *interval*, is a *relativistic invariant*, and therefore invariant with respect to changes of coordinate frames in the context of special relativity.
From (A.5) we have that if the two points are infinitesimally close to eachother we can define the infinitesimal interval as

$$\mathrm{d}s^2 = c^2\mathrm{d}t - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 \tag{A.7}$$

The invariance of such differential quantity is easy to show considering the previous case we stated where $\mathrm{d}s = \mathrm{d}\tilde{s} = 0$ we have, using basic intuition that

$$\mathrm{d}s^2 = a(V)\mathrm{d}\tilde{s}^2 \tag{A.8}$$

Where $a(V)$ is a function of the relative velocity between the two considered frames. It cannot depend on direction due to the isotropy of space.

Consider now three inertial reference frames K, K_1, K_2, and let V_1, V_2 be the velocities of the frames K_1, K_2. We can therefore say, using (A.8) that

$$\begin{aligned} \mathrm{d}s^2 &= a(V_1)\mathrm{d}s_1^2 = a(V_2)\mathrm{d}s_2^2 \\ \mathrm{d}s_1^2 &= a(V_{12})\mathrm{d}s_2^2 \end{aligned} \tag{A.9}$$

Where we defined the velocity between K_1, K_2 as V_{12}. Rewriting the equation we have

$$\mathrm{d}s^2 = a(V_1)a(V_{12})\mathrm{d}s_2^2 = a(V_2)\mathrm{d}s_2^2$$

Equating the coefficients of the differential $\mathrm{d}s_2$, we have

$$a(V_{12}) = \frac{a(V_2)}{a(V_1)} \tag{A.10}$$

The previous equation then might be true if and only if $a(V_{12})$ depends only on the angle between the velocities V_1, V_2. This cannot be true due to the isotropy of spacetime, as we stated for the previous problem, and therefore $a(V)$ might only be a constant function. Taking $a(V_{12}) = 1$ for consistency between frames of reference, we have finally demonstrated that the differential spacetime interval is invariant

$$\mathrm{d}s = \mathrm{d}\tilde{s} \tag{A.11}$$

This definition of $\mathrm{d}s$ gives rise to three kinds of intervals:

1. *Spacelike intervals* if $s_{12}^2 < 0$
2. *Timelike intervals* if $s_{12}^2 > 0$
3. *Light-like intervals* if $s_{12}^2 = 0$

These three distinctions let us answer two previously impossible questions: is it possible to find a reference frame where two events happen at the same time or at the same place in our three-dimensional perception?. The answer is surprisingly *yes*. It depends on the kind of the interval between the two points.
Let's work with the first assumption, taken two events in spacetime E_1, E_2, defined $t_{12} = t_2 - t_1$ and l_{12} as our usual 3D distance between the events, we have

$$s_{12}^2 = c^2 t_{12}^2 - l_{12}^2$$

Let's now search a system where $l'_{12} = 0$. In order to have this, using that $s_{12} = s'_{12}$ we have

$$s_{12} = c^2 t_{12} - l_{12}^2 = c^2 t_{12}^{'2} = s_{12}^{'2} > 0$$

I.e. the spacetime interval between the frame of reference at rest with respect to the two events and the new unknown frame of reference is timelike.
Analogously, if we wanted to find a new system where the two events happen at the same time, we might have set $t'_{12} = 0$, therefore getting

$$s_{12} = c^2 t_{12} - l_{12}^2 = l_{12}^{'2} = s_{12}^{'2} < 0 \tag{A.12}$$

§§ A.2.1 Spacetime Diagrams

The idea of spacetime and absoluteness of the velocity of interactions can be described well by a 2D spacetime diagram. Taken an origin for our system of coordinates (ct, x) we have that, considering v as the slope of a constant wordline, that $|v| < c$.

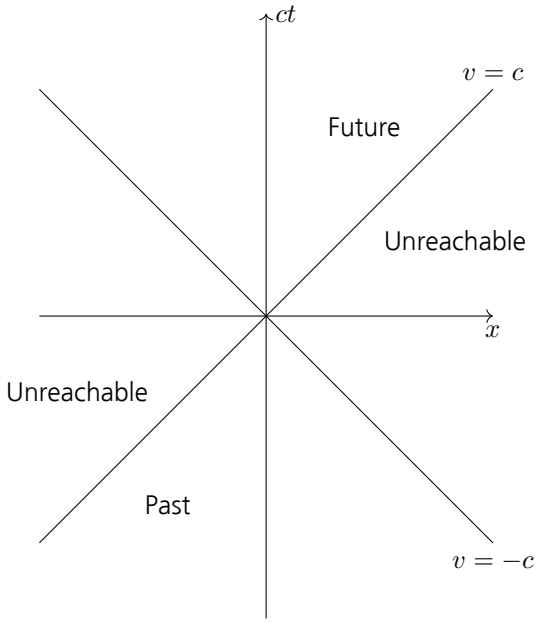


Figure A.2: Simple spacetime diagram. Note how all the events beyond the asymptote (or *horizon*) $v = \pm c$ are inaccessible from 0

Thought in higher dimensions we have that all the past and future of an event are enclosed inside a cone bordered by our horizon $|v| = c$ which separates physical impossibilities from the actual physical past and future of what we're considering.
Note that if $v = \pm c$ we must have $x = \pm ct$, giving us a spacelike interval for our diagram.
Considering instead past and future it's also easy to see that the past is always spacelike, since $c^2t^2 - x^2 < 0$, and that the future is always timelike. Note also that past and future must be absolute

§ A.3 Proper Time

Since time is not a relativistic invariant, we need to search for a good substitute of it. Given a clock fixed at the origin of some inertial frame K'. After some time $\mathrm{d}t$, the clock has moved (in our system) by the following quantity

$$\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2}$$

By definition, in K' this clock is at rest, therefore we have

$$\mathrm{d}x' = \mathrm{d}y' = \mathrm{d}z' = 0$$

Imposing the invariance of intervals we have that

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 = c^2\mathrm{d}t'^2 \tag{A.13}$$

Therefore, it must be true that

$$\mathrm{d}t' = \mathrm{d}t\sqrt{1 - \frac{\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2}{c^2\mathrm{d}t^2}} \tag{A.14}$$

This is the expression for the passing of time in the system where the clock is at rest, and it's called the *proper time* of the clock, usually indicated with τ. Writing the sum of differentials as $\mathrm{d}r^2$ and using the definition of v^2, we have that

$$\mathrm{d}\tau = \mathrm{d}t\sqrt{1 - \frac{v^2}{c^2}} = \frac{\mathrm{d}s}{c} \tag{A.15}$$

Integrating and using the fundamental theorem of calculus, we have that a given time interval will be "felt" differently by the clock, where

$$\Delta\tau = \int_{\tau_1}^{\tau_2} \sqrt{1 - \frac{v^2}{c^2}}\mathrm{d}t < \Delta t \tag{A.16}$$

This tells us that a moving clock will tick slower than a clock at rest (note also on how this definition depends directly on the chosen frame).
This difference of measured time is known as *time dilation*.

§ A.4 Formalization of the Principle of Relativity

All of what we found before can be crammed into the most fundamental element of relativity: coordinate transformations.
Consider two reference frames $K,\ (ct, x, y, z)$ and $K',\ (ct, x, y, z)$. Mathematically, what we call interval is the usual 4D distance in a seminegative definite metric, and due to its invariance we must have that all coordinate transformations between these two systems must be rototraslations (isometries). Translations can be immediately ignored since they only move the origin of the system, and therefore we choose our faithful rotations in order to find these coordinate transformation laws.
All the possible rotations are between the planes xy, xz, yz and tx, ty, tz. All rotations xy, xz, yz are our usual 3D rotations and are of no use, therefore we choose the rotations tx, ty, tz. Taking tx as the chosen one we have that the spacetime interval is

$$s^2 = c^2t^2 - x^2$$

Therefore, all searched rotations *must preserve* this relationship. The first idea one might have is to look at the symmetry of the system and deduce immediately that such rotation must be hyperbolic in nature. We therefore define the following

$$\begin{pmatrix} x \\ ct \end{pmatrix} = \begin{pmatrix} \cosh\psi & \sinh\psi \\ \sinh\psi & \cosh\psi \end{pmatrix} \begin{pmatrix} x' \\ ct' \end{pmatrix} \tag{A.17}$$

Taking $x' = 0$ it all reduces to this single equation

$$\frac{x}{ct} = \frac{V}{c} = \tanh\psi \tag{A.18}$$

It's common to indicate such value with the pure number β, called the *Lorentz Boost*, where

$$\beta = \frac{V}{c}$$

Solving (A.18) we have that

$$\beta = \frac{\sinh\psi}{\sqrt{1+\sinh^2\psi}} = 0 \implies \sinh^2\psi = \frac{\beta}{\sqrt{1-\beta^2}} \tag{A.19}$$

And

$$\cosh^2\psi = 1+\sinh^2\psi \implies \cosh\psi = \frac{1}{\sqrt{1-\beta^2}} = \gamma \tag{A.20}$$

Where γ is known as the *Lorentz/Gamma Factor*.
Substituting back into (A.18) we have back our searched transformations

$$\begin{pmatrix} x \\ ct \end{pmatrix} = \begin{pmatrix} \gamma & \beta\gamma \\ \beta\gamma & \gamma \end{pmatrix} \begin{pmatrix} x' \\ ct' \end{pmatrix} \tag{A.21}$$

Note that the inverse transformation is simply given imposing $\beta \to -\beta$.
The complete transformation between the two reference frames will finally be a 4D linear system as follows

$$\begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \gamma & \beta\gamma & 0 & 0 \\ \beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} \tag{A.22}$$

These transformations are known as *Lorentz Transformations* and are the fundamental transformations between frames of reference in special relativity. These transformations formalize the principle of relativity. For $v << c$ these transformations bring back the usual Galilean transformations corrected by a first order factor in c, as we expected

$$\begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & \beta & 0 & 0 \\ \beta & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} \tag{A.23}$$

§§ A.4.1 Length Contraction and Time Dilation

Using Lorentz transformations it's possible to mathematically formalize all relativistic effects. One of such is known as *length contraction*, where the measured length of an object depends on the chosen reference frame.
As a matter of example take a "rigid" rod in a system K, long Δx, and consider the system K' where the rod is at rest. In this system we have

$$\Delta x' = x'_1 - x'_2 = \gamma(x_2 - x_1) - \gamma\beta c(t_2 - t_1) = \gamma\Delta x - \gamma\beta c\Delta t \tag{A.24}$$

Since we're measuring the length directly, we can say without problems that $\Delta t = 0$, and we get

$$\Delta x' = \gamma \Delta x = \frac{\Delta x}{\sqrt{1-\beta^2}} = \frac{\Delta x}{\sqrt{1-\frac{v^2}{c^2}}} \tag{A.25}$$

Therefore, for $\beta \neq 0$ we have $\Delta x' < \Delta x$. We call $\Delta x = l_0$ as the proper lenght of this rod.
Note that a major consequence of this is that a rigid body in the classical sense of the term cannot be conceived in Special Relativity.
A second effect that we stated before and didn't formalize properly is that of time dilation. Taken a clock at rest in a system K' and two events happening at some coordinate (x', y', z') of K'. We have that the time elapsed between the two events will be $\Delta t' = t'_2 - t'_1$, and therefore, using Lorentz transformations we get, in K

$$\Delta t = \gamma \left(t'_1 + \frac{\beta}{c} \Delta x' \right) \tag{A.26}$$

Imposing that the events happen at the same place (x', y', z') we have $\Delta x' = 0$ and therefore

$$\Delta t = \gamma \Delta t' \tag{A.27}$$

Therefore, the clock in the still frame is measuring smaller time intervals, and the time measured is dilated.

§§ A.4.2 Velocity Transformations

As we have seen velocities have an upper bound which is the speed of light. It's possible to find the transformations of velocities from the transformations (A.21) and applying them to differentials.
We have

$$\begin{pmatrix} \mathrm{d}t \\ \mathrm{d}x \\ \mathrm{d}y \\ \mathrm{d}z \end{pmatrix} = \begin{pmatrix} \gamma & \frac{\beta\gamma}{c} & 0 & 0 \\ \frac{\beta\gamma}{c} & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathrm{d}t' \\ \mathrm{d}x' \\ \mathrm{d}y' \\ \mathrm{d}z' \end{pmatrix} \tag{A.28}$$

Rearranging the terms we have finally

$$\begin{cases} v_x = \dfrac{v'_x + \beta c}{1 + \frac{\beta}{c} v'_x} \\ v_y = \dfrac{v'_y}{\gamma \left(1 + \frac{\beta}{c} v'_x \right)} \\ v_z = \dfrac{v'_z}{\gamma \left(1 + \frac{\beta}{c} v'_x \right)} \end{cases} \tag{A.29}$$

Approximating for $v << c$ we get the usual velocity composition formula with an added relativistic correction

$$\begin{cases} v_x \approx v'_x + V \left(1 - \dfrac{v_x^{2'}}{c^2} \right) \\ v_y \approx v'_y - v'_x v'_y \dfrac{\beta}{c} \\ v_z \approx v'_z - v'_x v'_z \dfrac{\beta}{c} \end{cases} \tag{A.30}$$

Or, in vector form

$$v^i = v^{i'} + V^i - \frac{v^{i'}}{c^2}(V^i v'_i) \tag{A.31}$$

Note how v and v' are tied asymetrically in the transformation. Consider now a simple planar motion in the xy plane, where $v^i = (v_x, v_y, 0)$, we can find the law of transformation of angles considering that v^i can be rewritten in polar coordinates, as follows

$$\begin{cases} v_x = v\cos\theta \\ v_y = v\sin\theta \\ v_z = 0 \end{cases}$$

Applying the transformations, we have

$$\begin{cases} v\cos\theta = \dfrac{v'\cos\theta' + \beta c}{1 + \frac{\beta}{c}v'\cos\theta'} \\ v\sin\theta = \dfrac{v'\sin\theta'}{\gamma\left(1 + \frac{\beta}{c}v'\cos\theta'\right)} \end{cases} \tag{A.32}$$

Where we used that the motion in the new system will be still planar.
Rewritten in other terms, we have

$$\tan\theta = \frac{\frac{v'\sin\theta'}{\gamma\left(1+\frac{\beta}{c}v'\cos\theta'\right)}}{\frac{v'\cos\theta'+\beta c}{1+\frac{\beta}{c}v'\cos\theta'}} = \frac{v'\sin\theta'}{\gamma\left(v'\cos\theta' + \beta c\right)} \tag{A.33}$$

Which explicitates the change of direction of velocity between different coordinate systems.

§ A.5 4-Vectors

[1]

As we have already suggested before, the 4-tuple $x^\mu = (ct, x, y, z)$ can be seen as a set of coordinates in spacetime, or as a radius vector. The square of vectors in spacetime can be seen as a non-euclidean scalar product as follows

$$x^\mu x_\mu = g_{\mu\nu}x^\mu x^\nu = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 \tag{A.34}$$

Where $g_{\mu\nu}$ is the metric tensor of spacetime

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{A.35}$$

From what we wrote for special relativity itself, we have a new definition

[1] From here on, all greek indexes $(\mu, \nu, \sigma, \cdots)$ are to be intended as spacetime indexes, and latin indexes $(i, j, k, \cdots)$ as usual 3D indexes if not otherwise stated

Definition A.5.1 (4-Vector)**.** A *4-vector* is a 4-tuple that transforms between coordinate frames using Lorentz transformations, as

$$a^\mu = \Lambda^\mu_\nu a^\nu \tag{A.36}$$

Where λ^μ_ν is the already defined transformation matrix of the Lorentz transformations.

Using the metric tensor one can transforms between covariant vectors and contravariant vectors using $a_\mu = g_{\mu\nu}a^\nu$, and due to the semidefinite signature of the metric one has that $a^i = -a_i$, where a^i is the spatial part of the vector. Note also that inserting it into the formula for a scalar product ($a^\mu b_\mu$) one gets back what we had defined before.
It's also possiible to define 4-scalars, which are relativistic invariants. One of such 4-scalars is the square of a 4-vector or the scalar product between 2 4-vectors.
Another way of writing 4-vectors is with a tuple composed as follows

$$a^\mu = (a^0, a^i) \tag{A.37}$$

Where the first component is known as the *polar* component of the 4-vector, and the second is known as the *axial* component of the 4-vector. Therefore we can write

$$\begin{aligned} x^\mu &= (ct, x^i) \\ x_\mu &= (ct, -x_i) \end{aligned} \tag{A.38}$$

§§ A.5.1 4-Velocity and 4-Acceleration

It's possible to define a 4-vector analogue to the velocity of a particle. Indicating with τ the proper time we define the 4-velocity u^μ as

$$u^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} \tag{A.39}$$

Since $\mathrm{d}\tau = \frac{c}{\gamma}\mathrm{d}t$ we have

$$u^\mu = \frac{\gamma}{c}\frac{\mathrm{d}x^\mu}{\mathrm{d}t}$$

In other words

$$u^\mu = \left(\gamma, \frac{\gamma}{c}v^i\right)$$

Note that the square of u^μ is a relativistic invariant and special in nature due to its unitary value, in fact

$$u^\mu u_\mu = \gamma^2 - \gamma^2\frac{v^2}{c^2} = 1$$

The 4-acceleration w^μ is defined analogously derivating again with respect to the proper time, hence

$$w^\mu = \frac{\gamma}{c}\,\mathrm{d}u^\mu\, t = \left(\frac{\gamma}{c}\frac{\mathrm{d}\gamma}{\mathrm{d}t}, \frac{\gamma}{c^2}\frac{\mathrm{d}\gamma v^i}{\mathrm{d}t}\right) \tag{A.40}$$

Deriving with respect to time we have firstly that

$$\frac{\mathrm{d}\gamma}{\mathrm{d}t} = \frac{v^i a_i}{\left(1-\frac{v^2}{c^2}\right)^{\frac{3}{2}}} = \frac{\gamma^3}{c^2}v^i a_i$$

And therefore

$$w^\mu = \frac{\mathrm{d}u^\mu}{\mathrm{d}\tau} = \frac{\gamma}{c}\left(\frac{\gamma^3}{c^2}v^i a_i, \frac{\gamma}{c^3}v^j a_j v^i + \frac{\gamma}{c}a^i\right) \tag{A.41}$$

It's possible to demonstrate that $w^\mu u_\mu = 0$, i.e. that 4-velocity and 4-acceleration are always mutually orthogonal. In fact

$$\frac{\mathrm{d}}{\mathrm{d}\tau}u^\mu w_\mu = \frac{\mathrm{d}u^\mu}{\mathrm{d}\tau}u_\mu + \frac{\mathrm{d}u_\mu}{\mathrm{d}\tau}u^\mu = 2u^\mu w_\mu = 0$$

§ A.6 Exercises

Exercise A.6.1 (Uniformly Accelerated Motion)**.** Solve the motion of an uniformly accelerated particle in the context of Special Relativity.
Consider that the 4-acceleration is constant only in the frame comoving with the particle.
SOLUTION
We have that in the comoving frame $\gamma = 1$ and $v = 0$, therefore

$$w^\mu = \left(0, \frac{\dot{v}^i}{c^2}\right)$$

Since a is constant we rotate the 3D system in order to get $a \parallel x$, therefore getting

$$w^\mu = \left(0, \frac{a}{c^2}, 0, 0\right)$$

Note that we can also define a 4-scalar

$$w^\mu w_\mu = -\frac{a^2}{c^2}$$

Changing to the fixed frame of reference, we have

$$w^{\mu'} = \frac{\gamma}{c}\left(\frac{\gamma^3}{c^2}v^i\dot{v}_i, \frac{\gamma^3}{c^2}v^j\dot{v}_j v^i + \frac{\gamma}{c}\dot{v}^i\right) = \frac{\gamma^4}{c^2}\left(\frac{v^i\dot{v}_i}{c}, \frac{v^2}{c^2}\dot{v}^i + \frac{\dot{v}^i}{\gamma^2}\right)$$

Using that

$$\left(\frac{v^2}{c^2} + \frac{1}{\gamma^2}\right)\dot{v}^i = \dot{v}^i$$

We end up with the following simplified result

$$w^{\mu'} = \frac{\gamma^4}{c^2}\left(\frac{1}{c}\dot{v}^i v_i\right)$$

Which gives us the following differential equation

$$w^\mu w_\mu = \frac{\gamma^8}{c^4}\left(\frac{1}{c^2}(v^i\dot{v}_i)^2\right) - \frac{\gamma^8}{c^4}\dot{v}^2 = -\frac{a^2}{c^4}$$

Simplifying the LHS we get

$$\frac{\gamma^8}{c^4}\left(\frac{v^2}{c^2}\dot{v}^2-\dot{v}^2\right)=\frac{\gamma^8}{c^4}\left(\frac{v^2}{c^2}-1\right)=-\frac{\gamma^6}{c^4}\dot{v}^2$$

Therefore, putting it back into the first equation, we get

$$-\gamma^6\dot{v}^2=-a^2 \implies \gamma^3\frac{\mathrm{d}v}{\mathrm{d}t}=a$$

Note that using the derivative of γ with respect to time we can rewrite the LHS as the derivative of a product, in fact

$$\frac{\mathrm{d}(\gamma v)}{\mathrm{d}t}=\frac{\gamma^3}{c^2}v^2\dot{v}+\gamma\dot{v}=\dot{v}\left(\frac{\gamma^3}{c^2}v^2+\gamma\right)=\gamma^3\dot{v}\left(\frac{v^2}{c^2}+\frac{1}{\gamma^2}\right)=\gamma^3\dot{v}$$

Therefore, finally

$$\frac{\mathrm{d}(\gamma t)}{\mathrm{d}t}=a \implies \gamma v(t)=at+c$$

Imposing that $v(0)=0$ we get $c=0$ and therefore, solving for $v(t)$, we have

$$\frac{v(t)}{\sqrt{1-\frac{v^2}{c^2}}}=at \implies v^2=a^2t^2-\frac{a^2t^2}{c^2}v^2 \implies v^2=a^2t^2\left(1+\frac{a^2t^2}{c^2}\right)^{-1}$$

Therefore

$$v(t)=\frac{at}{\sqrt{1+\frac{a^2t^2}{c^2}}}$$

Then, by direct integration we can find $x(t)$

$$x(t)=\int\frac{at}{\sqrt{1+\frac{a^2t^2}{c^2}}}\mathrm{d}t=\frac{c^2}{2a}\int\frac{1}{\sqrt{1+w^2}}\mathrm{d}w=\frac{c^2}{2a}\left(2\sqrt{1+w}+k\right)$$

Where we used the substitution $w=\frac{a^2t^2}{c^2}$. Imposing the initial condition that $x(0)=0$ we get $k=-1$, and therefore

$$x(t)=\frac{c^2}{a}\left(2\sqrt{1+\frac{a^2t^2}{c^2}}-1\right)$$

The proper time of the particle is

$$\tau=\frac{1}{c}\int_{s_0}^{s}\mathrm{d}s=\int_{t_0}^{t}\frac{1}{\gamma}\mathrm{d}t=\int_0^t\sqrt{1-\frac{v^2}{c^2}}\mathrm{d}t$$

From the definition of $v(t)$ we have that

$$\gamma=\frac{1}{1-\frac{a^2t^2}{c^2\left(1+\frac{a^2t^2}{c^2}\right)}}$$

Therefore our integral becomes

$$\tau = \int_0^t \sqrt{1 - \frac{a^2t^2}{c^2\left(1 + \frac{a^2t^2}{c^2}\right)}}\mathrm{d}t = \frac{a}{c}\int_0^{\frac{a}{c}t} \sqrt{1 - \frac{z^2}{1+z^2}}\mathrm{d}z = \frac{a}{c}\int_0^{\frac{a}{c}t} \frac{1}{\sqrt{1+z^2}}\mathrm{d}z = \frac{a}{c}\arcsin\left(\frac{at}{c}\right)$$

Where we used the substitution $\frac{at}{c} = z$

B Waves

§ B.1 Plane Harmonic Waves

The wave equation in 4d is the following second order partial differential equation

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = \frac{1}{u^2}\frac{\partial^2 f}{\partial t^2} \tag{B.1}$$

Taken the simpler case of a wave equation moving through a single spatial dimension, i.e.

$$\frac{\partial^2 f}{\partial x^2} = \frac{1}{u^2}\frac{\partial^2 f}{\partial t^2}$$

We already know that it has a simple solution in terms of cosines (or sines)

$$f(x,t) = f_0 \cos(kx - \omega t) \tag{B.2}$$

Given that

$$\frac{\omega}{k} = u$$

The solution (B.2) is of particular importance in treating electromagnetic waves (or waves in general). It tells us that the wave $f(x,t)$ varies sinusoidally with the distance x and it's harmonic in t for any given point in space.
Fixing the position and moving only through time, going to a new time $t + \Delta t$, we have that due to the previous constraint the point x will have moved by $\Delta x = u\Delta t$.
It's exactly like adding a phase to the wave (hence the name *phase velocity* of u). Note that if we used another solution, say

$$f_r(x,t) = f_0 \cos(kx + \omega t)$$

We would describe a wave going back, with displacement $\Delta x = -u\Delta t$ for some given time.
In terms of electrodynamics, all these constants have names:

- u is the phase velocity
- λ is the wavelength
- ω is the angular frequency
- k is the angular wavenumber

There are also other derived values that have given name, which are

- T the period, for which $\lambda = uT = \frac{2\pi}{k}$
- ν the frequency, for which $\nu = \frac{u}{\lambda} = \frac{\omega}{2\pi} = T^{-1}$
- σ the spectroscopic wavenumber, for which $\sigma = \frac{1}{\lambda}$

Going back to waves in 3 spatial dimensions, the general solution of the wave equation (B.1) is known as a *plane harmonic wave*, and has the following mathematical shape

$$f(x, y, z, t) = f_0 \cos\left(k^i x_i - \omega t\right) \tag{B.3}$$

Here the angular wavenumber (or wavenumber for simplicity), is a vector, and is known as the *propagation vector* k^i. The magnitude of this vector *is the actual wavenumber*, i.e.

$$k = \sqrt{k^i k_i}$$

The physical meaning of this solution resides mostly inside the cosine argument, $k^i x_i - \omega t$.
Setting it as constant we get the equation of a plane in space, which are called the *surfaces of constant phase*. The normals of these surfaces (planes) are proportional and perpendicular to the wavevector, and these planes "move" in that direction at a rate equal to the phase velocity u

§§ B.1.1 Alternative Representations of the Wavefunction

We can think immediately about alternative representations of the wavefunction (B.3). One of these is by using the definition of the constant phase surfaces we gave before, in fact, as we said $k^i = k\hat{n}^i$, and therefore

$$f(x^i, t) = f_0 \cos\left[(x^i \hat{n}_i - ut)k\right]$$

Remember that $ku = \omega$.
Another one, is by using complex functions. Using the Euler identity for the complex exponential we can write

$$f(x^i, t) = F_0 e^{i(k^i x_i - \omega t)}$$

Note that since generally $F_0 \in \mathbb{C}$, the actual physical quantity is the real part of f.
One use of the complex wavefunction is immediate when dealing with spherical waves, where $x^i = r$ and

$$f(r, t) = \frac{1}{r}\cos(kr - \omega t) = \mathfrak{Re}\left\{\frac{1}{r} e^{i(kr - \omega t)}\right\} \tag{B.4}$$

Ignoring the real part on the right it's clear that the second will be extremely easier to manipulate.

§ B.2 Group Velocity

Suppose now that we have two harmonic waves ϕ, ψ with different angular frequencies $\omega_1 = \omega + \Delta\omega$ and $\omega_2 = \omega - \Delta\omega$.
In general also the wavenumbers will differ, and we'll call them for simplicity $k_1 = k + \Delta k$, $k_2 = k - \Delta k$.

Supposing that the waves have the same amplitude A and are traveling in the same direction (say z), the superposition of the two will be

$$\Psi = \psi + \phi = A\left(e^{i((k+\Delta k)z-(\omega+\Delta\omega)t)} + e^{i((k-\Delta k)z-(\omega-\Delta\omega)t)}\right)$$

Collecting terms

$$\Psi = Ae^{i(kz-\omega t)}\left(e^{i(z\Delta k-t\Delta\omega)} + e^{-i(z\Delta k-t\Delta\omega)}\right)$$

Recognizing a $2\cos(\cdots)$ on the right we have that

$$\Psi = 2Ae^{i(kz-\omega t)}\cos\left(z\Delta k - t\Delta\omega\right)$$

The result of this superposition is a new wave η, with amplitude $B = 2A$ multiplied by a "modulation envelope" given by the cosine.
From the previous results it's clear that this envelope doesn't travel at the phase velocity, in fact, we have a new propagation velocity, known as the *group velocity*

$$u_g = \frac{\Delta\omega}{\Delta k} \tag{B.5}$$

At the limit, we have that $\omega'(k) = u_g$, and since $\omega = ku$ or $\omega = kc/n$ we have

$$u_g = \frac{\mathrm{d}}{\mathrm{d}k}\left(\frac{kc}{n}\right) = \frac{c}{n} - \frac{ck}{n^2}\frac{\mathrm{d}n}{\mathrm{d}k} = u\left(1 - \frac{k}{n}\frac{\mathrm{d}n}{\mathrm{d}k}\right)$$

Note that we wrote $n = n(k)$! The refraction index is in general dependent on the angular frequency ω, and therefore on k! As an example, you can see how different wavelengths of light behave passing through glass. They go through when λ is in the range of visible light, but are completely opaque in infrared, i.e. n *must* vary with frequency (or wavelength, or wavenumber, or …)
From the previous monologue and from the definitions we can then say that

$$\begin{aligned} u_g &= u\left(1 - \frac{k}{n}\frac{\mathrm{d}n}{\mathrm{d}k}\right) \\ u_g &= u - \lambda\frac{\mathrm{d}u}{\mathrm{d}\lambda} \\ \frac{1}{u_g} &= \frac{1}{u} - \frac{\lambda_0}{c}\frac{\mathrm{d}n}{\mathrm{d}\lambda_0} \end{aligned} \tag{B.6}$$

Where λ_0 is the vacuum wavelength.
Phase velocity and group velocity can only be equal in vacuum ($n = 0$), where $u_g = u = c$.

Bibliography

[Fow75] Grant R. Fowles. *Introduction to Modern Optics*. 2ed. Dover Publications Inc., New York, 1975. isbn: 978-0-486-65957-2.

[Goo96] Joseph W. Goodman. *Introduction to Fourier Optics*. 2ed. McGraw-Hill, 1996. isbn: 0-07-024254-2.

[Gri17] D. J. Griffiths. *Introduction to Electrodynamics*. Cambridge University Press, 2017. isbn: 978-1-108-42041-9.

[Jac98] J. D. Jackson. *Classical Electrodynamics*. John Wiley and Sons, Inc., 1998. isbn: 0-471-30932-X.

[LL71] L. D. Landau and E. M. Lifshits. *Course of Theoretical Physics, Vol.2, The Classical Theory of Fields*. MIR, 1971.

[Zan12] A. Zangwill. *Modern Electrodynamics*. Cambridge University Press, 2012. isbn: 978-0-521-89697-9.