Electromagnetism

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

Matteo Cheri

Notes on Electromagnetism

January 9, 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.

January 9, 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

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

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

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

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

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

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 intensity 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 intensity* M^i as follows

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

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 intensity, 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.

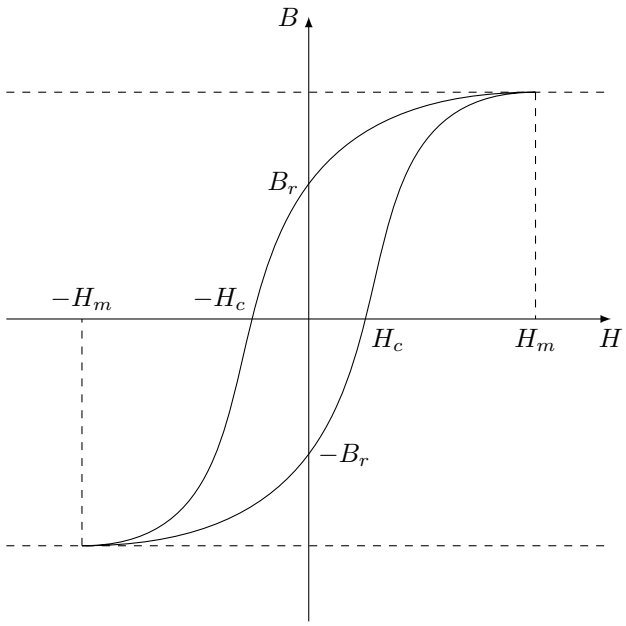

Figure 5.1: Example of an hysteresis cycle, without the first magnetization curve.

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

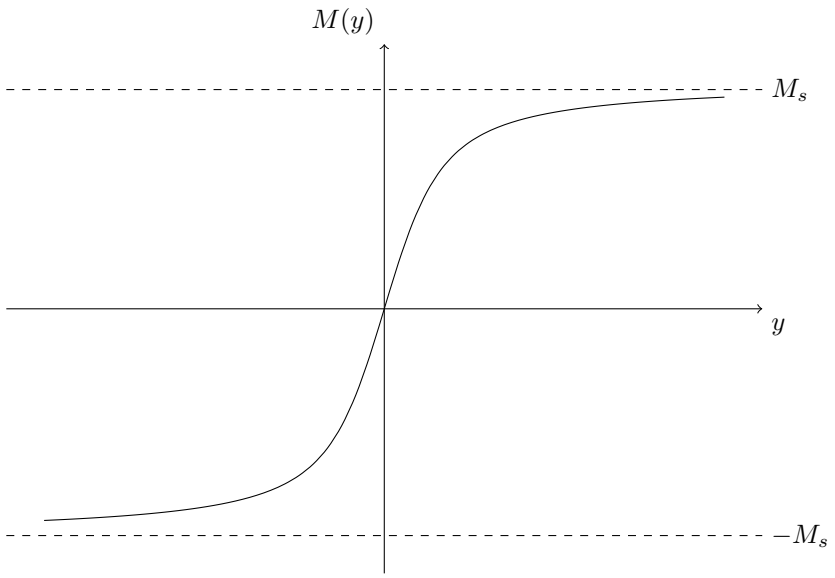

Figure 5.2: Langevin function for Magnetization

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

$$\begin{cases}e^{ikl}+\dfrac{E'_R}{E_1}e^{-ikl}=\dfrac{E_T}{E_1}\\[2ex]\dfrac{E_1}{E'_R}e^{ikl}+e^{-ikl}=\dfrac{E_T}{E'_R}\end{cases}$$

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

$$\begin{cases}E+E_R=E_1+E'_R\\n_0\left(E-E_R\right)=n_1\left(E_1-E'_R\right)\end{cases}\qquad\begin{cases}E'_R=\dfrac{n_1-n_T}{2n_1}E_Te^{ikl}\\[2ex]E_1=\dfrac{n_1+n_T}{2n_1}E_Te^{-ikl}\end{cases}\tag{8.80}$$

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 intensity 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 intensity, but an average. Also, the interfering waves must not have the same frequencies and can come from different sources, therefore the total intensity 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 intensity 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 intensity 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 intensity 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 intensity 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 intensity 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

$$R = \left| \frac{E_R}{E_0} \right|^2 = R_1 \frac{\left|1 - e^{i\delta}\right|^2}{|1 - R e^{i\delta}|}^2$$

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 intensity 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 intensity around these peaks, i.e. at $\omega - \omega_m$, we have

$$I_T = I_0 T^2\frac{1}{T^2+4(1-T)\sin^2\left(\frac{dm}{\pi c}(\omega-\omega_m)\right)} \approxeq I_0\frac{\frac{\sigma^2}{4}}{\frac{\sigma^2}{4}+(\omega-\omega_m)^2} \tag{9.43}$$

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 intensity, 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 intensity 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

$$\int_0^s e^{\frac{i\pi z^2}{2}} \mathrm{d}z = \int_0^s \cos\left(\frac{\pi z^2}{2}\right) \mathrm{d}z + i \int_0^s \sin\left(\frac{\pi z^2}{2}\right) \mathrm{d}z = C(s) + iS(s) \tag{10.53}$$

<++>

§ 10.4 Fourier Theory of Diffraction

§§ 10.4.1 Apodization

§§ 10.4.2 Spatial Filtering

§§ 10.4.3 Phase Gratings

11 Optics of Solids

12 Geometric Optics

Part IV

Old Chapters

13 EMW-old

§ 13.1 Propagation of Electromagnetic Waves

Consider a random point in spacetime, where only an electromagnetic field is present.
Here we have

$$\rho = 0$$
$$J^i = 0$$

Using H^i instead of B^i, since, in free space

$$B^i = \mu_0 H^i$$

We get Maxwell's equation for what's known as a free electromagnetic field

$$\begin{cases} \partial_i E^i = 0 \\ \epsilon^i{}_{jk} \partial^j E^k = -\mu_0 \dfrac{\partial H^i}{\partial t} \\ \partial_i H^i = 0 \\ \epsilon^i{}_{jk} \partial^j H^k = \epsilon_0 \dfrac{\partial E^i}{\partial t} \end{cases} \tag{13.1}$$

The absence of sources here is given by the two divergence relations. Note also that these coupled PDEs are valid both in the dynamic and static case.
These equations, although it might not be that clear from the system, are completely separable. Taken the two curl equations we have, using the relations found before, that

$$\begin{cases} \epsilon^i{}_{jk} \epsilon^k{}_{lm} \partial^j \partial^l E^m = -\mu_0 \epsilon_0 \dfrac{\partial^2 E^i}{\partial t^2} \\ \epsilon^i{}_{jk} \epsilon^k{}_{lm} \partial^j \partial^l H^m = -\mu_0 \epsilon_0 \dfrac{\partial^2 H^i}{\partial t^2} \end{cases}$$

Using the Levi-Civita identity we have that the double curl becomes

$$\partial^i \left(\partial_j E^j \right) - \partial^j \partial_j E^i$$

Which, if substituted inside the previous system, using that $\partial_i(H^i, E^i) = 0$ gives back the already well known wave equation

$$\begin{aligned} \frac{1}{c^2}\frac{\partial^2 E^i}{\partial t^2} - \partial^j \partial_j E^i = \Box E^i = 0 \\ \frac{1}{c^2}\frac{\partial^2 H^i}{\partial t^2} - \partial^j \partial_j H^i = \Box H^i = 0 \end{aligned} \tag{13.2}$$

The solution will be what's known as an *electromagnetic wave*, a wave moving in space composed by both an electric and magnetic field. From the wave equation we can already say that it's moving at speed c, as it should. To be precise it's what we know as *light*.

§§ 13.1.1 Electromagnetic Waves in Dielectric Media

In case that we're dealing with the movement of this wave in dielectric or magnetic media, we must remember that the fields *will be different*. The equations will be the same in shape, but will need some tweaking.
The tweaking is not actually in the fields, but in the constants. Given the speed of an electromagnetic wave in the vacuum is c, where

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$$

In a media, we will have the *speed of propagation* of the waves, u, given by the substitution of $\mu_0\epsilon_0$ with $\mu\epsilon$ (remember that $\mu = \mu_r\mu_0$ and $\epsilon_r\epsilon_0$)

$$u = \frac{1}{\sqrt{\mu\epsilon}} = \frac{1}{\sqrt{\mu_r\mu\epsilon_r\mu_0}} = \frac{c}{\sqrt{\mu_r\epsilon_r}} \tag{13.3}$$

From this we define the *index of refraction* of a medium, n as

$$n = \frac{c}{u} = \sqrt{\mu_r\epsilon_r} \tag{13.4}$$

So that, the propagation speed can be written also

$$u = \frac{c}{n}$$

And the wave equations become

$$\begin{cases} \frac{n^2}{c^2}\frac{\partial^2 E^i}{\partial t^2} - \partial^j \partial_j E^i = 0 \\ \frac{n^2}{c^2}\frac{\partial^2 H^i}{\partial t^2} - \partial^j \partial_j H^i = 0 \end{cases} \tag{13.5}$$

In general, it's clear that both fields solve the same basic wave equation

$$\partial^i \partial_i f = \frac{1}{u^2}\frac{\partial^2 f}{\partial t^2}$$

The solution of this equation (in terms of complex exponentials) is

$$f(x^i, t) = e^{i(k^i x_i - \omega t)} \tag{13.6}$$

What happens when we apply the same differential operators that appear in Maxwell's equation to this solution?
Well we simply take the derivatives and see that we're actually dealing with an eigenfunction of the differential operator, with eigenvalues

$$\begin{cases} \partial_i f = ik_i f \\ \partial_t f = -i\omega f \end{cases} \tag{13.7}$$

Thus, in terms of electromagnetic fields we can write Maxwell's equations as follows

$$\begin{cases} k^i E_i = 0 \\ \epsilon^i{}_{jk} k^j E^k = \mu\omega H^i \\ k^i B_i = 0 \\ \epsilon^i{}_{jk} k^j H^k = -\epsilon\omega E^i \end{cases} \tag{13.8}$$

It's clear that these three vectors compose a mutually orthogonal triad. Therefore, if we consider the magnitudes of these vectors from the third equation, we get

$$H = \frac{\epsilon\omega}{k} E = \epsilon u E \tag{13.9}$$

Where we used $\omega/k = u$. If we rewrite them in terms of the refraction index $n = c/u$, defined $Z_0 = \sqrt{\mu_0/\epsilon_0}$ the "free space impedance" we also have

$$H = \frac{n}{Z_0} E$$

§§ 13.1.2 Energy Flow

As we saw before, the Poynting vector is defined as the cross product of E with H, and is a representation of the flux of electromagnetic energy per unit area. If we take an electromagnetic field which solves the wave equation, as

$$\begin{aligned} E^i &= E_0^i \cos\left(k^i x_i - \omega t\right) \\ H^i &= H_0^i \cos\left(k^i x_i - \omega t\right) \end{aligned} \tag{13.10}$$

We have that the Poynting vector for such fields is

$$S^i = \epsilon^i{}_{jk} E^j H^k = S_0^i \cos^2\left(k^i x_i - \omega t\right)$$

Where $S_0 = E_0 \times H_0$. Since, as we saw before k^i is perpendicular to both E and H, it must be parallel to the Poynting vector. Taken the average (remember that $\left\langle \cos^2(\theta) \right\rangle = 1/2$) we get

$$\left\langle S^i \right\rangle = \frac{1}{2} S_0^i = I \frac{k^i}{k} = I\hat{n} \tag{13.11}$$

Where I is known as the *irradiance* and has, clearly, value

$$I = \frac{1}{2} E_0 H_0 = \frac{n}{2Z_0} E_0^2$$

This is known also as *Malus' Law*.
The irradiance, defined as before, is nothing more than the rate of flow of energy, and it's proportional to the square of the amplitude of the electric field.
For isotropic media, then the direction of the energy flux is defined by both S^i and k^i.

§ 13.2 Polarization of Electromagnetic Waves

§§ 13.2.1 Linear Polarization

Definition 13.2.1 (Linear Polarization)**.** Consider a general plane harmonic wave with the following solution to Maxwell's equations:

$$E^i = E_0^i e^{i(k^i r_i - \omega t)}$$
$$H^i = H_0^i e^{i(k^i r_i - \omega t)}$$

If both $E_0^i,\ H_0^i \in \mathbb{R}^3$ are constant (real constant vectors), then the wave is said to be *linearly polarized*.

Definition 13.2.2 (Polarizer)**.** A *polarizer* is an optical element that generates linearly polarized light. One of such instrument is the Polaroid filter.
A polarizer is said to have two main axes, one transmission axis and one blocking axis. The transmission axis is the one that will let the component of the E^i field pass, therefore by definition polarizing the light wave. If the polarizer is *completely* transparent to the incoming light parallel to the transmission axis it's known as an *ideal polarizer*

Consider now some randomly polarized light that passes through an ideal linear polarizer, and suppose that it arrives such that E^i arrives at an angle θ with respect to the transmission axis. The transmitted magnitude then will simply be the projection of E^i onto the axis.
Called τ our axis, then

$$E_\tau = E\cos\theta$$

And since $I \propto E^2$, the transmitted intensity is

$$I_\tau = I\cos^2\theta$$

By the same reasoning as before, for unpolarized light then, since $\langle\cos^2\theta\rangle = 1/2$ we have that, the transmitted intensity through this polarizer is exactly half the incoming intensity, i.e.

$$I_\tau = \frac{1}{2}I_u$$

Definition 13.2.3 (Partial Polarization)**.** A light wave is said to be *partially polarized* if it's made by a mixture of polarized and unpolarized light. The *degree of polarization* P is defined as the ratio between the intensity of polarized light I_p and the total intensity I_T (the sum of the intensity of polarized and unpolarized light)

$$P = \frac{I_p}{I_T}$$

For the special case of partial *linear* polarization, it can be calculate as the difference between the maximum intensity minus the minimum intensity of light, normalized with respect to the total intensity

$$P_{pl} = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}$$

It's clear that $P \in [0, 1]$, where $P = 0$ indicates completely unpolarized light and $P = 1$ indicates completely polarized light

§§ 13.2.2 Circular and Elliptical Polarization

Consider two orthogonal waves $\psi_1,\ \psi_2$ with equal amplitudes $A_1 = A_2$. Due to their orthogonality there's a $\pi/2$ phase shift between the two, and can be written as:

$$\psi_1 = E_0 \cos(kz - \omega t)\,\hat{x}^i \psi_2 \quad = E_0 \sin(kz - \omega t)\,\hat{y}^i \tag{13.12}$$

Using the principle of superposition, then, the total field (i.e., total wave) is then

$$E^i = E_0^i \left(\cos(kz - \omega t)\,\hat{x}^i + \sin(kz - \omega t)\,\hat{y}^i\right) \tag{13.13}$$

This is clearly a solution to the Maxwell equation $\Box E^i = 0$ and therefore can be seen as a valid electromagnetic wave, especially one which rotates around the z axis with angular frequency ω. The direction of rotation defines if it's a *right circularly polarized wave* $(-y)$ or a *left circularly polarized wave* $(+y)$.
Using complex notation and noting that the $\pi/2$ shift can be defined with a multiplication by i of the y component, hence

$$E^i = \left(\hat{x}^i \pm i\hat{y}^i\right) e^{i(kz - \omega t)} \tag{13.14}$$

The circular polarization is just a special case of an elliptical polarization, where $A_1 \neq A_2$, where

$$E_0^i = E_0 \hat{x}^i + iE_1 \hat{y}^i \quad \Longrightarrow \quad E^i = E_0^i e^{i(kz - \omega t)} \tag{13.15}$$

It's also clear how this notation generalizes polarization altogether, in fact, if E_0^i is a real vector we get back our usual linear polarization that we treated before.
A question arises now, what optical element can create elliptical polarization from a general light wave? The answer is *quarter-wave plates*.

Definition 13.2.4 (Quarter Wave Plate)**.** A *quarter wave plate*, also known as a $\lambda/4$*-plate* is an optical element made of two different refracting transparent crystals, which combined give a polarizer with two different transmission axes.
The axis with the highest refraction index (n_1) is known as the *fast axis*, while the axist with the smallest refraction index (n_2), is known as the *slow axis*.
The instrument is built such that the optical thickness (nd, where d is the thickness), obeys the following relation

$$n_1 d - n_2 d = \frac{\lambda_0}{4}$$

With λ_0 being the vacuum wavelength of the considered wave. Solving for d, we then have the quarter-wave plate relation

$$d = \frac{\lambda_0}{4(n_1 - n_2)} \tag{13.16}$$

§§ 13.2.3 Jones Calculus

Consider a general plane harmonic electromagnetic wave, i.e.

$$E_0^i = E_{0x}\hat{x}^i + E_{0y}\hat{y}^i \tag{13.17}$$

Since in the most general case $E_{0x}, E_{0y} \in \mathbb{C}$ we might choose to use the $\hat{x}^i, \hat{y}^i$ basis and work directly in $\mathbb{C}^2$. We can define then the *Jones Vector* as follows

Definition 13.2.5 (Jones Vector)**.** We define a *Jones vector* as a vector $v^i \in \mathbb{C}^2$ defined as follows:

$$\begin{pmatrix} E_{0x} \\ E_{0y} \end{pmatrix} = \begin{pmatrix} |E_{0x}|e^{i\phi_x} \\ |E_{0y}|e^{i\phi_y} \end{pmatrix} \tag{13.18}$$

Then, this vector is the most general way of defining a plane harmonic electromagnetic wave. As we have seen before, then we have

1. For linearly polarized waves in one direction (x or y):

$$A\begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad A\begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{13.19}$$

2. For linearly polarized waves (45 degrees):

$$A\begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{13.20}$$

3. For left-circularly polarized waves:

$$A\begin{pmatrix} 1 \\ i \end{pmatrix} \tag{13.21}$$

4. For right-circularly polarized waves:

$$A\begin{pmatrix} 1 \\ -i \end{pmatrix} \tag{13.22}$$

This representation in terms of complex vectors is really useful for calculating the polarization (and amplitude) of the wave resulting from the superposition of two differently polarized waves. I.e.: suppose you have a RCP (Right Circularly Polarized) wave and a LCP (Left Circularly Polarized) wave with unitary amplitude in some units. After the superposition we get:

$$\begin{pmatrix} 1 \\ -i \end{pmatrix} + \begin{pmatrix} 1 \\ i \end{pmatrix} = 2\begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Then, RCP+LCP=x-Linear polarization.
The action of optical elements can then be seen as the action of a matrix upon these vectors. Then, as an example, we can represent a linear polarizer on the x direction as:

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{13.23}$$

And so on for other optical elements.
In general, we have:

1. Linear Polarizers
   - Horizontal polarizers
   $$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{13.24}$$
   - Vertical polarizers
   $$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{13.25}$$
   - $\pm 45^\circ$ polarizers
   $$\frac{1}{2}\begin{pmatrix} 1 & \pm 1 \\ \pm 1 & 1 \end{pmatrix} \tag{13.26}$$
2. $\lambda/4$ Plates
   - Vertical fast axis
   $$\begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \tag{13.27}$$
   - Horizontal fast axis
   $$\begin{pmatrix} 1 & 0 \\ 0 & -i \end{pmatrix} \tag{13.28}$$
   - $\pm 45^\circ$ fast axis
   $$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & \pm i \\ \pm i & 1 \end{pmatrix} \tag{13.29}$$
3. $\lambda/2$ Plates
   - Horizontal or vertical fast axis
   $$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{13.30}$$
4. Retarders
   $$\begin{pmatrix} e^{i\phi} & 0 \\ 0 & e^{i\phi} \end{pmatrix} \tag{13.31}$$
5. Phase Changers
   $$\begin{pmatrix} e^{i\phi_x} & 0 \\ 0 & e^{i\phi_y} \end{pmatrix} \tag{13.32}$$
6. Circular Polarizers

- Left Circular Polarizer

$$\frac{1}{2}\begin{pmatrix}1 & -i\\ i & 1\end{pmatrix} \tag{13.33}$$

- Right Circular Polarizer

$$\frac{1}{2}\begin{pmatrix}1 & i\\ -i & 1\end{pmatrix} \tag{13.34}$$

In general, for a train of optical devices $(A^i_j)_1, \cdots, (A^i_j)_n$, where A_i are complex matrices, we have that the resulting wave R^i will be simply the following product, given I^i as our incident wave:

$$\prod_{\alpha=1}^{n} (A^i_j)_\alpha I^j = R^i \tag{13.35}$$

§§ 13.2.4 Orthogonal Polarization

Given two waves $E^i_1, E^i_2 \in \mathbb{C}^2$, they're said to be *orthogonally polarized* if, the complex scalar product between the two is null.
I.e.

$$\langle E^i_1, E^j_2 \rangle = E^i_1 \overline{E}^2_i = 0 \tag{13.36}$$

14 REF-old

§ 14.1 Reflection and Refraction

§§ 14.1.1 Snell's Law

The experimental effect of reflection and refraction can be studied in two ways mainly:

1. Using wave mechanics
2. Using electromagnetism

The first treatment is pretty simple to guess. Experimentally, what we see is an incoming (incident) wave coming from a place with some refraction index n_1, passing through a boundary plane, before reaching the second place where there is a new different refraction index $n_2 \neq n_1$. Experimentally we might consider the final result as three waves, ψ, the incident wave, ψ_{rr}, the refracted wave, and ψ_{re}, the reflected wave.
Being all these waves, we can write them in terms of solutions to the wave equation in absence of sources, complex exponentials.
Ignoring the amplitude factor we have

$$\begin{aligned} \psi(x^i,t) &= e^{i(k^i x_i - \omega t)} \\ \psi_{rr}(x^i,t) &= e^{i(k^i_{rr} x_i - \omega t)} \\ \psi_{re}(x^i,t) &= e^{i(k^i_{re} x_i - \omega t)} \end{aligned} \tag{14.1}$$

Since ωt is the same for every wave (as it should) we must have

$$k^i x_i = k^i_{rr} x_i = k^i_{re} x_i$$

And, if we put ourselves in the reference frame of the incidence plane (perpendicular to the boundary plane between the two zones), calling the angle of incidence θ, the angle of reflection θ' and the angle of refraction ϕ, this condenses to

$$k \sin\theta = k_{re} \sin\theta' = k_{rr} \sin\phi \tag{14.2}$$

By definition, k and k_{re} are traveling in the same medium with n_1 as the refraction index, hence $k = k_{re}$ and we get through the previous equality

$$\theta' = \theta$$

We're left with a single equation then

$$k \sin\theta = k_{rr} \sin\phi$$

By definition of k, we can also say

$$\frac{k_{rr}}{k} = \frac{\omega/u_2}{\omega/u} = \frac{\omega n_2}{c}\frac{c}{\omega n_1} = \frac{n_2}{n_1} = n$$

I.e., the ratio between the modulus of the wavevector of the incoming wave and the modulus of the wavevector of the refracted wave is equal to the ratio of the two refraction indexes, aso known as the *average refraction index* n.
Combining it all we get *Snell's law*

$$\sin\phi = n \sin\theta \tag{14.3}$$

§§ 14.1.2 Fresnel Equations

Snell's law can let us calculate with ease the reflection and refraction angles, but by definition it doesn't convey any information about the magnitudes of the waves. For this, we need electromagnetic theory. Using the second equation of (13.8) we know that the magnetic vectors are orthogonal both to the electric and wavevector fields, and they're equal to

$$H^i = \frac{1}{\mu\omega}\epsilon^i{}_{jk} k^j E^k \tag{14.4}$$

Consider now two possible cases.

1. TE polarization (or S, σ polarization), where the E^i field is perpendicular to the incidence plane
2. TM polarization (or P, π polarization), where the H^i field is perpendicular to the incidence plane (hence E^i is parallel to the plane)

Considering the two section as dielectrics, it's possible to use the boundary conditions for E and H, and knowing how the normal and tangential components behave while passing through the plane we can have more information on the amplitude of the two resulting waves (reflected and refracted).
For TE (S, σ) polarization we therefore have, remembering that the tangential component of the electric field is conserved

$$\begin{cases} E + E_{re} = E_{rr} \\ H_{re}\cos\theta - H\cos\theta = -H_{rr}\cos\phi \\ -kE\cos\theta + k_{re}E_{re}\cos\theta = -k_{rr}E_{rr}\cos\phi \end{cases} \tag{14.5}$$

While for TM (P, π) polarization, we have

$$\begin{cases} H - H_{re} = H_{rr} \\ E\cos\theta + E_{re}\cos\theta = E_{rr}\cos\phi \\ kE - k_{re}E_{re} = k_{rr}E_{rr} \end{cases} \tag{14.6}$$

Solving the system of equations, we can define *four* coefficients, two for TE polarization and one for TM polarization. Indicating them as $t_{s/p}$ and $r_{s/p}$ (respectively s/p-polarization transmission coefficient

and s/p-reflection coefficient), as the ratio of the intensity of the transmitted/reflected electric field with the initial incident field, i.e.

$$\begin{aligned} t_s &= \left(\frac{E_{rr}}{E}\right)_{TE} \quad r_s = \left(\frac{E_{re}}{E}\right)_{TE} \\ t_p &= \left(\frac{E_{rr}}{E}\right)_{TE} \quad r_p = \left(\frac{E_{re}}{E}\right)_{TM} \end{aligned} \tag{14.7}$$

Solving the equations, we get for the r coefficient

$$\begin{aligned} r_s &= \frac{\cos\theta - n\cos\phi}{\cos\theta + n\cos\phi} \\ r_p &= \frac{\cos\phi - n\cos\theta}{n\cos\theta + \cos\phi} \end{aligned} \tag{14.8}$$

Where, as usual, $n = n_2/n_1$ is the average refraction index.
A special case comes from *normal incidence*, i.e. when $\theta = \phi = 0$. The two coefficients become

$$r_s = r_p ? \frac{1-n}{1+n} \tag{14.9}$$

It's clear now, since n can be also greater than one, that two main cases must exists, one where $n < 1$ and therefore the phase mustn't change, and one where $n > 1$, where phase must change by π. It's easy to believe this is true in the realm of complex numbers, where we can write $e^{i\pi} = -1$, which gives the aforementioned phase change by π. The second case, where $n > 1$ will be treated with more detail later in the section on *total internal reflection*.
With these coefficients, using Snell's law, we can get *Fresnel's equations*, which give us a way to calculate the amplitudes of the electromagnetic waves which get reflected/refracted. Hence, using $n = \sin\theta/sin\phi$ and cleaning up what we get via trigonometric identities, we have

$$\begin{aligned} r_s &= -\frac{\sin(\theta-\phi)}{\sin(\theta+\phi)} \\ t_s &= \frac{2\cos\theta\sin\phi}{\sin(\theta+\phi)} \\ r_p &= -\frac{\tan(\theta-\phi)}{\tan(\theta+\phi)} \\ t_p &= \frac{2\cos\theta\sin\phi}{\sin(\theta+\phi)\cos(\theta-\phi)} \end{aligned} \tag{14.10}$$

Or, by eliminating ϕ using Snell's law

$$\begin{aligned} r_s &= \frac{\cos\theta - \sqrt{n^2 - \sin^2\theta}}{\cos\theta + \sqrt{n^2 - \sin^2\theta}} \\ r_p &= \frac{\sqrt{n^2 - \sin^2\theta} - \cos\theta}{\cos\theta - \sqrt{n^2 - \sin^2\theta}} \end{aligned} \tag{14.11}$$

Defined these coefficients, for what we have already said about intensity and by the shape of these coefficients, we can define a new parameter as follows

Definition 14.1.1 (Reflectance). Given the reflection coefficients (for s or p polarization) we define the *reflectance* as the fraction of light energy that gets reflected, i.e.

$$\begin{aligned} R_s &= |r_s|^2 = \left|\frac{E_{re}}{E}\right|^2_{TE} \propto I_r^{TE} \\ R_p &= |r_p|^2 = \left|\frac{E_{re}}{E}\right|^2_{TM} \propto I_r^{TM} \end{aligned} \tag{14.12}$$

For normal incidence, therefore, it just reduces to the following

$$R_s = R_p = \left(\frac{n-1}{n+1}\right)^2 \tag{14.13}$$

§ 14.2 Total Internal Reflection

§§ 14.2.1 External and Internal Reflection

Going back to our two possible cases of outcome for reflection in normal incidence, we can define the following

1. Internal reflection, for $n > 1$
2. External reflection, for $n < 1$

It's clear from Fresnel's equations that when $n > 1$, all values of θ are possible and both r and t coefficients are real, while for $n < 1$ there exists a θ_c, known as the *critical angle* for which $\sin\theta_c = n$, where for $\theta < \theta_c$ transmission *can happen*.
For $\theta > \theta_c$ the coefficients $r_s, r_p \in \mathbb{C}$ and trasmission *doesn't happen*, since it cannot. Here, in this case, since $\sin\theta > n$, we get

$$\begin{aligned} r_s &= \frac{\cos\theta - i\sqrt{\sin^2\theta - n^2}}{\cos\theta + i\sqrt{sin^2\theta - n^2}} \\ r_p &= \frac{i\sqrt{sin^2\theta - n^2} - n^2\cos\theta}{n^2\cos\theta + i\sqrt{\sin^2\theta - n^2}} \end{aligned} \tag{14.14}$$

Note that, by definition then:

$$R_s = R_p = \overline{r}_s r_s = \overline{r}_p r_p = 1 \tag{14.15}$$

Also, if it wasn't clear, the critical angle can be calculated with ease as follows

$$\theta_c = \arcsin(n) \tag{14.16}$$

Hence we are in a regime of so called *total reflection*, or, to be precise, *total internal reflection*, which we will call *TIR* for ease.
It's possible to use TIR experimentally for transmitting light through continous fibers (see fiber optics), and create *light circuits*.
Consider a single fiber as a solid cylindrical dielectric, immersed in a medium for which $n_{ext} < n_{int}$.

We find ourselves in the case of $n < 1$, and if the angle of incidence of light to the fiber optic α follows $\alpha > \theta_c$, the light is trapped inside the dielectric in a regime of TIR.
By definition, then

$$\alpha = \arcsin\left(\sqrt{n_{int}^2 - n_{ext}^2}\right) \tag{14.17}$$

§§ 14.2.2 Brewster Angle

For p (or TM) polarization, there exists another "critical angle" for which $r_p = 0$. From the Fresnel formula in TIR regime for r_p, this angle, known as *Brewster angle* or *polarization angle* θ_B is defined as

$$\theta_B = \arctan(n) \tag{14.18}$$

For $\theta = \theta_B$ all unpolarized light gets refracted with TE (or s) polarization. ($r_p = 0$, there is no polarization)
A cool application of this principle is the *Brewster window*, an optical element for which the incoming light is incident at $\theta = \theta_B$ on some glass plate with two parallel faces. No light gets reflected, and what gets refracted is TE polarized.

§§ 14.2.3 Evanescent Wave

Although the electromagnetic wave in TIR regime gets reflected completely, there exists an electromagnetic field beyond the boundary of TIR, this field defines what's known as the *evanescent wave*. Consider the mathematical function which defines a refracted (transmitted) wave:

$$E_t^i = E_{rr}^i e^{i(k_{rr}^i r - \omega t)}$$

Choosing a reference system with the x axis orthogonal to the transmission axis we have

$$k_{rr}^i r_i = k_{rr} x \sin\phi - k_{rr} y \cos\phi$$

Using Snell's law for writing $\cos\phi$ in terms of $\sin\theta$ we have, simplifying k_{rr} to k

$$k^i r_i = kx \sin\phi - iky\sqrt{\frac{\sin^2\theta}{n^2} - 1} \tag{14.19}$$

Inserting it back to our transmitted E field, we get, given the following definitions

$$\begin{cases} \alpha = k\sqrt{\dfrac{\sin^2\theta}{n} - 1} \\ k_1 = \dfrac{k}{n}\sin\theta \end{cases}$$

An evanescent wave decaying on the y direction

$$E_t^i = E_{rr}^i e^{-\alpha|y|} e^{i(k_1 x - \omega t)} \tag{14.20}$$

§§ 14.2.4 Phase Changes

In the case of TIR, as we implied before, the complex nature of the reflection coefficients imply a phase change in function of the incidence angle.
By definition in TIR $|r_s| = |r_p| = 1$, and since $r_s, r_p \in \mathbb{C}$, by definition of complex angle we write both in terms of complex exponentials as follows

$$\begin{cases} r_s = e^{i\delta_s} = \dfrac{ae^{-i\alpha}}{ae^{i\alpha}} \\ r_p = e^{i\delta_p} = \dfrac{be^{-i\beta}}{be^{i\beta}} \end{cases} \tag{14.21}$$

The arguments of the two complex numbers, δ_s, δ_p, for how we defined them are exactly equal to 2α and 2β for how we defined them.
In terms of tangents, we have therefore

$$\tan\alpha = \tan\frac{\delta_s}{2}$$
$$\tan\beta = \tan\frac{\delta_p}{2}$$

Defining the tangent in terms of r and its complex conjugate we have, since

$$z + \overline{z} = 2\mathfrak{Re}(z)$$
$$z - \overline{z} = 2i\mathfrak{Im}(z)$$

That

$$\begin{aligned} \tan\alpha &= \frac{\mathfrak{Im}(r_s)}{\mathfrak{Re}(r_s)} = \frac{\sqrt{\sin^2\theta - n^2}}{\cos\theta} \\ \tan\beta &= \frac{\mathfrak{Im}(r_p)}{\mathfrak{Re}(r_p)} = \frac{\sqrt{\sin^2\theta - n^2}}{n^2\cos\theta} \end{aligned} \tag{14.22}$$

Defining the *phase difference* Δ as follows

$$\Delta = \delta_p - \delta_s$$

We have

$$\tan\Delta = \frac{\cos\theta\sqrt{\sin^2\theta - n^2}}{\sin^2\theta} \tag{14.23}$$

§ 14.3 Reflection Matrix

Using Jones calculus in the regime of reflection and refraction, we can imagine the s and p polarization components of the electromagnetic wave as a basis for $\mathbb{C}^2$, with the most general vector given by (p, s). Here we can define a (isotropic) *reflection matrix* R^i_j as:

$$R^i_j = \begin{pmatrix} -r_p & 0 \\ 0 & r_s \end{pmatrix} \tag{14.24}$$

And a *transmission matrix* T_j^i analogously as

$$T_j^i = \begin{pmatrix} t_p & 0 \\ 0 & t_s \end{pmatrix} \tag{14.25}$$

In the various regimes of reflection this matrix can be more or less complex.
Taken the case of normal incidence, as we saw we have

$$r_s = r_p = \frac{1-n}{1+n}$$

And therefore

$$R_j^i = \frac{1-n}{1+n} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \tag{14.26}$$

For what's known as *near grazing incidence* (i.e. when $r_p = r_s$ for external reflection and $r_p = -r_s$ for internal reflection) we have

$$R_j^i = \pm\delta_j^i \tag{14.27}$$

While for TIR, where $r_s = e^{i\delta_s}$ and $r_p = e^{i\delta_p}$, we have

$$R_j^i = \begin{pmatrix} e^{-i\delta_p} & 0 \\ 0 & e^{-i\delta_s} \end{pmatrix}$$

Note that, when a vector gets reflected in TIR regime, we have the following result:

$$\begin{pmatrix} A' \\ B' \end{pmatrix} = \begin{pmatrix} e^{-i\delta_p} & 0 \\ 0 & e^{-i\delta_s} \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} = e^{-i\delta_p} \begin{pmatrix} A \\ Be^{i\Delta} \end{pmatrix} \tag{14.28}$$

I.e., the p component of the (generally elliptically polarized wave) gets a phase change by exactly $\Delta = \delta_p - \delta_s$

15 INT-old

§ 15.1 Coherent Waves

§§ 15.1.1 Interference

Consider two interacting waves with electric fields $E^i_{(1)}$ and $E^i_{(2)}$. In general these two waves are solutions of the Maxwell wave equations, therefore, the most general shape they can have is the following:

$$\begin{cases} E^i_{(1)} = E^i_1 e^{ik^i_1 r_i - i\omega t + i\phi_1} \\ E^i_{(2)} = E^i_2 e^{ik^i_2 r_i - i\omega t + i\phi_2} \end{cases} \tag{15.1}$$

Where $\phi_1, \phi_2 \in [0, 2\pi]$ are two general phase factors.

Definition 15.1.1 (Mutually Coherent Waves)**.** Given two general waves, they're said to be *coherent* if and only if, given the phases ϕ_1 and ϕ_2, we have

$$\phi_1 - \phi_2 = k \tag{15.2}$$

Where $k \in \mathbb{R}$ is a constant real number.

Getting back to our interference pattern, we have that the total field that will be measured must be a linear superposition of these two general field. Therefore, the measured total intensity I is, (considering two linearly polarized waves, for ease of calculation))

$$I = \|E\|^2 = \left(E^i_{(1)} + E^i_{(2)}\right)\overline{\left(E^{(1)}_i + E^{(2)}_i\right)} = E^2_1 + E^2_2 + 2\mathfrak{Re}\left(E^i_{(1)}\overline{E^{(2)}_i}\right)$$

Evaluating the third addendum on the previous expression, using the general wave solution (15.1), we get

$$2\mathfrak{Re}\left(E^i_{(1)}\overline{E^{(2)}_i}\right) = 2E^i_1 E^2_i \mathfrak{Re} e^{i(k^i_1 - k^i_2)r_i + i(\phi_1 - \phi_2)}$$

Evaluating the right hand side of the previous equation, we have that the intensity will in general be calculated (for linearly polarized waves) as

$$I = I_1 + I_2 + 2E^i_1 E^2_i \cos\left[\left(k^i_1 - k^i_2\right) r_i + i(\phi_1 - \phi_2)\right] \tag{15.3}$$

Where I_1 and I_2 are the partial intensities of the single waves. The third term, is known as the *interference term*, which, since it's comprised between $[-1, 1]$, makes the intensity oscillate between $I_1 + I_2 \leq I \leq I_1 + I_2$ depending on

$$\theta(r^i) = i(k_1^i - k_2^i)r_i + i(\phi_1 - \phi_2) \tag{15.4}$$

The result are interference fringes. It's also clear that this can happen only and only if the waves are mutually coherent. If they weren't $\phi_1 - \phi_2$ would vary randomly, and, by definition

$$\left\langle \cos\left(\theta(r^i)\right)\right\rangle = 0$$

Which implies that the interference term is in general null, and $I = I_1 + I_2$ always.
Interference fringes can be obtained also through spatial variations, since the argument of the cosine depends both on the phase difference and the distance r.
Interference fringes also cannot appear in the case of two orthogonally polarized waves, because by definition their scalar product is zero, eliminating the interference term, independently from their mutual coherence.

§§ 15.1.2 Young's Double Slit Experiment

In 1802 dr. Young et al. managed to build an experiment which managed to show the interference pattern caused by the interference of two coherent electromagnetic waves.
The experimental setup is pretty simple, a coherent source (point-like) S emits a single electromagnetic wave, which passes through two slits S_1 and S_2, placing a detector (in case of light a simple wall is enough), it's possible to observe regular peaks of intensity.
Theoretically it all depends on the phase difference between the two waves that come out of each slit. Taken a random point P on the screen, we can say that it will be at a distance d_1 from S_1 and d_2 from S_2. The phase difference will obviously depend on the difference of the two distances from the chosen point.
For convenience, we search for the points of max intensity, (i.e., when $\cos\theta(r) = 1$ and the interference term is the maximum possible). Due to coherence between the two fields, we can say without any problem that

$$\theta(r) = k^i r_i = k_0(d_2 - d_1)$$

Hence, our intensity minimums will be

$$\max(I) = I|_{\cos\theta(r)=1} \implies k_0(d_2 - d_1) = \pm 2n\pi, \quad n \in \mathbb{N}$$

Remembering that $k_0 = 2\pi\lambda^{-1}$, we get the final simple result

$$|d_2 - d_1| = n\lambda \tag{15.5}$$

I.e., the distance between the intensity peaks is strictly tied to the distance between the slits and the wavelength of the electromagnetic wave, where the successive peaks will always be at integer multiples of the wavelenght.
In general, for two slits distant h between each other, chosen a reference frame for which the origin is the middle point between the two slits and the screen is distant x from this, taken a general point (x, y) on it we will have that

$$|d_2 - d_1| = \sqrt{x^2 + \left(y + \frac{h}{2}\right)^2} + \sqrt{x^2 + \left(y - \frac{h}{2}\right)^2}$$

Therefore, solving the previous equation, searching for the intensity peaks, we have, after a second order approximation that

$$|d_2 - d_1| \approx x\left(1 + \frac{1}{2x^2}\left(y + \frac{h}{2}\right)^2\right) - x\left(1 + \frac{1}{2x^2}\left(y - \frac{h}{2}\right)^2\right) \approx n\lambda$$

And solving

$$y \approx \frac{nx\lambda}{h} \tag{15.6}$$

I.e., the distance between peaks is approximately an integer multiple of the distance from the slits times the wavelength, over the distance between the two slits themselves.

§§ 15.1.3 Michelson Interferometer

Another slightly more complex system which analyzes the interference phenomena, is the Michelson interferometer. Michelson's interferometer was developed in 1880 by Michelson et al.
It's known as a *division of amplitude* interferometer, for the simple reason that it divides the beam in two (ideally equal) parts after passing through a so called *beam splitter*. A beam splitter is a particular optical device, which divides an electromagnetic wave in two (ideally equal) parts, sending one in the transverse direction.
For building up this interferometer, the setup is a source of coherent electromagnetic waves, a beam splitter, two mirrors and a detector. Placed the beam splitter at the center of the setup, with one of the two mirrors M_2 placed at a distance d_2 from the splitter, and the other one, M_1 at a distance d_1 on the other transmission axis of the splitter. Opposite to the mirror there will be our detector D, which will measure the interference fringes caused by the interaction of the two beams when they end up interacting again before the reflection at the beam splitter.
The observed fringes directly depend on the optical path of the two mirrors, which is modified by the simple motion of one of the two.
This setup is equivalent to having two point-like sources distant $d = |d_2 - d_1|$ from each other, interfering and reaching a detector.
Due to the two waves being ideally identical due to the inner workings of the beam splitter, (i.e., in this case $k_1 = k_2$) the measured intensity at the detector will depend only on the phase shift induced by the optical distance traveled, and

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos(\phi_1 - \phi_2)$$

Again, the interference peaks will be when $\cos(\phi_1 - \phi_2) = 1$, i.e., when

$$\phi_1 - \phi_2 = k_0 d = \frac{2\pi d}{\lambda} \tag{15.7}$$

With intensity

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \tag{15.8}$$

§ 15.2 Partial Coherence

In reality, what we treated before is an approximation of reality. In fact, in general, two interfering electromagnetic waves *aren't necessarily coherent*, or at least they aren't always coherent.
By definition of coherence, this means that the phases of the signals vary randomly through time, which brings forward the idea of averaging the total amplitude.

Definition 15.2.1 (Time Average)**.** We define the *time average* of a periodic function $f(t)$, as the following improper integral

$$\langle f \rangle = \lim_{T\to\infty} \frac{1}{T} \int_0^T f(t)\mathrm{d}t \tag{15.9}$$

Note how the final result doesn't depend on time.

Applied to what we found for the amplitude, we then have

$$\langle I \rangle = I_1 + I_2 + 2\mathfrak{Re}\left(\left\langle E_1^i \overline{E_i^2} \right\rangle\right) \tag{15.10}$$

Where we conveniently defined $I_1 = \langle E_1^2 \rangle$ and $I_2 = \langle E_2^2 \rangle$.
Looking back at the previous experimental setups, it's clear how interference depends on the difference of length of the optical paths of the two interfering waves, due to our choice of time averaging we can see the difference as an added needed time to reach the detector, let's say τ, for which, if E_1 reaches it at a time t, E_2 will reach it at a time $t+\tau$. Considering that the interference is seen when both waves have reached the detector, we have that the interference term is the following

$$2\mathfrak{Re}\left(\left\langle E_1^i(t)\overline{E_i^2(t+\tau)} \right\rangle\right) \tag{15.11}$$

The integral inside the real part operator, can then be redefined as follow, as a new function depending on τ

Definition 15.2.2 (Correlation Function)**.** The *correlation function* $\Gamma_{12}(\tau)$ is defined as follows, forgetting the vectorial nature of electromagnetic waves for a second,

$$\Gamma_{12}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T E_1(t)\overline{E_2(t+\tau)}\,\mathrm{d}t \tag{15.12}$$

Clearly, $\Gamma_{12}(\tau) = \Gamma_{21}(\tau)$, and from this we can also define the *autocorrelation function* $\Gamma_{11}(\tau)$ and $\Gamma_{22}(\tau)$. It's also evident that

$$\begin{aligned} \Gamma_{11}(0) &= I_1 \\ \Gamma_{22}(0) &= I_2 \end{aligned} \tag{15.13}$$

Another, more comfortable definition is the following

Definition 15.2.3 (Degree of Spatial Coherence)**.** Defined the correlation function $\Gamma_{12}(\tau)$ as before, we define the *degree of spatial coherence* $\gamma_{12}(\tau)$ as the normalized correlation function

$$\gamma_{12}(\tau) = \frac{\Gamma_{12}(\tau)}{\sqrt{\Gamma_{11}(0)\Gamma_{22}(0)}} = \frac{\Gamma_{12}(\tau)}{\sqrt{I_1 I_2}} \tag{15.14}$$

Clearly, $-1 \leq \gamma_{12}(\tau) \leq 1$

Then, the irradiance or amplitude I, can be rewritten as follows

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\mathfrak{Re}\left(\gamma_{12}(\tau)\right) \tag{15.15}$$

In general, as we'll see later, γ_{12} is a complex-valued function periodic in τ. It's also obvious that the interference bands will appear only and only if $\gamma_{12}(\tau) \neq 0$.
Two waves will then said to be

1. Completely coherent, if $|\gamma_{12}(\tau)| = 1$
2. Completely incoherent, if $|\gamma_{12}(\tau)| = 0$
3. Partially coherent, if $0 \leq |\gamma_{12}(\tau)| \leq 1$

The minimum and maximum values of amplitude in a regime of partial or complete coherence then, can be found as follows

$$\begin{aligned} I_{max} &= I_1 + I_2 + 2\sqrt{I_1 I_2}|\gamma_{12}(0)| \\ I_{min} &= I_1 + I_2 - 2\sqrt{I_1 I_2}|\gamma_{12}(0)| \end{aligned} \tag{15.16}$$

Definition 15.2.4 (Fringe Visibility)**.** Another important value in the analysis of interference is the *visibility of fringes* $\mathcal{V}$, defined as follows

$$\mathcal{V} = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{15.17}$$

Or

$$\mathcal{V} = \frac{2\sqrt{I_1 I_2}|\gamma_{12}(0)|}{I_1 + I_2} \tag{15.18}$$

Note that, by definition, then if $I_1 = I_2 = I$, the fringe visibility is simply the absolute value of the degree of coherence

$$\mathcal{V} = |\gamma_{12}(0)|$$

This constant, by definition, is $0 \leq \mathcal{V} \leq 1$, where a value of 0 implies that there is no interference, while a value of 1 implies the maximum possible interference between the waves

§§ 15.2.1 Coherence Time and Coherence Length

In order to understand better the regime of partial coherence, we take a quasimonochromatic electromagnetic wave, with a random phase $\phi(t)$, which is a periodic step function which changes value every *coherence time* τ_0.
Suppose that this electromagnetic wave gets split at some point, and the two resulting waves travel two different optical paths long d_1, d_2, and rejoin at some point creating an interference pattern. In general, the two waves will have the following mathematical shape

$$E^i(t) = E_0^i e^{-i\omega t + i\phi(t)}$$

Indicating the amplitude $I_{1,2} = |E_0|^2$ we get that then, the *self-degree of coherence* is

$$\gamma(\tau) = \frac{\left\langle E^i(t) E_i(t+\tau)\right\rangle}{I} = \lim_{T\to\infty}\frac{1}{T}\int_0^T e^{-i\omega t + i\phi(t)} e^{i\omega(t+\tau) - i\phi(t+\tau)}\mathrm{d}t$$

Simplifying the exponentials inside we get

$$\gamma(\tau) = e^{i\omega\tau} \lim_{T\to\infty} \frac{1}{T} \int_0^T e^{i(\phi(t)-\phi(t+\tau))} \mathrm{d}t$$

For evaluating this integral we need to define the behavior of $\phi(t) - \phi(t+\tau)$. Firstly we know that it's a step function which varies randomly periodically with period τ_0, hence, the difference will be zero until we find ourselves with $\phi(t+\tau)$ (or vice-versa, $\phi(t)$) in a different step. Mathematically, for the first period

$$\begin{cases} \phi(t) - \phi(t+\tau) = 0 & 0 < t < \tau_0 - \tau \\ \phi(t) - \phi(t+\tau) = \Delta & \tau_0 - \tau < t < \tau_0 \end{cases}$$

With $\Delta \in [0, 2\pi]$ being a random value. Then, for a single period, the integral before becomes

$$\frac{e^{i\omega\tau}}{\tau_0} \left(\int_0^{\tau_0-\tau} \mathrm{d}t + \int_{\tau_0-\tau}^{\tau_0} e^{i\Delta} \mathrm{d}t \right) = \left(\frac{\tau_0 - \tau}{\tau_0} e^{i\omega\tau} + \frac{\tau}{\tau_0} e^{i\Delta} \right)$$

Time-averaging the result, we get the normalized autocorrelation function for these general waves. Remembering the randomness of $e^{i\Delta}$ we have that its average will be 0, leaving only the first part, which, simplified becomes

$$\gamma(\tau) = \begin{cases} \left(1 - \frac{\tau}{\tau_0}\right) e^{i\omega\tau} & \tau < \tau_0 \\ 0 & \tau \geq \tau_0 \end{cases} \tag{15.19}$$

In this case, $I_1 = I_2$ (by definition), hence $|\gamma| = \mathcal{V}$, i.e.

$$\mathcal{V} = |\gamma(\tau)| = \begin{cases} 1 - \frac{\tau}{\tau_0} & \tau < \tau_0 \\ 0 & \tau \geq \tau_0 \end{cases} \tag{15.20}$$

This clearly gives the name *correlation time* to τ_0, it's evident how the waves become completely uncorrelated after for $\tau \geq \tau_0$, and the fringes are not visible! Analogously, we can define a related length called *coherence length*, as follows

$$l_c = c\tau_0 \tag{15.21}$$

This value corresponds to the maximum length that of an uninterrupted wave train

§§ 15.2.2 Spectral Resolution of Finite Wave Trains

In nature, strictly monochromatic sources of electromagnetic waves do not exist, and the best we can hope for is a finite wave train with frequencies spread around some mean value $\langle\omega\rangle = \omega_0$.
By definition of finite wave train, there should be some way to analyze the relationship between the frequency spread (or line width) and the coherence of a source.
Fourier calculus comes in our help for this task. Called $f(t)$ our wavefunction we define the Fourier pair $g(\omega)$ as

$$g(\omega) = \hat{\mathcal{F}}[f]$$

In our case, the Fourier transform, transforms a wavefunction from the time space to the frequency space. As we should know from mathematics, the Fourier operator is invertible, and the frequency-time pair can be defined in integral form as follows

$$f(t) = \hat{\mathcal{F}}^{-1}[g] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} g(\omega) e^{-i\omega t} \mathrm{d}\omega$$
$$g(\omega) = \hat{\mathcal{F}}[f] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(t) e^{i\omega t} \mathrm{d}t$$

Getting back to our analysis, consider $f(t)$ being the wavefunction of some finite wave train with coherence time τ_0. By definition then, we have

$$f(t) = \begin{cases} e^{-i\omega_0 t} & -\frac{\tau_0}{2} < t < \frac{\tau_0}{2} \\ 0 & |t| > \frac{\tau_0}{2} \end{cases} \tag{15.22}$$

Its Fourier transform is

$$g(\omega) = \mathcal{F}[f] = \frac{1}{\sqrt{2\pi}} \int_{-\frac{\tau_0}{2}}^{\frac{\tau_0}{2}} e^{i(\omega-\omega_0)t} \mathrm{d}t = \frac{1}{\sqrt{2\pi}} \left[\frac{e^{i(\omega-\omega_0)t}}{i(\omega-\omega_0)} \right]_{-\frac{\tau_0}{2}}^{\frac{\tau_0}{2}} \tag{15.23}$$

Evaluating the integral and remembering the complex exponential formula for $\sin(z)$, we have

$$\hat{\mathcal{F}}[f] = \frac{1}{i(\omega-\omega_0)\sqrt{2\pi}} \left(e^{\frac{i(\omega-\omega_0)\tau_0}{2}} - e^{-\frac{i(\omega-\omega_0)\tau_0}{2}} \right) = \frac{1}{\omega-\omega_0} \sqrt{\frac{2}{\pi}} \sin\left(\frac{(\omega-\omega_0)\tau_0}{2} \right) \tag{15.24}$$

Or, using for convenience the sinc function we have

$$\hat{\mathcal{F}}[f] = \frac{1}{\omega_0-\omega} \frac{2}{\sqrt{\pi}} \mathrm{sinc}\left(\frac{(\omega-\omega_0)\tau_0}{2} \right)$$

Definition 15.2.5 (Power Spectrum)**.** The utility of defining the Fourier transform of a train waves comes in handy for defining a new entity, the so called *Power spectrum* of the wave, commonly indicated as $G(\omega)$, defined as the square modulus of the Fourier transform, i.e.

$$G(\omega) = \left| \hat{\mathcal{F}}[f] \right|^2 = |g(\omega)|^2 \tag{15.25}$$

Having defined the power spectrum of a wave, we have for our generic wave train then

$$G(\omega) = \frac{2}{\pi(\omega-\omega_0)^2} \sin^2\left(\frac{\omega-\omega_0}{2} \tau_0 \right) \tag{15.26}$$

The utility of defining the power spectrum comes from its mathematical properties, in fact we have $G(\omega) = 0$ if and only if $\omega = \omega_0$, and its maximum is, when

$$\max\left(G(\omega)\right) = G\left(\omega_M\right) \qquad \omega_M = \frac{\pi}{2\tau_0} \pm \omega_0 \tag{15.27}$$

Therefore, the width of the frequency distribution, $\Delta\omega$ is then

$$\Delta\omega = \frac{2\pi}{\tau_0} \tag{15.28}$$

From this, in terms of frequency, we have that

$$\Delta\nu = \frac{1}{t_0}$$

I.e., a sequence of wave trains lasting τ_0 will have the same exact power spectrum of the single pulse. In general, if the pulses also have random τ_0, we can say without loss of generality that it's approximately equal to $\langle\tau_0\rangle^{-1}$.
Reasoning in the opposite direction, given a spectral source with line width $\Delta\nu$, the coherence time of the wave trains $\langle\tau_0\rangle$ can be estimated as follows

$$\langle\tau_0\rangle = \frac{1}{\Delta\nu}$$

Which, in terms of coherence length l_c it gives us

$$l_c = \frac{c}{\Delta\nu}$$

Using also that $\Delta\nu/\nu = |\Delta\lambda|/\lambda$, we get that

$$l_c = \frac{\lambda^2}{\Delta\lambda} \tag{15.29}$$

With $\Delta\lambda$ being the width of the spectral line in terms of wavelength.

§ 15.3 Spatial Coherence

In this section we will analyze a different case. Before we dealt only with two fields that reach a single fixed point in space, from here on we will study the coherence of two or more fields reaching two or more *different* points in space.
Suppose again a quasimonochromatic point-like source S and 3 receiving points P_i, with P_3 being on the same director from the point P_1 and P_2 disposed in a different director from the source at the same distance d from it.
The coherence for the fields E_1, E_3 will be known as the *Longitudinal Spatial Coherence*, while the coherence between E_1, E_2 will be known as the *Transverse Spatial Coherence*. It's already clear for what we said before that the longitudinal spatial coherence will depend only on how far is P_1 from P_3, called this distance d_{13} we can also say that it depends on the value

$$t_{13} = \frac{d_{13}}{c}$$

Compared to the coherence time τ_0. It's clear that for whatever $E_1(t)$ we have, $E_3(t)$ will have the same time dependence of E_1, with a retardation of t_{13}.
If $t_{13} << \tau_0$ the two fields will be coherent, while for $t_{13} >> \tau_0$ the two fields will be little or completely

not coherent between the two.
For the fields E_1 and E_2 instead the time dependence will be the same, hence they will always be completely mutually coherent ($t_{12} = 0$), but only if the source S is point-like. It's clear that in reality there are no real point sources, and for an extended source (i.e., with spatial resolution) the size of the source itself must also be accounted for.
In general these extended sources can be considered as a cluster of point sources, and therefore we can immediately consider the more general case of multiple sources (two for ease of calculation) with two measuring points P_1, P_2, distant $d_{1a}, d_{1b}, d_{2a}, d_{2b}$ from the sources, in a regime of transverse spatial coherence.
Called these two sources S_a, S_b, we can define the two fields E_1, E_2 as the sum of the two fields going from the $i-$th source to the $j-$th point. I.e.

$$\begin{aligned} E_1 &= E_{1a} + E_{1b} \\ E_2 &= E_{2a} + E_{2b} \end{aligned} \tag{15.30}$$

Noting that the fields emitted from the source S_a *will not* be coherent (actually, will be completely incoherent) with the fields emitted from the source S_b, the correlation function between the two fields at the two receiving points will be

$$\gamma_{12}(\tau) = \frac{\left\langle E_1(t)\overline{E_2}(t+\tau)\right\rangle}{\sqrt{I_1 I_2}} = \frac{1}{\sqrt{I_1 I_2}}\left(\left\langle E_{1a}(t)\overline{E_{2a}}(t+\tau)\right\rangle + \left\langle E_{1b}(t)\overline{E_{2b}}(t+\tau)\right\rangle\right) \tag{15.31}$$

If the two fields are generic train waves, we already know that the self coherence of such pulse is

$$\gamma(\tau) = \left(1 - \frac{\tau}{\tau_0}\right) e^{i\omega\tau}$$

And considered the two different optical paths with different travel times for the waves, then, in our case

$$\gamma_{12}(\tau) = \frac{1}{2}\gamma(\tau_a) + \frac{1}{2}\gamma(\tau_b)$$

Where, obviously, we have

$$\begin{aligned} \tau_a &= \frac{d_{2a} - d_{1a}}{c} + \tau \\ \tau_b &= \frac{d_{2b} - d_{1b}}{c} + \tau \end{aligned}$$

After some algebra which I won't do so bear with me here, we have

$$\left|\gamma_{12}(\tau)\right|^2 = \left(1 - \frac{\tau_a}{\tau_0}\right)\left(1 - \frac{\tau_b}{\tau_0}\right)\left(\frac{1 + \cos\left[\omega\left(\tau_b - \tau_a\right)\right]}{2}\right) \tag{15.32}$$

Where the approximation $\tau_a - \tau_b << \tau_a, \tau_b$ has been used.
It's clear then that the visibility and coherence of waves in such system also depends on the difference of travel times for the waves, from the sources to the points, also in a periodic manner.
This clearly means that at the two points the waves will show periodically interference (i.e., coherence, $\gamma_{12} \neq 0$) also if the two sources emit electromagnetic radiation which in which is completely incoherent. This is called a *periodic spatial dependence*.

Consider now the special case of having the two points disposed in a symmetrical manner with respect to the sources, hence with $d_{1a} = d_{1b}$. Then

$$\tau_b - \tau_a = \frac{d_{2a} - d_{1b}}{c} \approx \frac{sl}{2cd} \tag{15.33}$$

Where, s is the distance between the sources, l is the distances between the points, d is the mean distance between the points and the sources. Here we used the approximation $r >> s, l$.
Or, extending the calculations, we have

$$r_{2a,b} = \sqrt{\left(l \pm \frac{s}{2}\right)^2 + d^2} = d\sqrt{1 + \frac{1}{d^2}\left(l \pm \frac{s}{2}\right)^2}$$

Approximating the root using the same approximation idea we had before, we have

$$d_{2a,b} \approx d + \frac{1}{2d}\left(l^2 + \frac{s^2}{4} \pm ls\right) \implies \tau_b - \tau_a \approx \frac{r}{c} + \frac{l^2}{2dc} + \frac{s^2}{8dc} \pm \frac{ls}{2dc}$$

The experimental setup in this case is really similar to the Young interferometer's case, and the variation of lateral coherence between the two will be described by a periodic bell shaped curve.
The dips of this curve, will be then for $|\gamma_{12}|^2 = 0$. Considering the main peak at the center, these peaks will be reached at some distance l_t, which is approximately the length of the area of maximum coherence. Then

$$\frac{\omega l_t s}{2dc} = \pi$$

Using $\omega/c = k_0 = 2\pi/\lambda$, which implies $\omega = 2\pi c/\lambda$ we have

$$\frac{2\pi l_t s}{2\lambda d} = \pi \implies l_t = \frac{\lambda d}{s} \tag{15.34}$$

Or, in terms of angular separation $\theta_s \approx s/d$

$$l_t = \frac{\lambda}{\theta_s} \tag{15.35}$$

The value l_t is known as the *transverse coherence width*.

§ 15.4 Multiple Beam Interference

The most general treatment of coherence comes from the study of multi-beam interference. For the treatment of coherence here, the division of amplitude is needed, and one way of obtaining this is by using multiple reflection between two semi-reflecting parallel plates.
Set up the experiment, with the two parallel plates fixed at some distance d between each other, we send the initial beam with amplitude E_0 towards them. Due to their being semi-reflecting, at contact with the first one the beam will be reflected at some angle θ and amplitude $E_0 r$, while the second will travel through it with amplitude $E_0 t$.
The process continues when the beam hits the second semi-reflecting surface, which will again repeat the same process, sending towards the first surface a beam with amplitude $E_0 tr$ and transmitting a beam with amplitude $E_0 t^2$.

Studying the process for n iterations, we get the following succession of amplitudes for the beams found inside the two surfaces:

$$E_0t, E_0tr, E_0tr^2, \cdots, E_0tr^n$$

While, the transmitted ones *outside* the second surface, will follow the following succession

$$E_0t^2, E_0t^2r^2, E_0t^2r^4, \cdots, E_0t^2r^{2n}$$

Each of the transmission will have an added phase change to the wave depending on the added optical path of the wave. In our case it comes from the $2n$-th reflection ($n \in \mathbb{N},\ n > 1$), since we want to see the actual phase displacement of the outgoing final beam, which we will then measure.
Hence, supposing that between the two plates $n = n_2/n_1 = 1$ we have

$$\delta = 2kd\cos\theta = \frac{4\pi}{\lambda}d\cos\theta \tag{15.36}$$

Or, in general, if $n \neq 1$

$$\delta = 2k_0dn\cos\theta = \frac{4\pi n}{\lambda_0}d\cos\theta \tag{15.37}$$

Then, accounting the phase shift, the final total amplitude of the beam will be the superposition (i.e. the sum) of the amplitudes of every beam that gets transmitted post multiple reflection, with a multiplied phase shift, i.e.

$$E_T = E_0t^2 + E_0t^2r^2e^{i\delta} + E_0t^2r^4e^{2i\delta} + \cdots$$

At the limit of infinite reflection, or, in practical terms approximating for many reflections, we have a geometric series for the final result of the amplitude, which is easily calculable with the usual mathematical result

$$E_T = E_0t^2\sum_{j=0}^{\infty} r^{2j}e^{ij\delta} = \frac{E_0t^2}{1 - r^2e^{i\delta}} \tag{15.38}$$

In terms of intensity $I_T = |E_T|^2$, and remembering that in general $r, t \in \mathbb{C}$, we have

$$I_T = |E_T|^2 = \frac{I_0|t|^4}{|1 - r^2e^{i\delta}|^2} \tag{15.39}$$

Called $\delta_r/2$ the phase shift for a single reflection (the half unity factor is just for ease of calculation later), we have that

$$r = |r|e^{i\frac{\delta_r}{2}}$$

And, writing $\Delta = \delta + \delta_r$ (the half factor here gets elided since we're considering only pairs of reflections), we have in terms of reflectance R and transmittance T

$$I_T = I_0\frac{T^2}{|1 - Re^{i\Delta}|^2}$$

Calculating the absolute value in the denominator as $z\overline{z} = |z|^2$ we have, after some simple algebra

$$\left|1 - Re^{i\Delta}\right|^2 = 1 - R\left(e^{i\Delta} + e^{-i\Delta}\right) + R^2 = 1 - 2R\cos(\Delta) + R^2$$

Rewriting the cosine as one minus sine squared of half the argument, we have

$$\left|1 - Re^{i\Delta}\right| = 1 - 2R + 4R\sin^2\left(\frac{\Delta}{2}\right) + R^2 = (1-R)^2\left(1 + \frac{4R}{(1-R)^2}\sin^2\left(\frac{\Delta}{2}\right)\right)$$

Defined the *coefficient of Finesse* F as

$$F = \frac{4R}{(1-R)^2} \tag{15.40}$$

We have that finally the total intensity of the different beams is

$$I_T = \frac{I_0 T^2}{(1-R)^2}\frac{1}{1 + F\sin^2\left(\frac{\Delta}{2}\right)} \tag{15.41}$$

The second factor on the product is known as Airy's function, which will define (up to an intensity coefficient) the interference lines observed from the convergence of all these beams. The physical meaning of the finesse coefficient then can be seen as a value defining the sharpness of the fringes. Note that, also, if $\Delta/2$ is a whole multiple of π, Airy's function will always be equal to one for all values of F. Keep this in mind since it will be an useful concept for when we'll treat the Fabry-Pérot interferometer and the concept of *free spectral range* of a Fabry-Pérot interferometer (or etalon, depending on the experimental configuration).
As it's clear from the shape of the function and the previous observation, we have that the condition for having the maximum of Airy's function is then

$$\frac{\Delta_{max}}{2} = N\pi$$

Where $N \in \mathbb{N}$ is the *order of interference*, i.e. it's an indicator of the path difference between two successive beams. In fact, from the definition of Δ we have

$$\Delta_{max} = 2N\pi = \frac{4\pi}{\lambda_0} nd\cos\theta + \delta_r \tag{15.42}$$

In the most general case, for which the reflecting properties are not equal, we they'll contribute with two different reflection coefficients $r_1, r_2 \in \mathbb{C}$, which will contribute each with a phase shift δ_1, δ_2, i.e.

$$\begin{aligned} r_1 &= |r_1|e^{i\delta_1} \\ r_2 &= |r_2|e^{i\delta_2} \end{aligned} \tag{15.43}$$

All the previous derivations hold if and only if we define the transmittance and reflectance as follows

$$\begin{aligned} T &= |t_1||t_2| = \sqrt{T_1 T_2} \\ R &= |r_1||r_2| = \sqrt{R_1 R_2} \end{aligned} \tag{15.44}$$

Going back to our study of the intensity function for multi-beam interference, defined $\mathcal{I} = I_T/I_0$ then we have that the maximum and minimum values for intensity will be the following

$$\begin{aligned} \mathcal{I}_{max} &= \frac{T^2}{(1-R)^2} \\ \mathcal{I}_{min} &= \frac{T^2}{(1+R)^2} \end{aligned} \tag{15.45}$$

Which, in the realistic case of absorption of energy, defined an absorption function A for which

$$A + T + R = 1$$

We can rewrite the maximum of $\mathcal{I}$ in terms of reflectance and absorption as

$$\mathcal{I}_{max} = \frac{(1 - A - R)^2}{(1 - R)^2}$$

Where we used the simple substitution $T = 1 - A - R$

§ 15.5 Fabry-Pérot Instruments

An instruments which uses the exact concept that we defined in the previous section is the *Fabry-Pérot Interferometer* and the *Fabry-Pérot Etalon*, invented by C.Fabry and A.Pérot in 1889.
The instrument is composed, excluding the source, by one collimating lens which redirects the electromagnetic waves towards two semi-reflecting plates, and a focusing lens which focuses all the transmitted waves towards a single point, in which we would usually put a photomultiplier, an amplifier and then a detector or recorder.
The main difference between the Etalon configuration and the Interferometer configuration of a Fabry-Pérot instrument is whether the semi reflecting plates are movable or fixed. The Etalon configuration has fixed plates, while an interferometer or *scanning Fabry-Pérot* has moving plates.
These plates are usually made of glass or quartz, and the reflecting surfaces are parallel and *as smooth as possible*. The flatness required for having a working Fabry-Pérot is from at least $\lambda/20$, up to $\lambda/100$, with λ being the wavelength of the studied beam.
Commonly, scanning Fabry-Pérot are used with point sources, and the transmitted electromagnetic waves are focused to a pinhole, while the etalon is used with broad sources, the final beams here get focused to a single point on the focal plane.
The result of the interference created in a Fabry-Pérot instrument are circular interference fringes. Each ring corresponds to values of constant θ (remember that $\delta = \delta(\theta)$), and they're also known as *fringes of equal inclination*.
We continue by defining a fundamental concept in Fabry-Pérot instruments, the *free spectral range*.

Definition 15.5.1 (Free Spectral Range)**.** The *free spectral range* of a Fabry-Pérot is defined as the phase separation between two adjacent orders of interference N, as for $N, N+1$. I.e., the phase values Δ inside this range satisfy

$$\Delta_{N+1} - \Delta_N = 2\pi \tag{15.46}$$

Substituting $\Delta_N = 2N\pi$ and $\Delta_{N+1} = 2(N+1)\pi = \Delta_N + 2\pi$, and their definitions in terms of $\delta(\theta)$, we get

$$\frac{4\pi}{\lambda_{N+1}} nd\cos\theta - \frac{4\pi}{\lambda_N} nd\cos\theta = 2\pi$$

Solving in terms of $\lambda_{N+1}^{-1} - \lambda_N^{-1}$ we have

$$\frac{1}{\lambda_{N+1}} - \frac{1}{\lambda_N} = \frac{1}{2nd\cos\theta} \tag{15.47}$$

Or, substituting $\omega = 2\pi c/\lambda$, in terms of angular frequency

$$\omega_{N+1} - \omega_N = \frac{\pi c}{nd\cos\theta} \tag{15.48}$$

Or, also using $\omega = 2\pi\nu$, in terms of frequency

$$\nu_{N+1} - \nu_N = \frac{c}{2nd\cos\theta} \tag{15.49}$$

Or, again, using $k_0 = \omega/c = 2\pi/\lambda = 2\pi\nu/c$, in terms of wavenumber

$$k_{N+1} - k_N = \frac{\pi}{nd\cos\theta} \tag{15.50}$$

§§ 15.5.1 Resolution of Fabry-Pérot Instruments

We will now treat the resolution of interference fringes with Fabry-Pérot instruments.
For simplicity in calculation we will consider a non-monochromatic wave for which the spectrum is composed by two exact frequencies ω, ω_1, and this spectrum will be studied with a Fabry-Pérot.
Each frequency will contribute to the total intensity, therefore the fringe intensity pattern will be given by the sum of the two. Assuming that the initial intensity I_0 is the same for both, we get the total intensity as the sum of the two Airy functions

$$I_T = \frac{I_0}{1 + F\sin^2\left(\frac{\Delta}{2}\right)} + \frac{I_0}{1 + F\sin^2\left(\frac{\Delta_1}{2}\right)} \tag{15.51}$$

As usual, the phase Δ is

$$\Delta = \delta(\theta) + \delta_r = \frac{4\pi n}{\lambda_0} d\cos\theta + \delta_r$$

Which, if we assume an almost-normal incidence of the electromagnetic wave to the plates ($\theta << 1$), we have at first order and in terms of frequencies

$$\Delta = \frac{2\omega d}{c} + \delta_r + \mathcal{O}\left(\theta^3\right)$$
$$\Delta_1 = \frac{2\omega_1 d}{c} + \delta_r + \mathcal{O}\left(\theta^3\right)$$

And here comes in our help what's known as Taylor's criterion, which states that two *equal* lines are resolved if and only if the individual curves cross at the point of half intensity of each, for which then the intensity at the saddle is equal to exactly twice the initial intensity.
Since we're exactly smack in the middle between the two single peaks, considered the range between the two $\Delta - \Delta_1$, the total intensity at this saddle will be the usual Airy function times $2I_0$ this time, considered at half range, i.e. at $(\Delta - \Delta_1)/2$, i.e.

$$I_T|_{\frac{\Delta-\Delta_1}{2}} = \frac{2I_0}{1 + F\sin^2\left(\frac{\Delta-\Delta_1}{4}\right)} = I_0 \tag{15.52}$$

Solving for $\Delta - \Delta_1$, we're left with

$$F \sin^2\left(\frac{\Delta - \Delta_1}{4}\right) = 1$$

Assumed (we hope so for a good resolution!) $\Delta - \Delta_1 << 1$, we can approximate to the first order the sine function, getting

$$F\left(\left(\frac{\Delta - \Delta_1}{4}\right)^2 + \mathcal{O}\big((\Delta - \Delta_1)^6\big)\right) = 1 \implies |\Delta - \Delta_1| \approx \frac{4}{\sqrt{F}} = 2\frac{1-R}{\sqrt{R}} \tag{15.53}$$

I.e., the minimum resolvable range depends directly to the inverse of the square root of the finesse coefficient, which is only determined by the physical characteristics of the instrument!
In terms of frequencies, remembering that $\Delta \approx \frac{2d}{c}\omega$, we have that the smallest resolvable interval of frequencies is then

$$|\omega - \omega_1| \approx \frac{2c}{d\sqrt{F}} = \frac{c}{d}\frac{1-R}{\sqrt{R}} \tag{15.54}$$

Therefore, the smallest resolvable interval of frequencies for a Fabry-Pérot instrument depends only on the reflectance of the plates R and their distance d.
We also give a new definition for two elements often used when treating Fabry-Pérot instruments

Definition 15.5.2 (Reflecting Finesse)**.** Given a Fabry-Pérot instrument, the *reflecting finesse* $\mathcal{F}$, is defined as the free spectral range divided by the smallest resolvable interval, i.e.

$$\mathcal{F} = \frac{\Delta_{N+1} - \Delta_N}{|\Delta - \Delta_1|} = \frac{\pi}{2}\sqrt{F} = \frac{\pi}{2}\frac{\sqrt{R}}{1-R} \tag{15.55}$$

Definition 15.5.3 (Resolving Power)**.** The *resolving power* RP of a Fabry-Pérot instrument is defined as the inverse of the minimum resolvable interval times the scanned frequency (or angular frequency, or also wavelength), i.e.

$$RP = \frac{\omega}{|\omega - \omega_1|} = \frac{\nu}{|\nu - \nu_1|} = \frac{\lambda}{|\lambda - \lambda_1|} \tag{15.56}$$

It's clear, by direct substitution and some algebra, that the resolving power is directly tied to the reflecting finesse as

$$RP = N\mathcal{F} = N\pi\frac{\sqrt{R}}{1-R} \tag{15.57}$$

I.e., it's directly proportional to the reflecting finesse and the order of interference. Clearly, the resolving power can be made arbitrarily big simply by choosing higher interference orders. Note that by definition of N, this can be accomplished by increasing the mirror separation, which unfortunately reduces the free spectral range of the instrument.
Another idea for increasing the RP is by increasing $\mathcal{F}$, bringing the reflectance closer and closer to 1. There's a clear limit to this, in fact it's physically limited by the absorption of the used material, which reduces the intensity of the transmitted fringes.

16 Diffraction

§ 16.1 Helmholtz Equation

The basis for understanding diffraction, comes obviously from Maxwell's equations.
Suppose that we're in a medium for which there are no *free* charges and no currents whatsoever ($M^i = 0$). Since in the most general case we treat non-monochromatic waves, we propose an Ansatz. Given any field (E,B,H.D), it can be decomposed in *phasors* (complex exponentials), times an amplitude part that in general depends from frequency ω.
Therefore, for any given ω we have

$$E^i(x^i, t) = E_\omega(x^i)e^{i\omega t} \tag{16.1}$$

Where E^i_ω are the so-called monochromatic components of the fields, and $\epsilon_\omega(x^i) = \epsilon_0\,(1 + \chi_\omega(r))$ is the dielectric permittivity.
Then, Maxwell's equation become, remembering that $D = \epsilon(x,t)E$ and $\mu_0 H = B$ ($D_\omega = \epsilon_\omega(x^i)E_\omega$ and $\mu_0 H_\omega = B_\omega$ with our Ansatz) in this case,

$$\begin{cases} \partial_i D^i = 0 \\ \partial_i H^i = 0 \\ \epsilon^i{}_{jk}\partial^j E^k = -\mu_0 \partial_t H^i \\ \epsilon^i{}_{jk}\partial^j H^k = \partial_t D^i \end{cases}$$

The "time-independent" monochromatic version then becomes

$$\begin{cases} \partial_i E^i_\omega = -\dfrac{E^i_\omega \partial_i \epsilon_\omega}{\epsilon_\omega} = -2E^i_\omega \partial_i \log(n_\omega) \\ \partial_i H^i_\omega = 0 \\ \epsilon^i{}_{jk}\partial^j E^k_\omega = -i\omega\mu_0 H^i_\omega \\ \epsilon^i{}_{jk}\partial^j H^k_\omega = i\omega\epsilon_\omega E^i_\omega \end{cases} \tag{16.2}$$

Where we defined the usual refraction index as follows

$$n_\omega(x^i) = \sqrt{\frac{\epsilon_\omega(x^i)}{\epsilon_0}}$$

Now, solving only for E^i_ω, using the properties of the $\epsilon^i{}_{jk}$ symbol, we get the following partial differential equation

$$\partial_j \partial^j E^i_\omega + 2\partial_j \left(E^i_\omega \partial_i \log(n_\omega) \right) + \omega^2 \mu_0 \epsilon_\omega E^i_\omega = 0 \tag{16.3}$$

We can rewrite $\epsilon_\omega \mu_0 \omega^2 = \omega^2/c^2 n^2_\omega = k_0^2 n^2_\omega = k^2$, and imposing that the refraction index varies slowly ($\Delta n_\omega / n_\omega << 1$), we get *Helmholtz's equation*, a time-independent counterpart to the usual wave equation, which can also be seen as an eigenvalue problem of the Laplace operator

$$\partial^j \partial_j E^i_\omega + k_0^2 n_\omega E^i_\omega = 0 \tag{16.4}$$

§§ 16.1.1 Normal Modes of the Electric Field

The solution of this differential equation isn't straightforward in general sets, so we begin to impose another Ansatz. Given the monochromatic field E_ω, we suppose that the solutions will be exponential solutions depending on the wavevector k^i, as follows

$$E^i_\omega(x^i) = E^i_{\mathbf{k}} e^{-ik^i x_i}$$

It's clear that in order to satisfy Gauss' equation we must have

$$\partial_i E^i_\omega = -ik_i E^i_{\mathbf{k}} e^{-ik^i x_i} = 0 \implies k_i E^i_\omega = 0$$

This clearly means that the direction of propagation of the waves is parallel to the wavevector, i.e. we have transverse propagation. Supposing that this direction coincides with the direction of the z axis, we can propose a new notation.

$$x^i = (r^\alpha_\perp, z)$$
$$k^i = (k^\alpha_\perp, k_z)$$

Where $\alpha = 1, 2$. Rewriting k_z in terms of $k_\perp$ we have

$$k^2 = k^2_\perp + k^2_z \implies \quad k_z = \sqrt{k^2 - k^2_\perp}$$

Which gives us a constraint on k, as

$$k \geq k_\perp$$

Note that there exist values for which $k_z \in \mathbb{C}$, they'll be treated later. Therefore, the general solution is

$$E^i_\omega(x^i) = E^i_{\mathbf{k}_\perp} e^{-ik^i_\perp r_{i,\perp} - iz\sqrt{k^2 - k^2_\perp}} \tag{16.5}$$

This is our well known plane wave propagation formula. A similar formula for H^i_ω can be retrieved by applying the same manipulations to Maxwell's equations, and the resulting Helmholtz equation will describe a field which vibrates in a direction orthogonal to E. The complete monochromatic solution to Helmholtz's equation will then be a superposition of every solution with $k_\perp \leq k$

$$E^i(x^i, t) = \iint_{k_\perp \leq k} E^i_{\mathbf{k}_\perp}(\omega) e^{i\omega t - ik^i_\perp r_{i,\perp} - iz\sqrt{k^2 - k^2_\perp}} \, \mathrm{d}^2 k_\perp \tag{16.6}$$

As we said, in general $k \in \mathbb{C}$ and the solutions found are said to be *evanescent modes*. All other solutions where $k \in \mathbb{R}$ are said to be *radiation modes*. Therefore, the integral can be extended without problems to all the possible values of $k_\perp$, i.e. $\mathbb{R}^2$

§ 16.2 Scalar Diffraction Theory

§§ 16.2.1 Green's Functions and Kirchhoff Diffraction

The previous section deals with Helmholtz's equation and its general solution. Usually tho, in real world application the using that solution is counterproductive, as it's not always readily integrable.
Looking again at Helmholtz's equation, it's clear that it's a set of three *uncoupled* partial differential equations. Said $u(r)$ as the generic component of the field, the equation is

$$\partial^i \partial_i u(r) + k^2 u(r) = 0 \tag{16.7}$$

Now, let's say we want to solve it in some particular (compact) set V, which has a random shape.
Due to the improbable boundary conditions we could choose in order to fit our problem inside the chosen set, we use *Green's functions*, for which, a solution $u(r)$ can then be expressed as follows, for some differential equation $\hat{L}h = f$

$$h(x) = \int G(x - x_0) f(x_0)\, \mathrm{d}x_0$$

Here G is Green's function for the differential operator $\hat{L}$, i.e. a function such that $\hat{L}G = \delta_{x_0}$, where $\delta_{x_0} = \delta(x - x_0)$ is Dirac's delta distribution.
Going back to our partial differential equation, if the source point of the electromagnetic waves is $r_0 \in V \subset \mathbb{R}^3$, the field $u(r)$ diverges at r_0, so it's a good idea to remove an infintesimal ball of radius ϵ from the set, i.e. we define a new set

$$\tilde{V} = V \setminus B_\epsilon(r_0)$$

This differential equation doesn't properly follow what we wrote before about Green's functions, since here Green's function would just be a solution of the problem, since

$$\hat{L}G = \left(\partial^i \partial_i + k^2\right) G(r) = 0$$

Therefore, Green's function is not unique. We will choose the simplest solution, i.e. a spherical wave emitting from r_0

$$G(r - r_0) = \frac{e^{ik(r - r_0)}}{|r - r_0|} \tag{16.8}$$

We can now use Green's theorem for finding a really important relation

$$\iiint_{\tilde{V}} \left(u(r)\partial^i \partial_i G - G(r - r_0)\partial^i \partial_i u\right) \mathrm{d}^3 x = -\iiint_{\tilde{V}} \left(uk^2 G - uk^2 G\right) \mathrm{d}^3 x = 0 \tag{16.9}$$

Where we applied Helmholtz's equation on the integral.
But also, for Green's theorem, said $\partial\tilde{V} = \partial V \cup \partial B_\epsilon(r_0)$, then

$$\oiint_{\partial\tilde{V}} (u(r)\partial_i G - G(r - r_0)\partial_i u)\, \hat{n}^i\, \mathrm{d}^2 s = \oiint_{\partial V \cup \partial B_\epsilon(r_0)} (u(r)\partial_i G - G(r - r_0)\partial_i u)\, \hat{n}^i\, \mathrm{d}^2 s = 0 \tag{16.10}$$

Or, using the properties of integrals and measures

$$\oiint_{\partial B_\epsilon(r_0)} (u(r)\partial_i G - G(r - r_0)\partial_i u)\, \hat{n}^i\, \mathrm{d}^2 s = -\oiint_{\partial V} (u(r)\partial_i G - G(r - r_0)\partial_i u)\, \hat{n}^i\, \mathrm{d}^2 s = 0 \tag{16.11}$$

The propagation problem now just reduced to calculating the integral on the left, which is analytically solvable. In $\partial B_\epsilon(r_0)$ it's clear that for our formulation of the problem $\hat{n}^i = -\hat{r}^i$, and therefore,

$$\frac{\partial G}{\partial n} = \partial_r G = -\left(ik - \frac{1}{|r-r_0|}\right)\frac{e^{ik(r-r_0)}}{|r-r_0|}$$

Therefore, Substituting into both integrals this result, and integrating on the left and sending $\epsilon \to 0$ (basically, going to r_0), we have

$$\lim_{\epsilon\to 0}\left[4\pi\epsilon^2\left(u(r_0)\left(\frac{1}{\epsilon} - ik\right)\frac{e^{ik\epsilon}}{\epsilon} - \hat{r}^i\partial_i u(r_0)\frac{e^{ik\epsilon}}{\epsilon}\right)\right] = 4\pi u(r_0)$$

I.e., the integral on the ball collapses to a constant times the field evaluated at the source. Fixing everything together we get *Kirchhoff-Helmholtz's integral theorem*

$$u(r_0) = \frac{1}{4\pi}\oiint_{\partial V}\left(\frac{\partial u}{\partial n}G(r-r_0) - \frac{\partial G}{\partial n}u(r)\right)\mathrm{d}^2 s \tag{16.12}$$

This theorem, clearly states that, at any point $r_0 \in V$, $u(r)$ can be determined only in terms of its boundary values. Writing explicitly Green's function, we have

$$u(r_0) = \frac{1}{4\pi}\oiint_{\partial V}\left(\frac{\partial u}{\partial n}\frac{e^{ik(r-r_0)}}{|r-r_0|} - u(r)\left(ik - \frac{1}{|r-r_0|}\right)\frac{e^{ik(r-r_0)}}{|r-r_0|}\right)\mathrm{d}^2 s \tag{16.13}$$

§§ 16.2.2 Fresnel-Kirchhoff Diffraction Formula

Since, as we calculated before, the wave at some point is defined from the values of the same at the boundary, we need to specify those.
Suppose that there's an opaque infinite wall with an aperture Σ. In order to evaluate the diffracted field we need to choose a useful surface ∂V over which we will perform the integration.
Chosen a point r_0 at which we will evaluate the field, we take $\partial V = S_1 + S_R$, where S_1 is a plane just after the aperture, and S_R is a spherical cap of radius R centered in r_0.
On S_R, we have that, if $R \to \infty$

$$\left.\frac{\partial G}{\partial n}\right|_{S_R} = \left(ik - \frac{1}{R}\right)\frac{e^{ikR}}{R} \approx ikG(r-r_0) \tag{16.14}$$

Then

$$\iint_{S_R}\left(G(r-r_0)\frac{\partial u}{\partial n} - u(r)\frac{\partial G}{\partial n}\right)\mathrm{d}^2 s \simeq \iint_{S_R} G(r-r_0)\left(\frac{\partial u}{\partial n} - iku(r)\right)\mathrm{d}^2 s$$

Written $\mathrm{d}^2 s = R^2 \mathrm{d}\Omega$, with Ω being the solid angle, we have that due to the functional shape of G, the function $|RG(r-r_0)|$ is uniformly bounded, which implies that the integral vanishes only if

$$\lim_{R\to\infty}\left[R\left(\frac{\partial u}{\partial n} - iku(r)\right)\right] = 0 \tag{16.15}$$

I.e., if the field $u(r)$ vanishes as dast as a diverging wave. This condition is known as «Sommerfield radiation condition», and *guarantees* that we're dealing only with outgoing waves from Σ and not

incoming. Note that this is clearly respected for spherical scalar waves.
We're left with only one integral now, which is the following

$$u(r_0) = \frac{1}{4\pi} \iint_{S_1} \left(G(r - r_0) \frac{\partial u}{\partial n} - u(r) \frac{\partial G}{\partial n} \right) \mathrm{d}^2 s \tag{16.16}$$

We now proceed to have two assumptions on the boundary behavior of u.

1. The values at the aperture of u are the same *with or without* the screen.
2. The values of u at the opaque wall is identically 0.

Then, the integral can be approximated to the following

$$u(r_0) = \frac{1}{4\pi} \iint_{\Sigma} \left(G(r - r_0) \frac{\partial u}{\partial n} - u(r) \frac{\partial G}{\partial n} \right) \mathrm{d}^2 s \tag{16.17}$$

This approximation holds *only* if $\lambda >> d$, with d being the distance from the aperture. If this is true, we can again use the approximation (16.14), and obtain the following simplified integral

$$u(r_0) = \frac{1}{4\pi} \iint_{\Sigma} \frac{e^{ik(r-r_0)}}{|r - r_0|} \left(\frac{\partial u}{\partial n} - u(r) \frac{\hat{n}^i r_i}{|r - r_0|} \right) \mathrm{d}^2 s$$

We suppose that $u(r)$ is a spherical wave coming from a secondary source r_1 with amplitude A, for which holds $\lambda >> d_1$, with d_1 being the distance from the source to the aperture, then, if we write r' as the distance between the source and the screen and r the distance between the source and the observation point, we have firstly that

$$\begin{cases} u(r) = \dfrac{A}{r'} e^{ikr' + i\omega t} \\ \dfrac{\partial u}{\partial n} \simeq \dfrac{ikA}{r'} e^{ikr' + i\omega t} \end{cases}$$

Then we have that

$$u(r_0) = \frac{ikAe^{i\omega t}}{4\pi} \iint_{\Sigma} \frac{e^{ik(r+r')}}{rr'} \left(\frac{r'^i \hat{n}_i}{r'} - \frac{r^i \hat{n}_i}{r} \right) \mathrm{d}^2 s$$

Rewriting the scalar products on the right as $\cos(\hat{n}, r)$, i.e. the cosines angles between the two vectors in question, we get, remembering that $k = 2\pi/\lambda$

$$u(r_0) = \frac{Ae^{i\omega t}}{2i\lambda} \iint_{\Sigma} \frac{e^{ik(r+r')}}{rr'} \left(\cos(\hat{n}, r) - \cos(\hat{n}, r') \right) \mathrm{d}^2 s \tag{16.18}$$

This integral is known as the «Fresnel-Kirchhoff diffraction formula». Note how this is a mathematical version of Huygens' principle, which states that a wavefront is generated at each point of it by spherical waves.
What we see in this integral is exactly that, the measured field at r_0 can be seen simply as a superposition of spherical waves times a correction factor, called «obliquity», which corrects the calculation in case of waves that do not arrive frontally to the aperture.
It's also important to note that if we switch the observer with the source and vice-versa, the result doesn't change. This is well known as «Helmholtz's reciprocity theorem».

§§ 16.2.3 Rayleigh-Sommerfeld Diffraction

Another formulation of diffraction comes from Rayleigh and Sommerfeld, which idea comes from the fact that if we chose a wave that, on Σ has $\partial_n u = u = 0$, then, being u harmonic, it must be zero everywhere. Therefore it's clear that Kirchhoff's formulation of diffraction, even though it's precise in predicting diffraction patterns, it still implies in its boundary condition that there is no field before the aperture.
Sommerfeld then, in order to solve this problem imposing a different choice of Green's function. Supposed that

1. Scalar theory holds, i.e. $\Delta n_\omega / n_\omega << 1$ (dispersion is negligible)
2. $u(r), G(r)$ satisfy Helmholtz's equation
3. Sommerfeld's radiation condition holds, i.e. $\lim_{R\to\infty} R\left(\frac{\partial u}{\partial n} - iku(r)\right) = 0$

We can choose a Green's function such that either $G(r)$ or its normal derivative are equal to zero on S_1, so that we don't have to impose conditions on both u and its normal derivative.
Suppose then that $G(r)$ is generated at the observation point, and there is a secondary source mirrored at the other side of the screen. Assuming that the wavelength of the waves emitted by these two sources are equal, we can say for sure that their phases are shifted by π or they're in phase. Thus, chosen the Green's function with a phase difference of π, we have

$$G_-(r) = \frac{1}{|r - r_0|} e^{ik(r-r_0)} - \frac{1}{|r - \tilde{r}_0|} e^{ik(r-\tilde{r})} \tag{16.19}$$

It's clear that $G(\Sigma) = 0$, hence the Kirchhoff integral reduces to the first Rayleigh-Sommerfeld solution

$$u_I(r_0) = -\frac{1}{4\pi} \iint_\Sigma u(r) \frac{\partial G_-}{\partial n} \, \mathrm{d}^2 s \tag{16.20}$$

With our definition, if again $d >> \lambda$ we have that $G_-(r) \approx 0$ everywhere, and

$$\frac{\partial G_-}{\partial n} \approx 2ik \cos(\hat{n}, r) \frac{e^{ik(r-r_0)}}{|r - r_0|} \tag{16.21}$$

If we chose the Green's function with no phase difference between the imaginary source and the observation point, we'd have

$$G_+(r) = \frac{1}{|r - r_0|} e^{ik(r-r_0)} + \frac{1}{|r - \tilde{r}_0|} e^{ik(r-\tilde{r})} \tag{16.22}$$

Which implies that now the normal derivative vanishes on Σ, and the second Rayleigh-Sommerfeld solution is

$$u_{II}(r) = \frac{1}{4\pi} \iint_\Sigma G_+(r) \frac{\partial u}{\partial n} \, \mathrm{d}^2 s \tag{16.23}$$

Taken the Fresnel-Kirchhoff diffraction formula (16.18), it's pretty easy to get back the two Rayleigh-Sommerfeld solutions by simple substitution.

Given that $d >> \lambda$ and $\tilde{d} >> \lambda$, with $\tilde{d}$ being the distance of the imaginary source from the aperture, we have that in this approximation, called $G_K(r)$ the Kirchhoff-Fresnel Green's function

$$\begin{cases} \dfrac{\partial G_-}{\partial n} = 2ik\cos(\hat{n}, r)\dfrac{e^{ik(r-r_0)}}{|r-r_0|} = 2\dfrac{\partial G_K}{\partial n} \\ G_+(r) = 2\dfrac{e^{ik(r-\tilde{r}_0}}{|r-\tilde{r_0}|} = 2G_K(r) \end{cases} \tag{16.24}$$

Which implies that

$$\begin{cases} u_I(r_0) = \dfrac{1}{i\lambda}\displaystyle\iint_\Sigma \frac{u(r)}{|r-r_0|}e^{ik(r-r_0)}\cos(\hat{n}, r)\,\mathrm{d}^2 s \\ u_{II}(r_0) = \dfrac{1}{2\pi}\displaystyle\iint_\Sigma \frac{1}{|r-\tilde{r}_0|}\frac{\partial u}{\partial n}e^{ik(r-\tilde{r}_0)}\cos(\hat{n}, \tilde{r})\,\mathrm{d}^2 s \end{cases} \tag{16.25}$$

Which, for a diverging spherical wave become

$$\begin{cases} u_I(r_0) = \dfrac{Ae^{i\omega t}}{i\lambda}\displaystyle\iint_\Sigma \frac{e^{ik(r+r')}}{rr'}\cos(\hat{n}, r)\,\mathrm{d}^2 s \\ u_{II}(r_0) = \dfrac{iAe^{i\omega t}}{\lambda}\displaystyle\iint_\Sigma \frac{e^{ik(r+r')}}{rr'}\cos\hat{n}, r\,\mathrm{d}^2 s \end{cases} \tag{16.26}$$

Note that if we compare the two theories, and define an *obliquity* factor $\psi(\theta, \theta')$ as

$$\psi(\theta, \theta') = \begin{cases} \psi_K = \dfrac{1}{2}\left(\cos(\theta) - \cos(\theta')\right) \\ \psi_I = \cos\theta \\ \psi_{II} = -\cos\theta' \end{cases} \tag{16.27}$$

The two theories converge, where Kirchhoff theory is the mathematical average of the two. In this situation the general diffraction integral is written as

$$u(r_0) = \frac{Ae^{i\omega t}}{i\lambda}\iint_\Sigma \frac{e^{ik(r+r')}}{rr'}\psi(\theta, \theta')\,\mathrm{d}^2 s \tag{16.28}$$

It's also important to note that, in order to work properly, both Rayleigh-Sommerfeld solutions need that the source and/or the measuring point are far away from the aperture (d»λ), also known as *far field approximation*, for which we can consider that the waves are *plane* and not spherical, thing that we cannot avoid with Kirchhoff's formulation.
Also, Rayleigh-Sommerfeld theory needs planar diffraction screens, which is not always the case.

§§ 16.2.4 Non-monochromatic Case

Considered Rayleigh-Sommerfeld theory in the case of non-monochromatic waves $u(r, t)$, the first thing that comes to mind is firstly decomposing the wave into monochromatic components and then superimpose all the components into an integral. It's clear that this integral is actually a Fourier transform of the field.

$$u(r, t) = \int_{\mathbb{R}} \hat{u}(r, \nu)e^{2\pi i\nu t}\,\mathrm{d}t$$

Where $\hat{u}$ is u's transform.
Introducing a parity transformation in frequency space $\nu \to -\nu$, and said r_0 and r_1 respectively the distances from the measuring point and from the source to the screen, we get from (16.18) that, using $\lambda\nu = c/n = u$, $k = 2\pi/\lambda = 2\pi\nu/u$

$$\hat{u}(r_0, -\nu) = -\frac{i\nu}{u}\int_{\mathbb{R}} \hat{u}(r_1, -\nu)\frac{e^{-2\pi i\nu\frac{r-r_0}{u}}}{|r-r_0|}\cos(\hat{n}, r)\,\mathrm{d}^2 s \tag{16.29}$$

Therefore, integrating with respect to time

$$u(r_0, t) = \iint_\Sigma \frac{\cos(\hat{n}, r)}{2\pi u|r-r_0|}\int_{\mathbb{R}} -2\pi i\nu\hat{u}(r_1, -\nu)e^{-2\pi i\nu\left(t-\frac{r-r_0}{u}\right)}\,\mathrm{d}t\,\mathrm{d}^2 s \tag{16.30}$$

Since the measuring point and the source aren't moving, using the properties of the Fourier transform, we can say that, the time integral is just the derivative of a transform, where, if we indicate the Fourier inverse operator as $\hat{\mathcal{F}}^{-1}$, gives

$$u(r_0, t) = \frac{1}{2\pi u}\iint_\Sigma \frac{\partial}{\partial t}\hat{\mathcal{F}}^{-1}\left[\hat{u}\right]\left(t - \frac{r-r_0}{u}\right)\frac{\cos(\hat{n}, r)}{|r-r_0|}\,\mathrm{d}^2 s \tag{16.31}$$

Writing explicitly the transform as it's defined, we then have

$$u(r_0, t) = \frac{1}{2\pi u}\iint_\Sigma \left.\frac{\partial u}{\partial t}\right|_{t=t_r}\frac{\cos(\hat{n}, r)}{|r-r_0|}\,\mathrm{d}^2 s \tag{16.32}$$

Where we evaluate the time derivative at the retarded time $t_r = t - r/u$, therefore also conserving causality.

§ 16.3 Fresnel Diffraction

Let's put ourselves again in the paraxial case, and assume that there are two planes, one containing the observed diffraction pattern, with coordinates (x, y), and another one containing the aperture, with coordinates (ξ, η). Using Rayleigh-Sommerfeld's first solution and imposing the paraxial approximation, for which, since we chose z as our direction of propagation, we could say, called θ our inclination angle, with respect to z

$$u(r_0) = \frac{z}{i\lambda}\iint_\Sigma u(r_{01})\frac{e^{ikr_{01}}}{r_{01}^2}\,\mathrm{d}\xi\,\mathrm{d}\eta$$

Where we have

$$r_{01} = \sqrt{z^2 + (x-\xi)^2 + (y-\eta)^2} \tag{16.33}$$

It's clear, that in order to have this approximation work we need two approximations

1. Scalar theory
2. $r_{01} >> \lambda$

The third approximation comes from the definition of r_{01} itself. Approximating the power series to the first order, in order to count for the slight curvature of the waves, we have

$$r_{01} \approx z\left(1+\frac{1}{2}\left[\left(\frac{x-\xi}{z}\right)^2+\left(\frac{y-\eta}{z}\right)^2\right]\right) = z+\frac{1}{2z}(x-\xi)^2+\frac{1}{2z}(y-\eta)^2 \tag{16.34}$$

The choice of keeping first order factors is based on the exponential factor itself. In fact, since $k \propto \lambda^{-1}$, for wavelengths of visible light, we have

$$kr_{01} \propto \frac{r_{01}}{\lambda} \propto 10^7$$

Therefore, rendering the approximation invalid. It should be noted tho that for a far field approximation we can say without problems that the term r_{01} at the denominator is approximately equal to z, while at the exponential we have to keep the first order. Then

$$\begin{cases} r_{01}^{-1} \approx z^{-1} \\ e^{ikr_{01}} \approx e^{ikz}e^{\frac{ik}{2z}\left[(x-\xi)^2+(y-\eta)^2\right]} \end{cases}$$

Fixing everything, and noting that since in this approximation the contributions come only from the aperture, we rewrite the Rayleigh-Sommerfeld integral as follows

$$u(x,y) = \frac{e^{ikz}}{i\lambda z}\iint_{\mathbb{R}^2} u(\xi,\eta)e^{\frac{ik}{2z}\left[(x-\xi)^2+(y-\eta)^2\right]}\,\mathrm{d}\xi\,\mathrm{d}\eta \tag{16.35}$$

Where $u(\xi,\eta)$ is the so called «aperture function».
Note that if we define the *kernel* $h(x,y)$ as follows

$$h(x,y) = \frac{e^{i(kz+\omega t)}}{i\lambda z}e^{\frac{ik}{2z}\left[(x-\xi)^2+(y-\eta)^2\right]} \tag{16.36}$$

Then, the field measured at the screen is simply the convolution between the aperture function and the kernel

$$u(x,y) = u(\xi,\eta) * h(x.y) \tag{16.37}$$

Working again on the expression inside the integral, it's possible to expand the squares and bring outside all common factors, yielding

$$\frac{ik}{2z}\left[(x-\xi)^2+(y-\eta)^2\right] = \frac{ik}{2z}(x^2+y^2)+\frac{ik}{2z}(\xi^2+\eta^2)-\frac{ik}{z}(x\xi+y\eta)$$

And the integral for a monochromatic source becomes

$$u_\omega(x.y) = \frac{e^{i(kz+\omega t)}e^{\frac{ik}{2z}(x^2+y^2)}}{i\lambda z}\iint_{\mathbb{R}^2}\left[u(\xi,\eta)e^{\frac{ik}{2z}(\xi^2+\eta^2)}\right]e^{-\frac{2\pi i}{\lambda z}(x\xi+y\eta)}\,\mathrm{d}\xi\,\mathrm{d}\eta = \frac{e^{i(kz+\omega t)}e^{\frac{ik}{2z}(x^2+y^2)}}{i\lambda z}\hat{\mathcal{F}}\left[u(\xi,\eta)e^{\frac{ik}{2z}(\xi^2+\eta^2)}\right](x,y) \tag{16.38}$$

Which then describes the observed pattern as a bidimensional Fourier transform of the aperture function times a quadratic phase factor.
This integral is known as the *Fresnel diffraction integral*. Remember that this integral is valid only and only if:

1. Scalar theory holds
2. The screen is at many wavelengths of distance from the aperture
3. The second order terms in the square root power expansion can be neglected

The third condition is satisfied, only if, expanding the root to the second order, as

$$ikr_{01} \approx ikz + \frac{ik}{2z}\left[(x-\xi)^2+(y-\eta)^2\right] + \frac{ik}{8z^3}\left[(x-\xi)^2+(y-\eta)^2\right]^2 \tag{16.39}$$

And, the last term is negligible, i.e., in terms of wavelength

$$\frac{\pi}{4\lambda z^3}\left[(x-\xi)^2+(y-\eta)^2\right]^2 << 1 \implies z^3 >> \frac{\pi}{4\lambda}\left[(x-\xi)^2+(y-\eta)^2\right]^2 \tag{16.40}$$

As an example, for a circular aperture of diameter 1 cm, an observation region of size 1 cm and electromagnetic waves with wavelenghts of 0.5 μm, the distance $z >> 25$cm for rendering this approximation valid. This can be reduced by imposing that the second order terms are adiabatic invariants for the integral. This approximation is also known as the *near field approximation* for this reason.

§§ 16.3.1 Fresnel Zones and Zone Plates

<++>

§ 16.4 Fraunhofer Diffraction

Fraunhofer diffraction imposes an even stricter approximation on the exponential, by ditching completely the first order terms, therefore, if

$$z >> \frac{k}{2}\left(\xi^2+\eta^2\right) \tag{16.41}$$

Then the integral gets even more simplified, and the resulting diffracted field becomes the exact Fourier transform of the aperture function $u(\xi,\eta)$, i.e., said $\nu_x = x/(\lambda z)$ and $\nu_y = y/(\lambda z)$, we have

$$u(x,y) = \frac{e^{i(kz-\omega t)}e^{\frac{ik}{2z}(x^2+y^2)}}{i\lambda z}\hat{\mathcal{F}}\left[u\right](\nu_x,\nu_y)$$

Or, explicitly

$$u(x,y) = \frac{e^{i(kz-\omega t)}e^{\frac{ik}{2z}(x^2+y^2)}}{i\lambda z}\iint_\Sigma u(\xi,\eta)e^{-\frac{2\pi i}{\lambda z}(x\xi+y\eta)}\,\mathrm{d}\xi\,\mathrm{d}\eta \tag{16.42}$$

This approximation, due to its zeroth order nature is known as the *far field approximation*, in fact, if we wanted to use this approximation with electromagnetic waves with wavelength of 0.6 μm (red light) and an aperture width of 2.5 cm the observation distance $z >> 1600$m. Another less stringent condition is the so called *antenna designer's formula*, for which, given an aperture of linear dimension D, the Fraunhofer approximation will be valid for

$$z > \frac{2D^2}{\lambda} \tag{16.43}$$

Still, also with this, in the previous case $z > 2000$m, but the requirement is less stringent.

§ 16.5 Examples of Fraunhofer Diffraction Integrals

17 Optics of Solids

§ 17.1 The General Wave Equation

<++>

§ 17.2 Dispersion and Absorption

§ 17.3 Propagation in Conducting Media

§ 17.4 Propagation in Crystals

§§ 17.4.1 Birefrengence

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

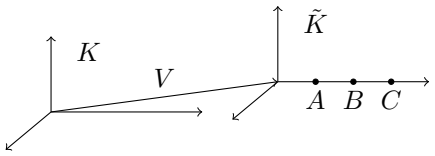

Figure A.1: The two frames K and $\tilde{K}$

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 $|v| < c$.

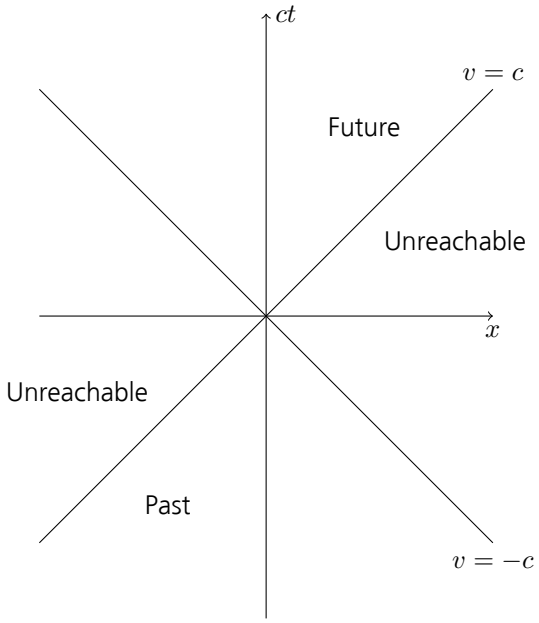

Figure A.2: Simple spacetime diagram. Note how all the events beyond the asymptote (or *horizon*) $v = \pm c$ are inaccessible from 0

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.
S O L U T I O N
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.