Introduction to near-field optics : angular spectrum, evanescent waves. 1

1.1 Introduction : From the wave equation to the Helmholtz equation

Transition to the monochromatic regime

We recall the expression of the wave equation in three dimensions :

$$\nabla^2 \mathbf{E} - \frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} = 0 \tag{1.1}$$

The wave equation is a second-order differential equation involving three spatial variables and one time variable. This coupling between spatial and temporal derivatives makes the equation difficult to solve. To simplify it, we can rewrite the field $\mathbf{E}(\mathbf{r},t)$ using its temporal Fourier transform — that is, express the total field as the sum of its spectral components (see Figure **??**) :

$$\mathbf{E}(\mathbf{r},t) = \int_{-\infty}^{+\infty} \tilde{\mathbf{E}}(\mathbf{r},\omega);\mathrm{e}^{-\mathrm{i}\omega t};\frac{\mathrm{d}\omega}{2\pi} \tag{1.2}$$

where $\tilde{\mathbf{E}}(\mathbf{r},\omega)$ is the Fourier component of the electric field at frequency $\frac{\omega}{2\pi}$. $\tilde{\mathbf{E}}(\mathbf{r},\omega)$ is no longer a function of time and represents the **spatial field distribution** associated with the spectral component at ω. Substituting this expression into the wave equation gives :

$$\nabla^2\left[\int_{-\infty}^{+\infty}\tilde{\mathbf{E}}(\mathbf{r},\omega),\mathrm{e}^{-\mathrm{i}\omega t},\frac{\mathrm{d}\omega}{2\pi}\right] - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\left[\int_{-\infty}^{+\infty}\tilde{\mathbf{E}}(\mathbf{r},\omega),\mathrm{e}^{-\mathrm{i}\omega t},\frac{\mathrm{d}\omega}{2\pi}\right] = 0$$

$$\int_{-\infty}^{+\infty}\nabla^2\left(\tilde{\mathbf{E}}(\mathbf{r},\omega),\mathrm{e}^{-\mathrm{i}\omega t}\right) - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\left(\tilde{\mathbf{E}}(\mathbf{r},\omega),\mathrm{e}^{-\mathrm{i}\omega t}\right)\frac{\mathrm{d}\omega}{2\pi} = 0$$

The time derivative $\frac{\partial^2}{\partial t^2}$ acts only on the complex exponential. The calculation thus becomes very simple :

$$\frac{\partial^2}{\partial t^2}\left(\tilde{\mathbf{E}}(\mathbf{r},\omega),\mathrm{e}^{-\mathrm{i}\omega t}\right) = -\omega^2\tilde{\mathbf{E}}(\mathbf{r},\omega),\mathrm{e}^{-\mathrm{i}\omega t} \tag{1.3}$$

Thanks to the Fourier transform, the second-order time derivative $\frac{\partial^2}{\partial t^2}$ is replaced by a simple multiplication by $-\omega^2$. We have thus partially linearized the wave equation.

The Helmholtz equation

Now that the time derivative has disappeared, we can rearrange the terms of the previous equations. We then obtain :

$$\int_{-\infty}^{+\infty}\left[\nabla^2\tilde{\mathbf{E}}(\mathbf{r},\omega)+\frac{\omega^2}{c^2}\tilde{\mathbf{E}}(\mathbf{r},\omega)\right]\mathrm{e}^{-\mathrm{i}\omega t},\frac{\mathrm{d}\omega}{2\pi}=0 \qquad (1.4)$$

Since the functions $t \longrightarrow \mathrm{e}^{-\mathrm{i}\omega t}$ ($\omega \in \mathbb{R}$) form a basis (with some care regarding the function space), if the integral over the entire spectrum is zero, then for each frequency ω, the term inside the integral must vanish. Hence :

Helmholtz equation in vacuum

$$\forall\omega, \qquad \nabla^2\tilde{\mathbf{E}}(\mathbf{r},\omega)+\frac{\omega^2}{c^2}\tilde{\mathbf{E}}(\mathbf{r},\omega)=0 \qquad (1.5)$$

This vector equation actually represents three scalar equations, each corresponding to one component of the electric field : $\nabla^2\tilde{E}_i(\mathbf{r},\omega)-\frac{\omega^2}{c^2}\tilde{E}_i(\mathbf{r},\omega)=0$, with $i=x,y,z$.

This equation is the **Helmholtz equation**. Since it is written for a fixed ω, it is also called the **propagation equation in the monochromatic regime**. By establishing it, we have shown that :

« $\mathbf{E}(\mathbf{r},t)$ is a solution of the wave equation »

is equivalent to saying that

« each of its spectral components $\tilde{\mathbf{E}}(\mathbf{r},\omega)$ ($\omega \in \mathbb{R}$) is a solution of the Helmholtz equation ».

In doing so, we have moved from a single equation that is difficult to solve because it involves partial derivatives with respect to both space and time (the wave equation), to an infinite set of equations, each of which is much simpler to solve. Indeed, the second-order time derivative has been linearized, replaced by a product.

Solving the Helmholtz equation in the 1D case, in vacuum

Let us look for the solution of the Helmholtz equation in 1D in the form of a wave propagating along z and polarized along x, that is, written as $\tilde{\mathbf{E}}(z,\omega)=\tilde{E}(z,\omega)\mathbf{u}_x$. Projecting the Helmholtz equation onto $\mathbf{u}_x$ gives :

$$\frac{\partial^2}{\partial z^2}\tilde{E}(z,\omega)+\frac{\omega^2}{c^2}\tilde{E}(z,\omega)=0 \qquad (1.6)$$

which is a second-order homogeneous differential equation*. Its general solution is of the form :

Solution of the 1D Helmholtz equation in vacuum

$$\tilde{\mathbf{E}}(z,\omega) = A_+ \mathrm{e}^{\mathrm{i}kz} \mathbf{u}x + A - \mathrm{e}^{-\mathrm{i}kz} \mathbf{u}_x \tag{1.7}$$

In the monochromatic case, this solution can be written as a function of space and time :

$$\mathbf{E}(z,t) = A_+ \mathrm{e}^{\mathrm{i}(kz-\omega t)} \mathbf{u}x + A - \mathrm{e}^{-\mathrm{i}(kz-\omega t)} \mathbf{u}x$$

where k is the positive quantity such that $k^2 = \frac{\omega^2}{c^2}$. The solution can thus be written as a **linear combination of two monochromatic plane waves**, one propagating along $+z$, and the other in the opposite direction $-z$, with $A+$ and A_- denoting the complex field amplitudes of each of these two plane waves.

The relationship between k and ω is called the **dispersion relation** :

Dispersion relation of plane waves in vacuum

$$k = \frac{\omega}{c} \tag{1.8}$$

k has the dimension of an inverse length and therefore designates, up to a factor of 2π, the spatial oscillation frequency of the plane wave (it is $k/2\pi$ that counts the number of oscillations per unit length : k appears as a spatial pulsation). We relate this frequency to the wavelength, which is the spatial period of oscillation.

Link between wave vector and wavelength in vacuum λ_0

$$k = \frac{2\pi}{\lambda} \rightarrow \lambda = \frac{2\pi c}{\omega} \tag{1.9}$$

The dispersion relation establishes a link between the temporal oscillation frequency of light† ($\omega = 2\pi f$, corresponding in the visible to frequencies around 500 THz) and its spatial oscillation frequency ($k = \frac{2\pi}{\lambda} = 2\pi\sigma$, with wavelengths around 500 nm). The connection between the two is ensured by the propagation speed of waves, here c.

Working in the monochromatic regime means fixing the working frequency ω. The wavelength is then deduced from the dispersion relation for this fixed ω.

* The resulting equation is well known : it is analogous to that of a harmonic oscillator with « frequency » $\frac{\omega}{c}$. But this « frequency » is expressed here in m^{-1} : it is a spatial frequency.

† Other wave phenomena, modeled by the same type of equations involving spatial and temporal derivatives, are also subject to dispersion relations. Consider the case of ocean waves : there exist models linking the propagation speed of water waves, the spacing between them, and the period between two successive crests at the same point. This link is mathematically expressed by the dispersion relation.

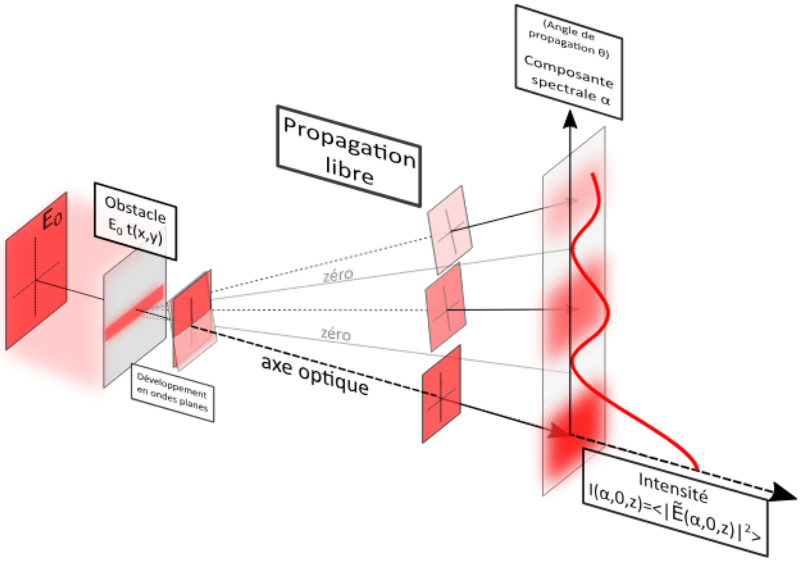

Figure 1.1 : Plane wave expansion and diffraction. The plane wave expansion reveals that an arbitrary field profile, as imposed by a slit illuminated by a single plane wave, can then be decomposed as a sum of plane waves. The diffraction pattern is the result of the propagation of these different plane waves, that acquire different dephasing along the propagation axis.

1.2 Plane-Wave Expansion : propagating and evanescent waves, far-field and near-field

The electromagnetic modes of free space are plane waves — functions whose spatial field distribution is perfectly known everywhere.

If we can express a given field in a particular plane as a sum of plane waves, then we can determine its distribution in another plane : farther away, the observed electric field is the superposition of these plane waves after propagation from the initial plane.

This approach, illustrated in Fig. **??**, is known as the **plane-wave expansion**, and it plays a fundamental role in wave optics and diffraction theory.

We wish to find the field distribution (x,y) in the plane $z = z_0$, $\tilde{\mathbf{E}}(x, y, z = z_0, \omega)$. From now on, the variable ω will be omitted, since it plays no essential role in the derivation. We start from the Helmholtz equation :

$$\nabla^2 \tilde{\mathbf{E}}(x,y,z) + \frac{\omega^2}{c^2} \tilde{\mathbf{E}}(x,y,z) = 0$$

The problem concerns propagation along the privileged direction z : we want to relate the field in the initial plane $z = 0$ to that in the plane $z = z_0$. To do so, we transform the Helmholtz equation into a one-dimensional form in z, using a spatial Fourier transform over x and y to eliminate the corresponding derivatives.

$$\tilde{\mathbf{E}}(x,y,z) = \iint \tilde{\mathbf{E}}(\alpha,\beta,z)\, \mathrm{e}^{\mathrm{i}(\alpha x+\beta y)}\, \frac{\mathrm{d}\alpha}{2\pi} \frac{\mathrm{d}\beta}{2\pi} \tag{1.10}$$

Here, α and β are the spatial-frequency counterparts of the temporal frequency ω. The inverse transform reads :

$$\tilde{\mathbf{E}}(\alpha,\beta,z) = \iint \tilde{\mathbf{E}}(x,y,z)\, \mathrm{e}^{-\mathrm{i}(\alpha x+\beta y)}\, \mathrm{d}x\mathrm{d}y \tag{1.11}$$

‡

Substituting into the Helmholtz equation gives :

$$\iint \left[\frac{\partial^2 \tilde{\mathbf{E}}(\alpha,\beta,z)}{\partial z^2} + \left(\frac{\omega^2}{c^2} - \alpha^2 - \beta^2 \right) \tilde{\mathbf{E}}(\alpha,\beta,z) \right] \mathrm{e}^{\mathrm{i}(\alpha x+\beta y)}\, \frac{\mathrm{d}\alpha}{2\pi} \frac{\mathrm{d}\beta}{2\pi} = 0 \tag{1.12}$$

Since the functions $\mathrm{e}^{\mathrm{i}(\alpha x+\beta y)}$ form an orthonormal basis, each spectral component must satisfy :

$$\forall\alpha, \forall\beta, \quad \frac{\partial^2 \tilde{\mathbf{E}}(\alpha,\beta,z)}{\partial z^2} + \left(\frac{\omega^2}{c^2} - \alpha^2 - \beta^2 \right) \tilde{\mathbf{E}}(\alpha,\beta,z) = 0, \tag{1.13}$$

or equivalently :

$$\frac{\partial^2 \tilde{\mathbf{E}}(\alpha,\beta,z)}{\partial z^2} + \gamma^2 \tilde{\mathbf{E}}(\alpha,\beta,z) = 0 \tag{1.14}$$

where we define

$$\gamma = \begin{cases} \sqrt{\dfrac{\omega^2}{c^2} - \alpha^2 - \beta^2}, & \text{if } \dfrac{\omega^2}{c^2} > \alpha^2 + \beta^2 \\ \mathrm{i}\sqrt{\alpha^2 + \beta^2 - \dfrac{\omega^2}{c^2}}, & \text{if } \dfrac{\omega^2}{c^2} < \alpha^2 + \beta^2 \end{cases} \tag{1.15}$$

Equation (1.14) is thus a one-dimensional Helmholtz equation describing propagation along z.

The two-dimensional Fourier transform has allowed us to eliminate the derivatives with respect to x and y. The general solution is then :

$$\tilde{\mathbf{E}}(\alpha,\beta,z) = A_+(\alpha,\beta)\mathrm{e}^{\mathrm{i}\gamma z} + A_-(\alpha,\beta)\mathrm{e}^{-\mathrm{i}\gamma z} \tag{1.16}$$

If γ satisfies the first condition in (1.15), the two terms represent waves **propagating** in the $+z$ and $-z$ directions respectively.

If γ is purely imaginary, they correspond to **evanescent waves** decaying exponentially with z. They decay from the diffraction plane, or in other words, from the plane containing the

‡ The sign convention in the exponential differs from that used for the temporal Fourier transform. This ensures that a positive spatial frequency $\alpha > 0$ corresponds to a wave propagating in the direction of increasing x, consistent with $\mathrm{e}^{\mathrm{i}(\alpha x-\omega t)}$.

material obstacle : the evanescent wave is said to be a bounded state of light, in the sense that its amplitude is bounded to the material obstacle.

The fact that, in the most general case, an electric field can be decomposed as a sum of propagating and evanescent waves, is one of the most important message of this chapter.

We now assume that the field propagates only toward increasing z, a standard boundary condition in free-space propagation. Thus :

$$A_{-}(\alpha,\beta) = 0.$$

The remaining amplitude $A_{+}(\alpha,\beta)$ is determined by the **initial condition** in the plane $z = 0$:

$$A_{+}(\alpha,\beta) = \tilde{\mathbf{E}}(\alpha,\beta,0).$$

Hence, the field at $z = z_0$ is given by :

$$\tilde{\mathbf{E}}(x,y,z) = \iint \left(\tilde{\mathbf{E}}(\alpha,\beta,0)\, \mathrm{e}^{\mathrm{i}(+\gamma z)}\right) \mathrm{e}^{\mathrm{i}(\alpha x+\beta y)} \frac{\mathrm{d}\alpha}{2\pi}\frac{\mathrm{d}\beta}{2\pi}$$

Combining the exponential leads to the key result :

Plane-Wave Expansion

$$\tilde{\mathbf{E}}(x,y,z = z_0) = \iint \tilde{\mathbf{E}}(\alpha,\beta,0)\, \mathrm{e}^{\mathrm{i}(\alpha x+\beta y+\gamma z_0)} \frac{\mathrm{d}\alpha}{2\pi}\frac{\mathrm{d}\beta}{2\pi} \tag{1.17}$$

This is the general solution to the problem of diffraction.

The field is thus expressed as the superposition of an infinite number of plane waves propagating in free space.

Each of these plane waves is defined by a pair (α,β), and by **a complex amplitude** $\tilde{\mathbf{E}}(\alpha,\beta,0)$, representing the modulus (weight) and phase of each plane wave in the decomposition. This is called the **plane-wave spectrum** of the initial field.

Plane-Wave Spectrum in the Initial Plane

$$\tilde{\mathbf{E}}(\alpha,\beta,0) = \iint \underbrace{\tilde{\mathbf{E}}(x,y,0)}_{\text{Given field}}\, \mathrm{e}^{-\mathrm{i}(\alpha x+\beta y)} \mathrm{d}x\mathrm{d}y \tag{1.18}$$

The wave vector k $= (\alpha,\beta,\gamma)$, where γ is related to (α,β) through Eq. (1.15). This corresponds to the dispersion relation for plane waves in free space : $k^2 = \alpha^2 + \beta^2 + \gamma^2 = \omega^2/c^2$.

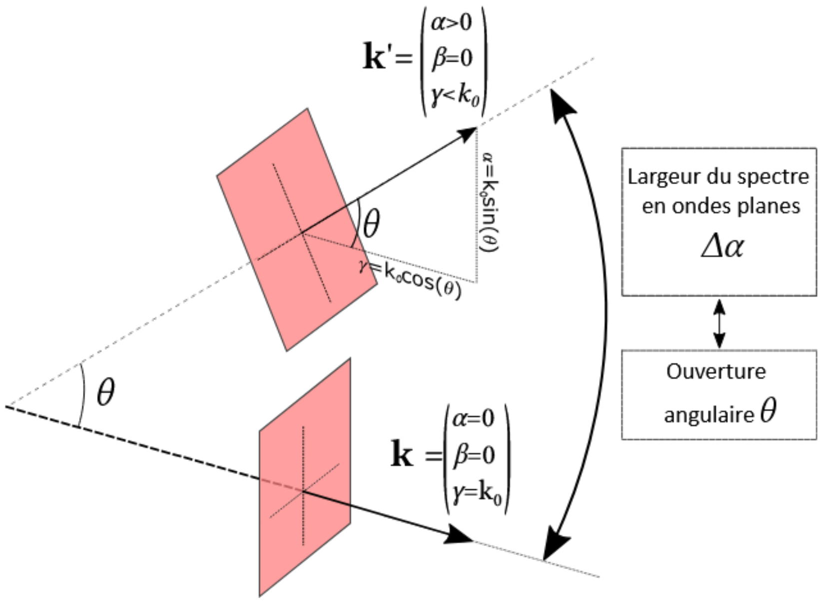

Figure 1.2 : Connection between propagation direction and wavevector components. The angular width of a beam is directly related to its width in terms of spatial frequencies...as long as the wavevector components are real.

Interpretation of the components of the plane wave expansion

We discuss here the mathematical and physical meaning of the different components of the plane wave spectrum.

Mathematically : (α,β) are two real numbers identifying one of the **spectral components** obtained in the plane-wave associated to a wave vector $\mathbf{k} = (\alpha,\beta,\gamma)$, and an amplitude $\tilde{\mathbf{E}}(\alpha,\beta,0)$.

The components α and β are the **transverse components** of the wave vector.

Two main physical meaning can be invoked :

Spatial frequencies

Each wave in the decomposition has its amplitude oscillating sinusoidally in space. This spatial oscillation is described by the term $\mathrm{e}^{\mathrm{i}(\alpha x+\beta y+\gamma z)} = \mathrm{e}^{\mathrm{i}(\alpha x+\beta y)}e^{i\gamma z}$: these are oscillations with frequencies $f_x = \alpha/2\pi$ along the x direction and $f_y = \beta/2\pi$ along the y direction.

Notion of Spatial Frequency

Thus, $\frac{\alpha}{2\pi}$ and $\frac{\beta}{2\pi}$ are called the **spatial frequencies** associated with the plane wave. A plane-wave decomposition can be viewed as a decomposition over a spectrum of spatial frequencies.

Strictly speaking, the quantities α and β are spatial angular frequencies. By a common abuse of language, they are sometimes also called spatial frequencies.

There is a direct relation between the spatial frequency content of an expanded spectrum and the size of an object or of its spatial features : by definition, a tiny object of size d will induce variations of the electric field on a scale comparable to d : this means that the diffracted field will contain components of spatial frequencies of the order of $\frac{2\pi}{d}$.

High spatial frequencies carry the information on the fine structures and details of a field, in analogy with high temporal frequencies explaining the fastest variations of a signal in time.

Propagation direction...or no propagation

(α,β) define the transverse components of the wave vector **k** and are directly associated with a **propagation direction** of a plane wave, provided $\alpha^2 + \beta^2 < \omega^2/c^2$.

The wave associated with the component ($\alpha = 0, \beta = 0$) propagates along the optical axis (z direction). As the values of α and β increase, the propagation angle relative to the optical axis also increases. The characteristic widths $\Delta\alpha$ and $\Delta\beta$ of the plane-wave spectrum are directly related to the beam's angular aperture, defining the extreme propagation directions of the plane waves involved.

When $\alpha^2 + \beta^2 > \omega^2/c^2$, γ is purely imaginary, corresponding to an **evanescent** wave with exponential decay along z. **The k-vector has now complex values : it can still be defined, but obviously, and there is no more direct geometrical interpretation of its physical meaning in terms of propagation direction**. This is simply because evanescent waves are...non-propagating wave, they cannot reach gently what is called the far-field in electromagnetism.

We see that this division in two ensembles of plane waves suggest that there is a boundary limit, that can be directly interpreted as a cut-off frequency when thinking in terms of spatial frequencies :

Propagation limit – spatial frequency cut-off

The domain of propagating waves is defined by the condition :

$$\alpha^2 + \beta^2 < \omega^2/c^2$$

where the boundary is a cut-off frequency.

In vacuum, only waves with spatial frequencies lower than ω/c can propagate into the far-field.

Conversely, all spatial frequency components larger than $\omega/c = \frac{2\pi}{\lambda}$ correspond to evanescent waves and will rapidly decay after the diffraction plane.

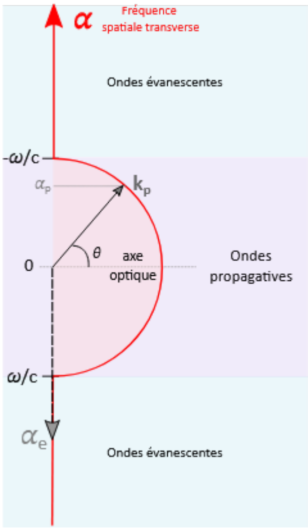

Figure 1.3 : Wavevector components in the propagative and evanescent regime.

In other words, all structures of the field in the diffraction plane $E(x,y,0)$ that are smaller than the wavelength are lost during propagation : this is a reformulation of the diffraction limit in terms of spatial Fourier components of the field.

A definition of the near-field

The previous discussion suggest that there is a domain of distance between a diffracting object and an observation plane where the amplitude of the rapidly decaying evanescent wave is still appreciable : we will try to define this region.

Let us first consider a diffracting object in the plane (x,z) with features on the order or smaller than the wavelength : after interaction with an incident plane waves, evanescent waves are diffracted by the object.

We consider a frequency much larger than the cut-off, corresponding to details of size d much smaller than the wavelength : $\alpha = \frac{2\pi}{d}$.

The longitudinal wavevector component is then :

$$\gamma = \sqrt{\frac{\omega^2}{c^2} - \alpha^2} \approx i\alpha$$

so that the associated plane wave decay is given by :

$$\exp(i\gamma z) \approx \exp(-\alpha z) = \exp(-\frac{2\pi z}{d})$$

whose typical decay length is given by $\delta = \frac{2\pi}{d}$.

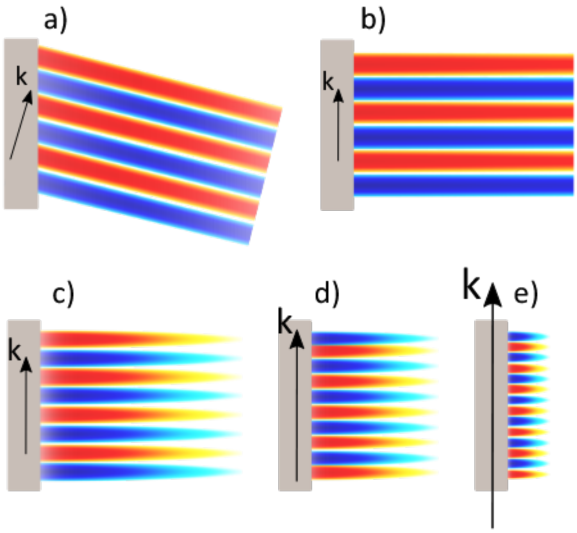

Figure 1.4 : Caption

Definition of the near-field

As often in electromagnetism, the relevant length scale is the wavelength λ.

As a rule of thumb, we see that, in order to detect evanescent waves corresponding to features smaller than the wavelength λ, the observation distance has to be smaller than $\frac{\lambda}{2\pi}$.

The distance $\frac{\lambda}{2\pi}$ is what defines the *near-field* region in optics.

1.3 Structure of the EM field in the near field region

The near-field structure correspond to the quasistatic limit : example of the electric dipole

In the far field region, the electric field produced by an oscillating dipole is called the **radiated** field. Radiation refers to the nature of propagating plane waves, that carry energy away, as indicated by their Poynting vector, always parallel to the wavevector (in isotropic media). The radiated field having the structure of a plane wave, we know that the electric and magnetic components are transverse, that is orthogonal to the observation direction, and propagate. Their respective amplitudes are easily related to each other using the plane-wave expressions of the vector operators, such as :$\mathbf{E} = i\mathbf{k} \times \mathbf{B}$, so that $E = cB$.

The field structure in the near field is radically different. In order to illustrate this point, we use an explicit example by looking at the analytical expression of the field produced by an oscillating dipole. Note that here, we consciously use the word "produce" and not "radiate" because the field now contains evanescent waves that are non propagating. The complex amplitude of the oscillating dipole electric field is given by Eq. (A.22) (with a complete derivation given as example in the Appendix) :

Complex amplitude of a field produced by an oscillating dipole $\mathbf{p_0}$

$$\tilde{\mathbf{E}}(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0} e^{ikr} \left[k^2 \frac{\mathbf{I} - \mathbf{u_r u_r}}{r} + (\mathbf{I} - 3\mathbf{u_r u_r}) \left(\frac{ik}{r^2} - \frac{1}{r^3} \right) \right] \cdot \mathbf{p}_0 \tag{1.19}$$

It is most important to notice that the field produced by an electric dipole contains three terms, respectively proportional to $1/r^3$, to $1/r^2$ and to $1/r^2$, with different preponderance depending on the observation distance with respect to the dipole.

Far-field limit

In the far field-region,the only preponderant term scales as $\frac{1}{r}$:

Field radiated by an oscillating dipole in the far-field limit

$$\tilde{\mathbf{E}}(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0} k^2 \frac{e^{ikr}}{r} [\mathbf{I} - \mathbf{u_r u_r}] \cdot \mathbf{p}_0 = \frac{1}{4\pi\varepsilon_0} k^2 \frac{e^{ikr}}{r} [\mathbf{p_0} - (\mathbf{p_0} \cdot \mathbf{r})\mathbf{r}] = \frac{1}{4\pi\varepsilon_0} k^2 \frac{e^{ikr}}{r} \mathbf{p}_{\mathbf{0},\perp} \tag{1.20}$$

where the subscript $\perp$ refers to the transverse component of the field (the vector itself minus its longitudinal component). Because it emerges from a point-like dipole, the field is *not* a plane wave : the wavefront are spherical (they are the surfaces of constant phase e^{ikr}) with an angular dependance of the amplitude. But they locally have a plane wave structure with a transverse character[§].

The $1/r$ amplitude decay is a necessary feature of radiated fields : because they carry energy away from a source, and in absence of decay (material losses along propagation), the outward flow of energy must be conserved : the average flux of Poynting vector through a closed sphere of radius R around the source must be constant for all R. The area of the sphere scales as $4\pi R^2$: this means that the field intensity must scale as $1/R^2$, and therefore the field amplitude scales as $1/R$.

Keep in mind that a $1/r$ dependence of a field amplitude allows you to directly identify a radiative contribution to the total field in the previous expansion.

Near-field limit - Electrostatic field of an electric dipole

Conversely, we see that in the limit $kr \ll 1$, the phase term is now negligible and only the r^{-3} is preponderant. The expression is now :

[§] Plane waves are the natural solution to field equations in complete vacuum (when there is no natural point-like origin for the generation of radiation).

Field radiated by an oscillating dipole in the near-field limit

$$\tilde{\mathbf{E}}(\mathbf{r}) = \frac{-1}{4\pi\varepsilon_0}\left[\frac{(\mathbf{I} - 3\mathbf{u_r}\mathbf{u_r})}{r^3}\right]\cdot \mathbf{p}_0 \quad \text{for} \quad kr \ll 1$$

Closest to the source, the field amplitude scales as $1/r^3$. It decays much faster than a radiative field. This correspond to a field that does not transport energy away from the source.

Upon inspection by connoisseurs, equation (A.2) is strictly the same as the electric field generated by an electrostatic dipole.

The regime $kr \ll 1$, corresponding to $r \ll \frac{\lambda}{2\pi}$: we are in the **near field**. The *electric* field produced by an oscillating electric dipole is strictly equivalent to the *electric* field instantaneously generated by the dipole as if it was a fixed electrostatic dipole, as defined in the quasistatic approximation.

This suggests another definition of the near-field : we define it as a spatial domain **where no retardation effects** are visible (hence the vanishing of the propagation exponential $\exp(ikr)$, and the consequence is that the field amplitudes, even if we formally still work in harmonic regime, is given at each time t by electrostatics.[¶]. This for sure has dramatic consequences : for example, it is very clear now that the fields are not exclusively transverse, as is the case with radiated fields in the far-field limit. Also, there are strong consequences in the respective amplitudes of the electric and magnetic fields, that are not constrained via $E = cB$. This is better illustrated by adding the exact expression of the magnetic field in the discussion.

Completing the asymptotic pictures with the magnetic field

An oscillating electric dipole also produces a magnetic field, that can be derived using the vector potential. The complete expression is :

Complex amplitude of the magnetic produced by an oscillating electric dipole $\mathbf{p_0}$

$$\tilde{\mathbf{B}}(\mathbf{r}) = \frac{\mu_0}{4\pi}\, e^{ikr}\left[(\frac{k^2}{r} + \frac{ik}{r^2})(\mathbf{u_r}\times\mathbf{I})\right]\cdot \mathbf{p_0} \qquad (1.21)$$

We identify easily a radiated magnetic field in $1/r$, that is coupled to the electric radiated field in the form of a electromagnetic plane wave.

There is a $1/r^2$ term similarly to the electric field (we have not discussed it so far, but may do so later on). However most importantly, one has to note that there is no $1/r^3$ term : in other words, in the near-field, it is expected that the electric field is much more intense than the magnetic field, that becomes asymptotically negligible compared to it.

¶ This result can be retrieved using another approximation : considering an infinite velocity of light, so that there is no retardation effects upon propagation.

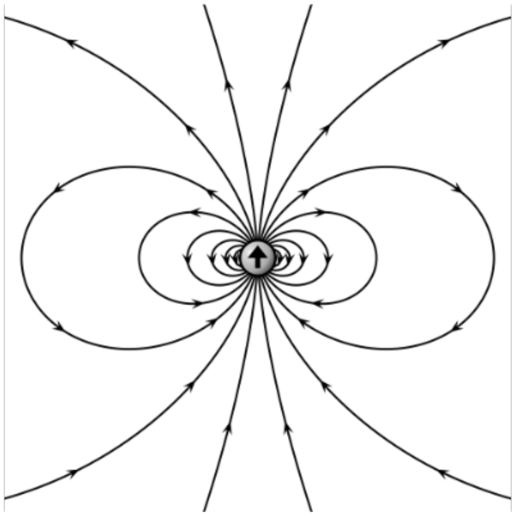

FIGURE 1.5 : Electrostatic dipole. Field lines of an electrostatic dipole. The field is non zero, even along the axis of the dipole, with a strong longitudinal component. In the direction perpendicular to the dipole orientation, the field is minimum.

Recall that we could identify that, in the near field E retains the character of the electrostatic field generated by an electric dipole. This makes completely sense : for an electrostatic dipole, the magnetic field is of course absolutely negligible !

The complete expression of the magnetic field therefore highlights the key message if this section : in the near field region of space, close to a distribution of charges (possibly induced by an external field illuminating matter), the EM field structure is directly given by the electrostatic (magnetostatic) field distribution induced by the charges (by the currents). It therefore strongly deviates from the usual plane wave structure, both in terms of field orientation (polarization) and in respective amplitudes of E and B.

Beyond the specific example of an oscillating dipole, a more general result can be derived by looking at the general expression of the electromagnetic potentials, the so-called retarded potential expression : in the near field, cancelling all retardation effects leads to the electrostatic potentials.

Near-field as the region where field structure follows circuit theory

The previous definition of the near field matches exactly the regime of the **quasistationary approximation in circuit theory**. Indeed, in electric circuits, when dealing with resistances, capacities and inductances, it is implied that all effects are instantaneous, in the sense that we consider no retardation of signals when propagating in the circuit (no phase added on the voltage or current signal when propagating in the wires !). This approximation regime is explicitly defined by $D \ll \lambda$ where D is the typical physical size of the circuit ! (This condition is most frequently forgotten). We therefore retain the features of electro- and magnetostatics, event if we work with time carying signals (the AC mode in a circuit).

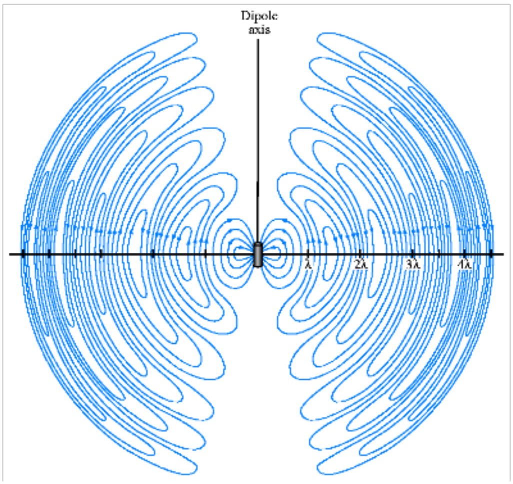

Figure 1.6 : Radiating dipole. Representation of the field lines of a radiating dipole. The near field region close to the dipole has the same structure as the electrostatic dipole.

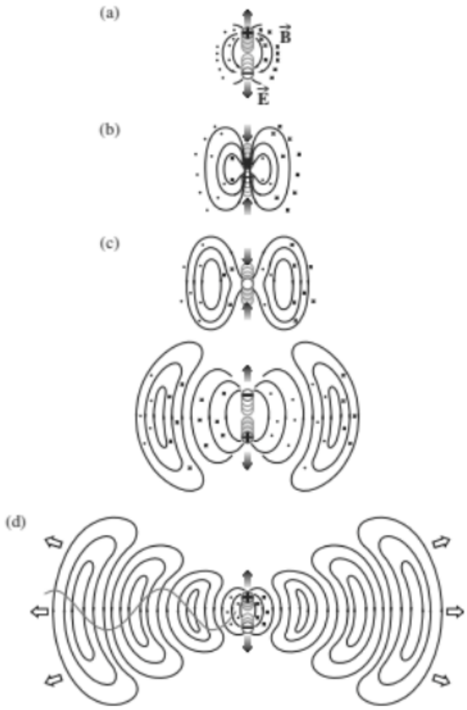

Figure 1.7 : Radiation, step by step. Waves are the result of the oscillation of charges (forming a radiating dipole), combined with a retardation effect along propagation. In the near field of the dipole, there is no retardation effect : the field has always the same structure as the one of an **electrostatic dipole**, but with a amplitude following the charge oscillation.

Circuit theory does not discuss in much details the structure of the EM field within and around a circuit. Few elements only are required to get a grasp on most of the physics of a circuit : resistances, capacitors, inductances.

Resistances are....resistive components. The definition of a resistive component is a component that dissipates electrical energy carried by the charges into another type of energy : it can be heat, or a mechanical motion....or light. This energy is therefore funneled out of the system and definitively lost. The amplitude of this resistive behavior is measured by the real part of the impedance of any given dipole – also called, the resistance.

Capacitors and inductances are **reactive** components : they are components that **store temporarily the energy provided by charges into a field – that field can then return this stored energy into the charges**. The ability of these two components to store and return energy is measured by their **reactance**, the imaginary part of the impedance.

Capacitors accumulate charges on their branches, therefore storing energy in a **strong electric field between the two branches**, with zero magnetic field. Conversely, **inductances** store energy by forming a **strong magnetic field**, **perpendicular to the coil axis**, with zero electric field.

This suggests ways to understand and predict the field structure in the near field region of an obstacle when light illuminates the object. One has first to consider the orientation of the illuminating field, then to picture the induced motion of charges in the object to figure out if the expected behavior is mostly one of a capacitor or of an inductance. Then, the structure of the EM field region can be inferred as corresponding to its quasistatic counterpart.

Let us take the example of a metallic sheet with a narrow, infinite subwavelength slit, perpendicular to the z axis. It definitely resembles a **capacitor**, with two subwavelength plates facing each other. When shining unpolarized, monochromatic light onto it, one can expect in the near field mostly a very intense electrostatic field component polarized perpendicular to the slit axis. Connecting to our initial discussion on plane wave expansion, this strong electric field is what results from the superposition of an incredible number of evanescent waves whose amplitude decays exponentially when moving away from the slit. All other EM field component (electric field along the slit direction, the two magnetic field polarizations) can exist, but are definitely much weaker.

1.4 Energetics of near and far field electromagnetics

The electromagnetic field carries energy

In classical electromagnetism, the electromagnetic field itself stores and transports energy. This energy can be viewed as being contained in the *electric field* E and in the *magnetic field* B.

The total electromagnetic **energy density** (energy per unit volume) is

$$u = u_E + u_B = \frac{1}{2}\varepsilon_0 E^2 + \frac{1}{2\mu_0} B^2, \tag{1.22}$$

where

$$u_E = \frac{1}{2}\varepsilon_0 E^2 \qquad \text{(electric energy density)}, \tag{1.23}$$

$$u_B = \frac{1}{2\mu_0} B^2 \qquad \text{(magnetic energy density)}. \tag{1.24}$$

The total electromagnetic energy in a volume V is therefore

$$U = \int_V u\, dV = \int_V \left(\frac{1}{2}\varepsilon_0 E^2 + \frac{1}{2\mu_0} B^2 \right) dV. \tag{1.25}$$

Electric and magnetic energy balance of near and far-field

In a most general fashion, for a time-varying electromagnetic field, energy contains both electric and magnetic contributions to the energy. We will show in this section that the Maxwell's equations can impose relations between the electric and magnetic field amplitudes, that, in consequence, affects the energy balance between both contributions.

We consider a 2D problem so that we can express the wavevector as $\mathbf{k} = k_x \mathbf{u_x} + \gamma \mathbf{u_z}$ for a propagation along z.

In the monochromatic regime, the electric field and magnetic field obey the Maxwell-Faraday equation : $i\omega \mathbf{B} = i\mathbf{k} \times \mathbf{E}$

As well as the Maxwell-Gauss equation : $i\mathbf{k} \cdot \mathbf{E} = 0$.

and finally, the dispersion relation is given by $k^2 = \frac{\omega^2}{c^2}$

Note that these equations are always valid...even when the wavevector has complex or imaginary components, which we expect to happen when considering evanescent waves in the near field. **Very importantly, note that the dispersion relation does not involve the squared modules of the wavevector, but the square of its complex amplitude !**

As suggested earlier, the main consequence is that in these situations, there is no clear geometrical interpretations (in terms of orthogonality between directions) to the dot or vector products.

The far-field regime : equipartition of energy for plane waves

In the far-field, the EM field has locally a plane-wave structure. Therefore, in what follows, we will regularly implicitly assimilate the physics EM field in the far-field with the physics of monochromatic propagating plane plane waves.

In the far-field,the wavavector component γ is real and positive, so that $k^2 = k_x^2 + \gamma^2 \frac{\omega^2}{c^2}$ with all terms being real and positive. This simple relation has several consequences

— In the far field, the amplitudes of the electric and magnetic field are related by $\omega^2|B|^2 = (k_x^2 + \gamma^2)|E|^2$, then $E = cB$. The relative amplitudes of electric field and magnetic field of a plane wave are perfectly constrained by this relation. A plane wave is necessarily associated to two fields **E** and **B** that oscillate in phase, with $E = cB$: for a given electric field amplitude, there can be no arbitrary magnetic field amplitude.

— By substituting this in the energy relations, we easily find $\frac{1}{2}\varepsilon_0 E^2 = \frac{1}{2\mu_0}B^2$, or in other words :

$$u_E = u_B = \frac{1}{2}\varepsilon_0 E^2 = \frac{1}{2\mu_0}B^2, \tag{1.26}$$

so energy is **equally shared** between the electric and magnetic fields, and propagates at the speed of light. This is another necessary consequence of the plane wace nature, that follows from the previous one : equipartition of energy between electric and magnetic contributions.

The near field regime

In the near field, we will show that the previous constraints break down, revealing once more the completely different physical behavior of propagating plane waves and near field composed of evanescent waves.

We first consider that we shine monochromatic light onto a small object, and we look at the near field in the vicinity of this object : in other words, we study the structure of evanescent waves bound to this object.

We first need to make an explicit choice of wave polarization, there is no general result. We consider working still in the in the (x,z) plane, and we arbitrarily choose to study s-polarized waves : this means focusing for now on waves with their electric field **E** is perpendicular to the plane of incidence, the (x,z) plane. In other words, the complex amplitude of the electric field reads $\mathbf{E} = (0,E,0)$.

We now use the Maxwell-Faraday equation with care, keeping for now all complex amplitudes :

$$\omega\mathbf{B} = E(-\gamma,0,k_x)$$

In the near field, the major contributions to the field come from evanescent waves : in other words, γ is purely imaginary, so that $\gamma = i|\gamma|$. Moving to the squared modules now makes a big difference with the far-field case :

$$\omega^2|B|^2 = (k_x^2 - \gamma^2)|E|^2$$

and there is no more simplification using the dispersion relation. We can relate both amplitudes of E and B in a more fashionable expression, valid only for evanescent wave, so for $k_x > k$:

$$|E| = \frac{c|B|}{\sqrt{2\frac{k_x^2}{k^2} - 1}}$$

We see that at the boundary, grazing angle incidence, we retrieve the propagating plane wave relation $|E| = c|B|$. However, for evanescent waves with $k_x \gg k$:

$$|E| = c|B| \frac{k}{\sqrt{2}k_x}$$

which shows that the s-polarized evanescent waves can have an electric wave much smaller than the magnetic field, compared to one of its plane wave counterpart.

The consequence for energy considerations is straightforward : there is no more necessary equipartition of electric and magnetic energy. For s-polarized waves, the magnetic energy is preponderant because the B field has intensity that goes much beyond the plane wave relation.

Energy balance in the near and far field regions

In the far-field, the radiated electromagnetic waves is similar to propagating plane waves, which obey the amplitude relation $E = cB$ as well as energy equipartition.

The near field contains essentially contributions from evanescent waves, who obey the same Maxwell's equation as propagating plane waves, but whose complex wavevector dramatically changes the structure. The electric and magnetic field decouple : they can have unrelated amplitudes.

In the near field, for s-polarization, as defined by what is considered the incidence plane, the EM field is essentially a magnetic field - and therefore following the above results, reduces to the magnetostatic field distribution, with preponderant magnetic energy and small electric field contribution.

Conversely, for p-polarization, the near field is essentially electric. Recall the example of the oscillating electric dipole, or to the example of a narrow infinite slit : the near field reduces to a very intense, capacitive electric field, as expected in electrostatics. The field orien

This has a very general consequence : if one wants to build up an extremely intense electric field, this means accumulating a strong amount of electric energy. Using some "intense" propagating beam for this purpose requires to pay the same amount of energy for the magnetic contribution and to build up also an intense magnetic field, because of the necessary equipartition of energy in propagating waves.

In the near-field, it is possible to buildup strong fields without this constraint. In the vinicity of a nanostructure under illumination, we basically add up field amplitude associated to the many evanescent waves accessible in this region of space. We have access to the energy stored in the near-field. The question is now : where does this energy come from ?

Recall that we work in the monochromatic regime, which is a steady-state, yet driven regime. In steady state, the incident energy ends up indeed being either reflected, transmitted (radiated) or absorbed (dissipated) and the total energy is conserved : there cannot be any storage and accumulation. The energy of the near-field was actually accumulated before reaching rhe steady state regime, when building up the near fields in the very first moments of the illumination of the structure, and before radiation could occur and balance the incoming flow of energy with an outgoing one.

The point of accessing this stored energy to reach spectacular field overintensities is especially important when trying to couple emitters or other objects carrying a dipole moment **d** to antennas or nanostructures, as their coupling scales as $\mathbf{d} \cdot \mathbf{E}$.

Completing the analogy between circuit physics, near and far field

To conclude this discussion on the energetics of the field, we go back to the microscopic description of the interaction of between light and matter in classical physics : an external monochromatic field drives the motion of charges, that we describe using the oscillating electric dipole picture. We already explained that the predominant field close to the charges corresponds to the fields observable in circuit physics. We now connect in the framework of circuit physics the mechanical motion of the charges, the buildup of EM fields, and radiation.

Meaning of electric energy

In electrostatics, assembling charges requires work against Coulomb forces ; this work becomes stored in the electric field :

$$U_E = \frac{1}{2}\varepsilon_0 \int E^2 \, dV. \tag{1.27}$$

For a capacitor, this gives the familiar expression

$$U_E = \frac{1}{2} C V^2,$$

showing that the energy is stored in the electric field between the plates.

Hence, u_E measures the amount of **potential energy per unit volume** stored in the electric configuration of charges at rest.

Meaning of magnetic energy

In circuits, work is required to establish a current in an inductor, because an induced electromotive force (EMF) resists the change of current (this is Lenz's law). The energy supplied is stored in the magnetic field :

$$U_B = \frac{1}{2\mu_0} \int B^2 \, dV = \frac{1}{2} L I^2. \tag{1.28}$$

Thus, u_B represents **energy per unit volume associated with currents and magnetic interactions**. It is analogous to kinetic energy in a mechanical system, since it is associated with the "motion" (currents) of charges.

Charge motion in an electric dipole as a damped harmonic oscillator

When driven by an external monochromatic field, our oscillating dipole continuously exchange energy between two forms : potential and kinetic energy. This is similar to the case of a mechanical harmonic oscillator – we however have here to consider two physical systems that have a time evolution : the charge itself, and the field that plays the role of the restoring force of the spring. In the case of the electric dipole, we saw that in the near field, the field is almost exclusively an electric field following the electrostatic field distribution.

In short, in the oscillatory motion of the charges of the electric dipole, energy is continuously exchanged between the electric field (storing potential energy), and the charge (storing kinetic energy). In a perfect mechanical oscillator, this would mean an evolution of the position of the mass and of its velocities described by two oscillations in phase quadrature : it describes perfect storing and restitution of energy with zero flux of energy in average with no dissipation. This illustrated that the near-field is reactive : it does not dissipate energy, it just acquired some that it periodically gives back to matter before retrieving later.

However, the laws of electromagnetism are a bit more complex than the mechanical motion of a spring and mass system. Two additional effects are linked to the oscillatory motion of the charge :

— First, because the **electric field and the dipole moment vary in time** (it matches the electrostatic field at all times but still varies), there is a buildup of a magnetic field by **induction** (Lenz law). This magnetic field also varies so that another field component is coupled to this magnetic field by induction as well. These two fields are unavoidable, they are present because of the time variation of the fields, in other words, because of the **first order time derivative of the dipol**e. We have encountered them without further commenting : in the case of the electric dipoles, these inductive contributions are actually the $1/r^2$ terms, present both in the electric and in the magnetic field produced by the electric dipole. They are significant in an intermediate region called the « **inductive near-field region** ». It can be shown that the corresponding E and B are in phase quadrature, which mean that they do not dissipate energy over time : the Poynting vector associated to this coupled electromagnetic field points periodically outwards from the dipole, then inwards, like energy is flowing away, then back onto the dipole, with no net outgoing energy flux in average. The fields in this region are said to be reactive, they retain the energy bounded to the system.

— A second contribution is related to the **second order time derivative of the dipole moment** : in other words, the charge acceleration, that is non zero because we do have an oscillatory motion. **Acceleration of charges** is what gives rise to **radiation**. The corresponding E and B fields are **in phase**, meaning that indeed, there is a net flow of energy outside of the system, in other words, **dissipation** : the **Poynting vector constantly points outwards**. The fields are said to be **resistive** : they carry energy definitely lost for the system.

This forms a picture to describe near and far field from an electric dipole, in terms of energy exchange in an equivalent circuit. An oscillating electric dipole behaves like a driven RLC circuit. The energy is periodically stored in the capacitor (= electrostatic field of the dipole) or in the inductance (=motion and kinetic energy of the charge, which is indeed current).

The unavoidable radiation by the oscillating charge is a damping process : radiation is modeled by a resistance in the equivalent circuit. The value of this resistance dissipating energy is not arbitrary. A resistance relates voltage and current amplitude via Ohm's law. Here the resistance value is imposed by Maxwell's equations via the $E = cB$ amplitude relation. The vacuum is said to have an impedance $Z_0 = (\mu_0/\epsilon_0)^{1/2} \approx 377\Omega$. Plane waves can only be formed by extracting equal amounts of electric and magnetic energy from the charges motion in the near field : this condition is ensured by the vacuum impedance.

1.5 Energy confinement in the near field

As in the previous parts of this chapter, we will explore near field physics through a comparison with the far field and the plane wave physics.

One of the key challenges in optics s to confine light, and therefore EM energy into the tiniest spot possible to enhance light-matter interactions.

Let us discuss what are the physical limits to this confinement.

How much can a plane wave confine light?

For plane wave confinement limit, we obviously have in mind that the diffraction limit, gives already an answer : it is not possible to focus light in a spot of transverse dimensions typically smaller than $\frac{\lambda}{2}$. We will first try to retrieve this resul using only energetic considerations.

FOr plane wave, the energy density oscillates and is equally shared between electric energy $\epsilon_0 E^2/2$ and magnetic energy $B^2/2\mu_0$.

The fields are related by the Max-well Faraday equation $\nabla \times \mathbf{E} = i\omega\mathbf{B}$.

We now suppose that the energy is confined in a volume of size a^3 where a is the lateral dimension. We can use this assumption ton introduce a length scale in Maxwell-Faraday equation : indeed, stating that the energy is confined over a lateral distance a is close to saying that the typical length scale of variation of the field amplitude is a. Recall that the curl of a field is a spatial derivative, it is close to the amplitude of the field over its variation scale :

$$\frac{E}{a} \approx \omega B$$

And now using the balance between electric and magentic energy :

$$\epsilon E^2 \approx E^2/[\mu_0(a\omega)^2]$$

which directly suggests $a^2 = \omega^2/c^2$ and therefore $a \approx \lambda/2\pi$!

The previous result illustrate what we otherwise well know : the length scale of variation of the electric field, and therefore of the confinement limit, is, at given frequency and for a plane wave, invariably given by the corresponding wavelength, as imposed by the dispersion relation : this is a fundamental limit of the energetics of the plane wave, described as a coupled EM field forming a harmonic oscillator.

Confinement limit in the near field

The previous limit falls in the near field for many reasons : as we saw, the balance between electric and magnetic field is not valid anymore, so that the typical length scales of the field can be much smaller than the wavelength ‖. This is what suggests the fact that spatial frequencies of evanescent waves can far exceed the vacuum wavevector.

In addition, let us recall that evanescent waves are generated through interaction with charges with matter only. The fact that evanescent waves can store energy in volumes much smaller than the wavelength is one of its very logical consequence : let us consider an hydrogen atom, with one electron and one proton, forming a radiating dipole The charges oscillate on a distance of typically one Angstrom : they form a mechanical oscillator, storing potential and kinetic energy in a very small volume, much smaller than the wavelength. The near field directly translates as an electrostatic field the deeply subwavelength features of this oscillator and therefore can also store electromagnetic energy on the same scales.

In summary, a solution to store energy at deeply subwavelength volumes is to use material systems, carrying charges whose motion can imply length scales much smaller that the wavelength. By our first definition of the near-field, these systems contribute to the radiation of many evanescent waves.

Concepts and key ideas

Introduction to near-field optics

- Plane-Wave Expansion of the EM field in a propagation problem.
- Propagating and evanescent plane, and there interpretation in terms of amplitude evolution, propagation directions, or spatial frequencies.
- The near field as a region close to a scattering object, with significant amplitude for evanescent waves, or of negligible retardation in the context of dipole radiation.
- Near field structure as being equivalent to electrostatic field structures ; in contrast with the important far field features (transverse plane wave structure)
- Evanescent waves and the near field as possibilities to confine and store energy below the diffraction limit, in the subwavelength regime.

‖ We will come back on this point when discussing the resolution limit in the next chapter.