Phys 431 - Lecture 10

Wave superposition

1.) Introduction

While the geometric optics discussion is valid in the limit where λ is much smaller than the length scale of the optical system, many more situations arise where this is no longer valid and the wave nature of light has to be incorporated. Phenomena like polarisation, interference and diffraction all rely on the concept of wave superposition. Thus, we want to discuss how phase, frequency, amplitude, etc. of the individual components influence the final optical disturbance. This discussion will rely on the properties of waves discussed in Part I of this course, and be primarily based on the principle of superposition. It states that the resulting disturbance is the algebraic sum of the different constituent waves and non-linear effects can be neglected. The component of the EM field that we will study is the electric field and we now focus on scalar quantities $E = E(\vec{r}, t)$.

2.) Adding waves of same frequency

Consider two harmonic waves of frequency ω traveling in x direction

propagation number

$$E_1 = E_{01}\cos(kx - \omega t + \varphi_1),$$

$$E_2 = E_{02}\cos(kx - \omega t + \varphi_2).$$

the different phases could e.g. be due to different path lengths $k\Delta x_i$

For convenience, we write this as $\alpha_i = \varphi_i + kx$

$$E_i = \mathrm{Re}\left[E_{0i}\cdot e^{i\varphi_i} e^{i(kx-\omega t)}\right] = \mathrm{Re}\left[E_{0i}\, e^{i\alpha_i} e^{-i\omega t}\right]$$

and thus obtain for the superposition of both fields

$$E = E_1 + E_2 = \mathrm{Re}\left[\underbrace{\left(E_{01}e^{i\alpha_1} + E_{02}e^{i\alpha_2}\right)}_{\text{amplitude } E_0 \text{ at } x} e^{-i\omega t}\right].$$

The amplitude E_0 at position x can be related to the intensity

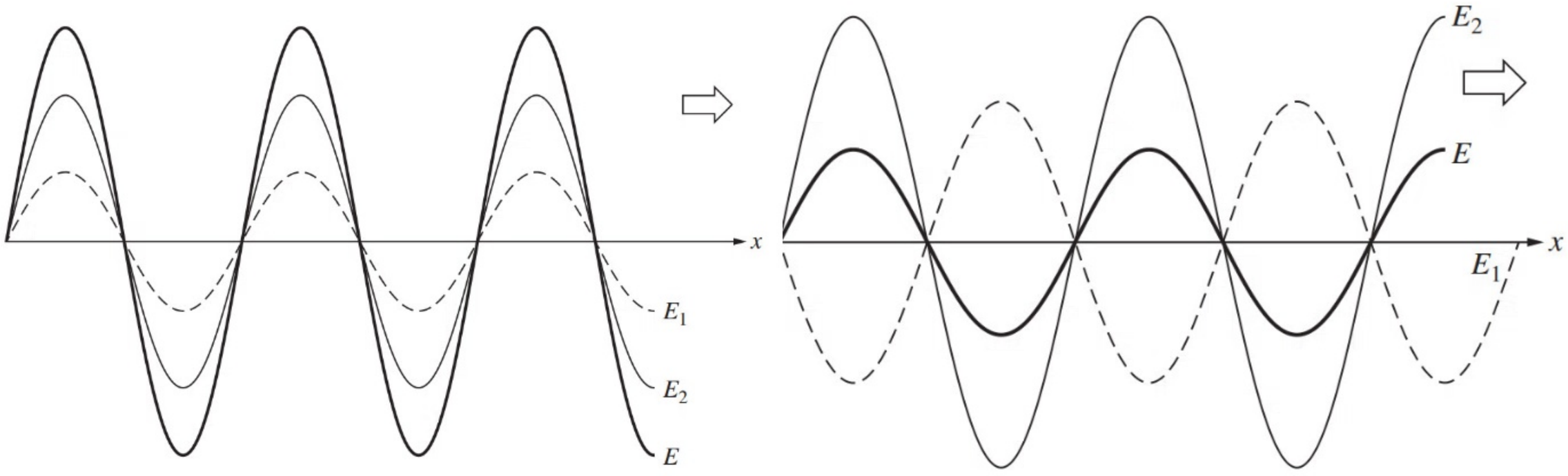
$$\langle \text{intensity} \rangle \propto |E_0|^2 = \left(E_{01}e^{i\alpha_1} + E_{02}e^{i\alpha_2}\right)\left(E_{01}e^{-i\alpha_1} + E_{02}e^{-i\alpha_2}\right)$$

$$= E_{01}^2 + E_{02}^2 + E_{01}E_{02}\left[e^{i(\alpha_1-\alpha_2)} + e^{-i(\alpha_1-\alpha_2)}\right]$$

$$= E_{01}^2 + E_{02}^2 + 2E_{01}E_{02}\cos(\alpha_1-\alpha_2)$$

where the second term in the first line is the complex conjugate. This is not simply the sum of the individual amplitudes squared but contains a cross term that is also known as the interference term. Its size is crucially influenced by the relative phase between the two interfering waves; i.e. if $\delta = \alpha_1 - \alpha_2 = 0, \pm 2\pi, \pm 4\pi, \ldots$ the resultant amplitude is maximum

Whereas for $\delta = \pm\pi, \pm 3\pi, \ldots$ the amplitude is minimal.



$\delta = 0, \pm 2\pi, \ldots$
constructive interference
for $E_{01} = E_{02}$ and $E_0 = 2E_{01}$

$\delta = \pm\pi, \pm 3\pi, \ldots$
destructive interference
for $E_{01} = E_{02}$ and $E_0 = 0$

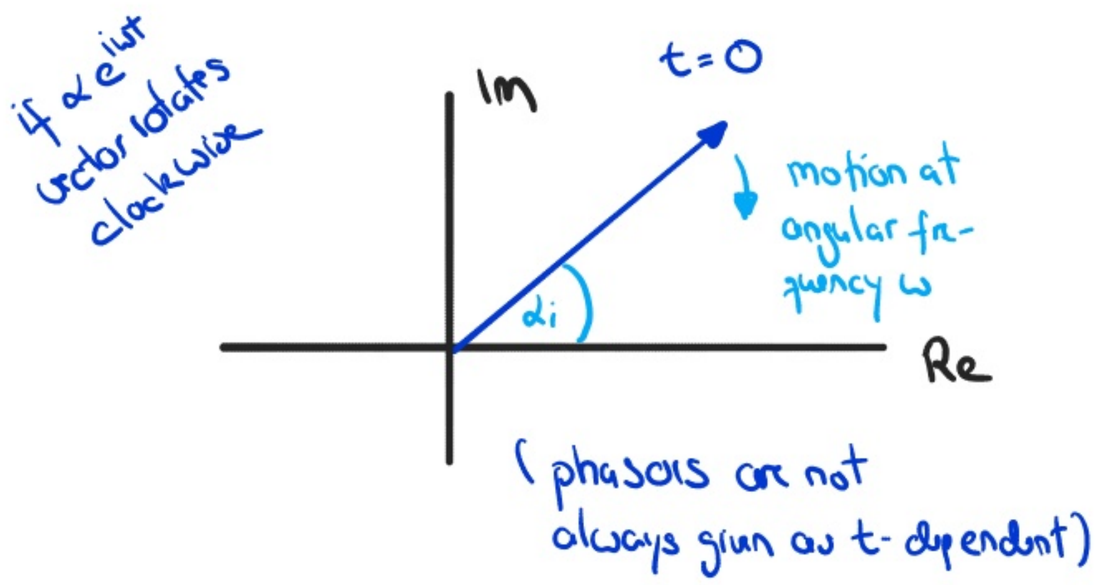
We can generalise the same concept to N waves to obtain

$$E = \sum_{j=1}^{N} E_j = \mathrm{Re}\left[\sum_j E_{0j} e^{i\delta_j} e^{-i\omega t}\right] = \mathrm{Re}\left[E_0 e^{-i\omega t}\right],$$

with a complex amplitude $E_0 = \sum_j E_{0j} e^{i\delta_j}$. The resulting disturbance is thus again harmonic and of the same frequency, but has a different amplitude and phase. As before, the intensity is proportional to $|E_0|^2$. Consider two light bulbs whose atoms emit photons that all together manifest themselves as an electromagnetic wave. As the underlying disturbances vary randomly and rapidly in phase, the different contributions from $\alpha_j(t)$ will average out to zero in $|E_0|^2$. Hence, the incoherent light of two light bulbs will not allow us to see any interference.

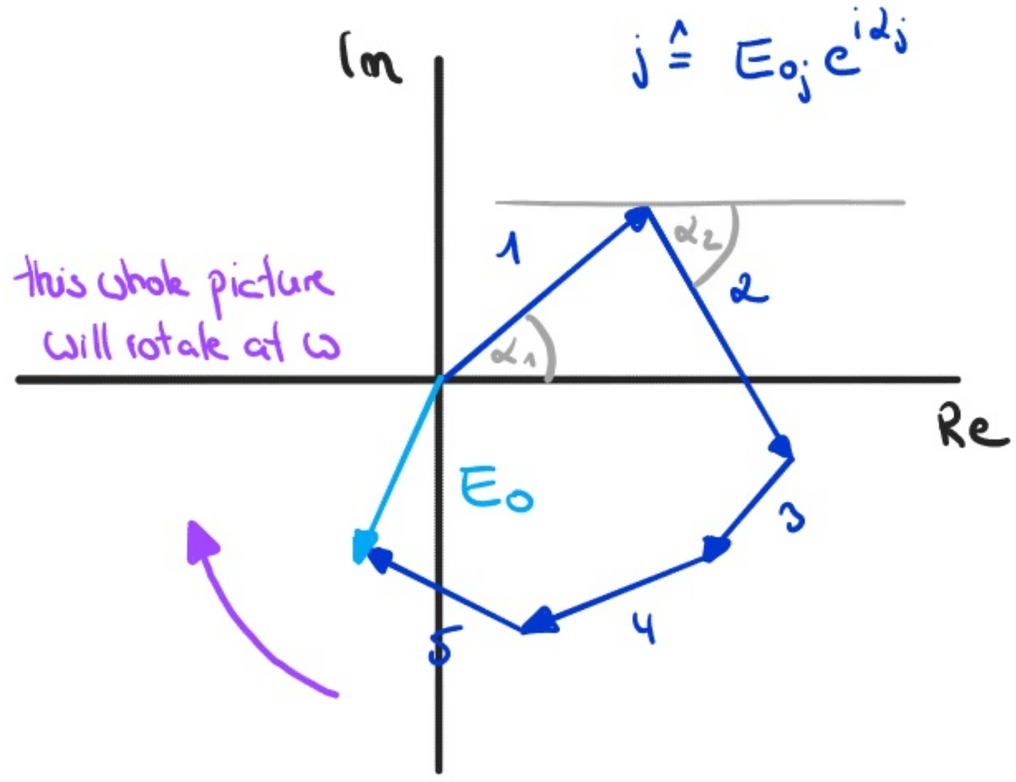
Only for coherent sources will we be able to see interference, which typically alters the energy distribution, i.e. there will be regions of flux density greater than the average as well as smaller than the average.

The amplitude summation in $E_0 = \sum_j E_{0j} e^{i\delta_j}$ can be represented graphically as an addition of vectors in the complex plane. You can essentially think of $E_{0j} e^{i\delta_j} e^{-i\omega t}$ as a vector rotating in the complex plane, i.e.



In this so-called phasor notation, E_0 can be evaluated by summing the individual components at $t=0$:

Since all of the vectors are rotating at the same angular frequency, the length of the complex amplitude will not change, only its angle w.r.t. the real axis.



This graphical techniques can be very useful when summing over many (or even $N \to \infty$) waves, especially if you find symmetries in the picture. E.g. in the above phasor diagram, we observe straight away that there is little constructive interference and E_0 is about the same size as E_{01}.

We can now consider what happens when two waves travelling in opposite directions hit each other. For the resultant field, we obtain

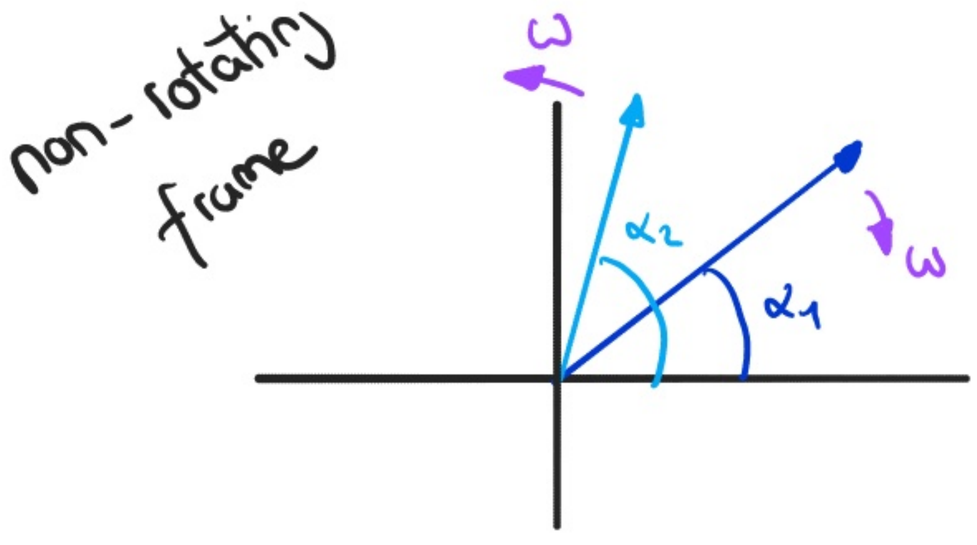
travelling in opposite direction

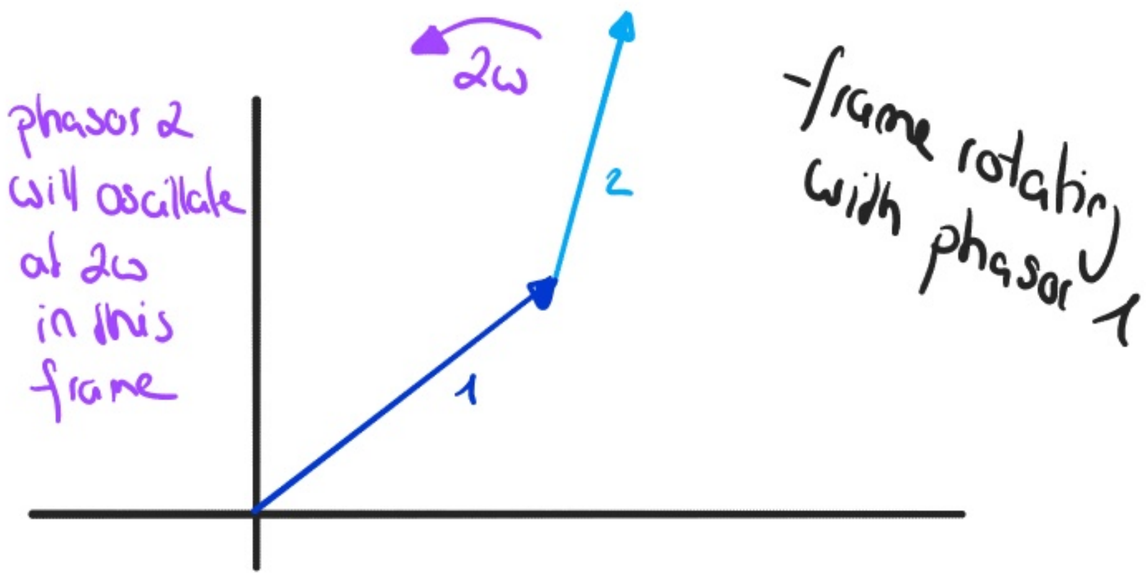
$$E = E_{01}\cos(kx - \omega t + \varphi_1) + E_{02}\cos(kx + \omega t + \varphi_2)$$

$$= \mathrm{Re}\left[E_{01}e^{i\alpha_1}\underline{e^{-i\omega t}} + E_{02}e^{i\alpha_2}\underline{e^{i\omega t}}\right]$$

phasors rotate in opposite directions

Using the phasor technique, we can sketch this in two ways





Using $E_{01} = E_{02}$, we find

$$E = E_{01}\,\mathrm{Re}\left[e^{i\alpha_1/2}\,e^{i\alpha_1/2}\,e^{-i\omega t} + e^{i\alpha_2/2}\,e^{i\alpha_2/2}\,e^{i\omega t}\right]$$

$$= E_{01}\,\mathrm{Re}\left[e^{i(\alpha_1+\alpha_2)/2}\left(e^{i(\alpha_1-\alpha_2)/2}\,e^{-i\omega t} + e^{i(\alpha_2-\alpha_1)/2}\,e^{i\omega t}\right)\right]$$

$$= E_{01}\,\mathrm{Re}\left[e^{i(\alpha_1+\alpha_2)/2}\,2\cos\left(\frac{\alpha_1-\alpha_2}{2} - \omega t\right)\right]$$

$$= 2E_{01}\cos\left(\frac{\varphi_1-\varphi_2}{2} - \omega t\right)\mathrm{Re}\left[e^{ikx}\,e^{i(\varphi_1+\varphi_2)/2}\right]$$

$$= 2E_{01}\cos\left(\frac{\varphi_1-\varphi_2}{2} - \omega t\right)\cos\left(kx + \frac{\varphi_1+\varphi_2}{2}\right).$$

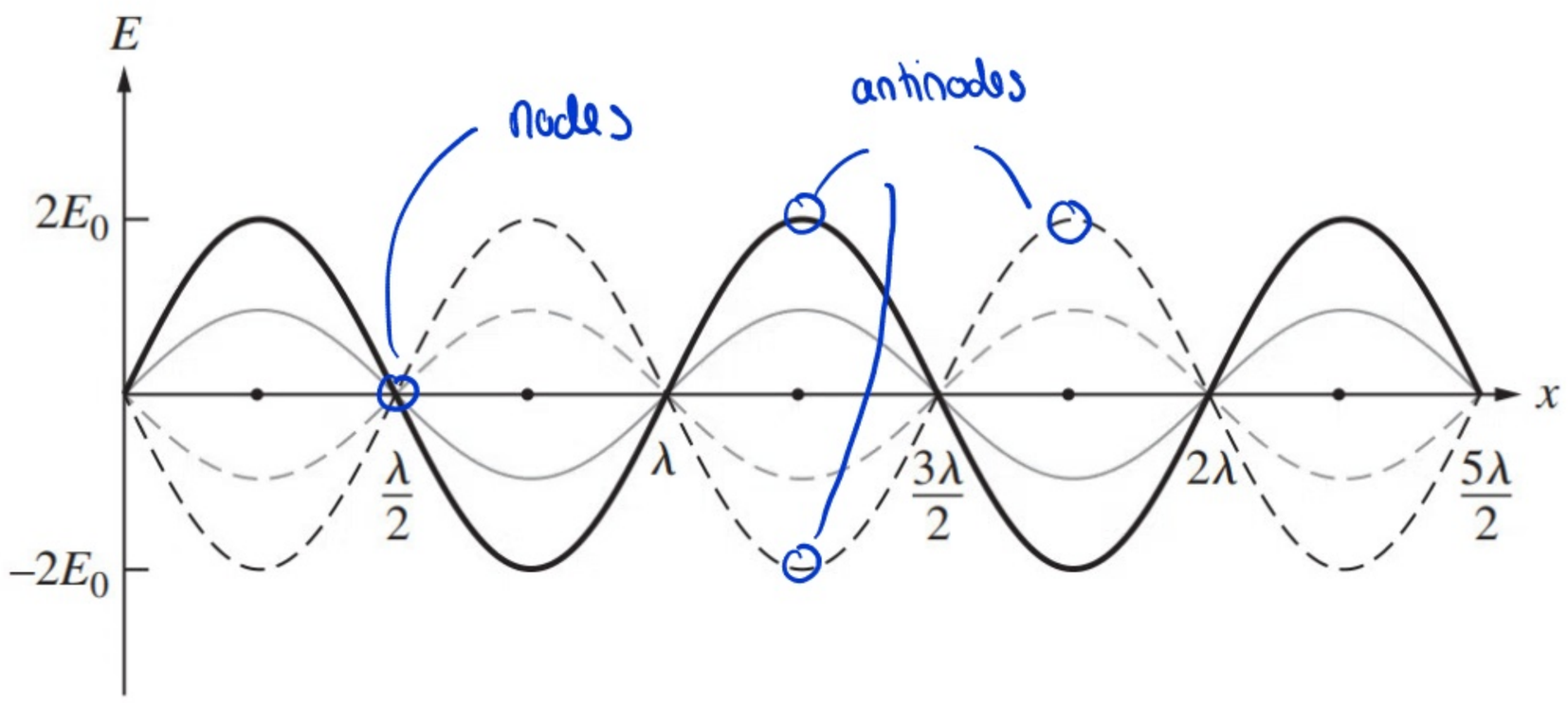
This is no longer a travelling wave, but a **standing wave** whose profile does not move through space. Its exact form will depend on the exact values of φ_1 and φ_2, which in turn are determined by the **boundary conditions** of the system of interest. Setting $\varphi_1 = 0$ without loss of generality, we e.g. require that $E(x,t)$ vanishes at a perfectly conducting surface for all times. If this surface is positioned at $x = 0$, we find that $E(0,t) = 0 \ \forall t$ is only possible if $\varphi_2 = \pi$ so that $\cos(\varphi_2/2) = 0$. In this case, the **total field** reduces to

$\cos(y \pm \pi/2) = \mp \sin(y)$

$$E(x,t) = 2E_{01}\cos(-\pi/2 - \omega t)\cos(kx + \pi/2)$$

$$= 2E_{01}\sin(\omega t)\sin(kx).$$

At certain points, this disturbance will vanish for all times (not only at $x = 0$). Since $k = 2\pi/\lambda$, $\sin(2\pi x/\lambda)$ vanishes for $x = \lambda/2$, λ, $3\lambda/2$, ... These points are **nodes** or **nodal points**. Halfway between each adjacent node, that is $x = \lambda/4, 3\lambda/4, 5\lambda/4, \ldots$, the amplitude has a maximum value of $2E_{01}$. These points are referred to as **antinodes**. Moreover, the disturbance will be zero for all values of x whenever $\sin(\omega t) = 0$, that is when $t = 0, \tau/2, 3\tau/2, \ldots$ where τ is the period of the waves (remember $\omega = 2\pi/\tau$). This can be illustrated as



Such standing mode patterns can also occur in higher dimensions with spatial distributions that will crucially depend on the exact shape of the respective boundary conditions.

Remember: Materials interact much more strongly with the electric field than with the magnetic field, that's why we use the E field to describe the 'light field' or 'optical disturbance'.

2.) Adding waves of different frequency

So far, we have restricted our attention to the superposition of waves that have the same frequency. Realistic disturbances are, however, rarely perfectly monochromatic. Light is very often quasi monochromatic consisting of a narrow range of frequencies.

Consider two waves of different frequency, travelling in the same

direction. We can write them as

$$E_1 = E_{01}\cos(k_1x - \omega_1 t), \quad E_2 = E_{02}\cos(k_2x - \omega_2 t).$$

Their sum will then be given by (using $E_{01} = E_{02}$)

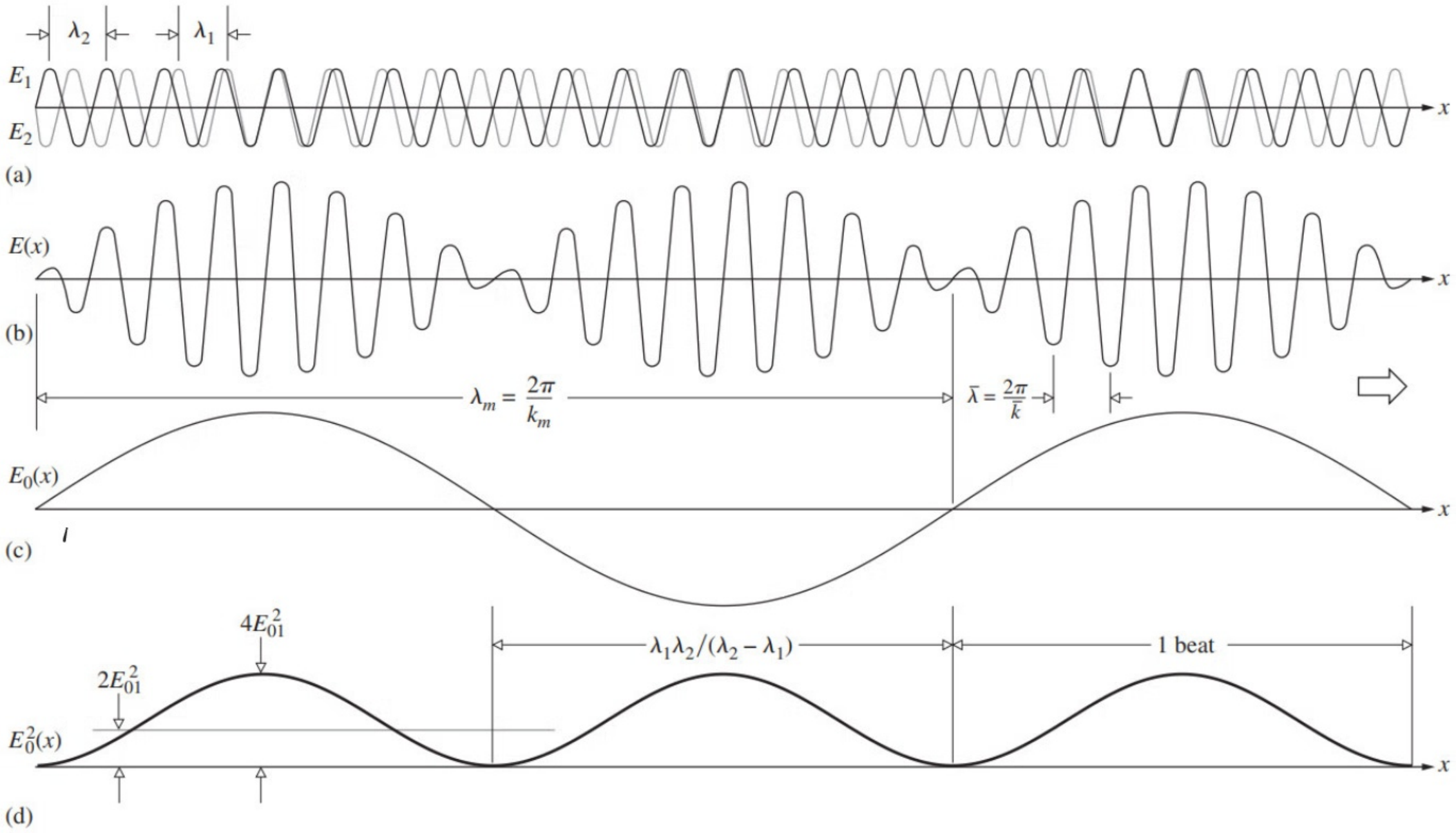
$$\begin{aligned} E &= E_{01}\cos(k_1x - \omega_1 t) + E_{02}\cos(k_2x - \omega_2 t) \\ &= 2E_{01}\cos\left[\tfrac{1}{2}(k_1+k_2)x - \tfrac{1}{2}(\omega_1+\omega_2)t\right] \\ &\quad \times \cos\left[\tfrac{1}{2}(k_1-k_2)x - \tfrac{1}{2}(\omega_1-\omega_2)t\right] \\ &= \underbrace{2E_{01}\cos\left(\frac{\Delta k x - \Delta\omega t}{2}\right)}_{\text{amplitude is modulated } E_0(x,t) \Rightarrow \text{'beating'}} \underbrace{\cos(\bar{k}x - \bar{\omega}t)}_{\text{travelling wave} \Rightarrow \text{'carrier'}}, \end{aligned}$$

where we have defined averages and differences as

carrier → $\bar{k} = \tfrac{1}{2}(k_1+k_2)$, and $\bar{\omega} = \tfrac{1}{2}(\omega_1+\omega_2)$,

beat → $\Delta k = k_1 - k_2$, $\Delta\omega = \omega_1 - \omega_2$.

The total disturbance can thus be interpreted as a travelling 'carrier' wave with a modulated amplitude $E_0(x,t)$. Note that if $\omega_1 \simeq \omega_2$ then $\Delta\omega \simeq 0$ and the time-modulation will be very small, while $E(x,t)$ overall might vary quite rapidly. See next page for figure.



The last row of this figure shows $E_0^2 \propto \cos^2(\Delta\omega t/2) = 1/2(1+\cos\Delta\omega t)$, illustrating that the average intensity oscillates about an average value of $2E_{01}^2$. This is the reason, why $\Delta\omega$ is also known as the beat frequency.

3.) Group velocity

Note that the carrier in the above picture can move at a different velocity than the overall modulation profile (or the 'envelope'), because the specific relationship between ω and k determines the velocity of the respective part. More precisely, when a number of different-frequency harmonic waves form a composite disturbance, the envelope will typically

travel at a speed different from that of the constituent waves. This introduces the important notion of the group velocity and how it differs from the phase velocity, we have discussed before. So generally speaking, if you recognise some constant feature in the shape of a pulse, the rate at which this feature moves is the group velocity of the waves as a whole.

Focusing on a disturbance of the form $E(x,t) = E_0(x,t)\cos(\overbrace{\bar{k}x - \bar{\omega}t}^{=\varphi})$.
Taking the amplitude to be constant for a moment, we can determine that each small peak in the carrier will move to the right with the usual phase velocity

$$v = -\frac{(\partial\varphi/\partial t)_x}{(\partial\varphi/\partial x)_t} = \frac{\bar{\omega}}{\bar{k}} .$$

The modulation envelope however advances at the group velocity v_g.
Focusing on $E_0(x,t) = 2E_{01}\cos\left(\frac{\Delta k x - \Delta\omega t}{2}\right)$, we extract the velocity

$$v_g = \frac{(\partial\varphi/\partial t)_x}{(\partial\varphi/\partial x)_t} = \frac{\Delta\omega}{\Delta k} .$$

As typically in dispersive media, we have a dispersion relation of the form $\omega = \omega(k)$, the previous expression equals the derivative provided that $\Delta\omega$ and Δk are small

$$\Rightarrow \quad v_g = \left.\frac{d\omega}{dk}\right|_{\bar{\omega}} .$$

Since $\omega = v \cdot k$, we can use the product rule to write

$$v_g = \frac{d}{dk}(kv) = v + k \cdot \frac{dv}{dk}$$ ← in a non-dispersive medium this is 0 and $v_g = v$

or equivalently with $\omega = ck/n$

$$\underline{\underline{v_g}} = \frac{d}{dk}\left(\frac{ck}{n}\right) = \frac{c}{n} - \frac{kc}{n^2}\frac{dn}{dk} = \underline{\underline{v\left(1 - \frac{k}{n}\frac{dn}{dk}\right)}}.$$

Hence, we obtain for normal optical media where $dn/dk > 0$, that $v_g < v$, as discussed before qualitatively.

Phys 434 - Lecture 11

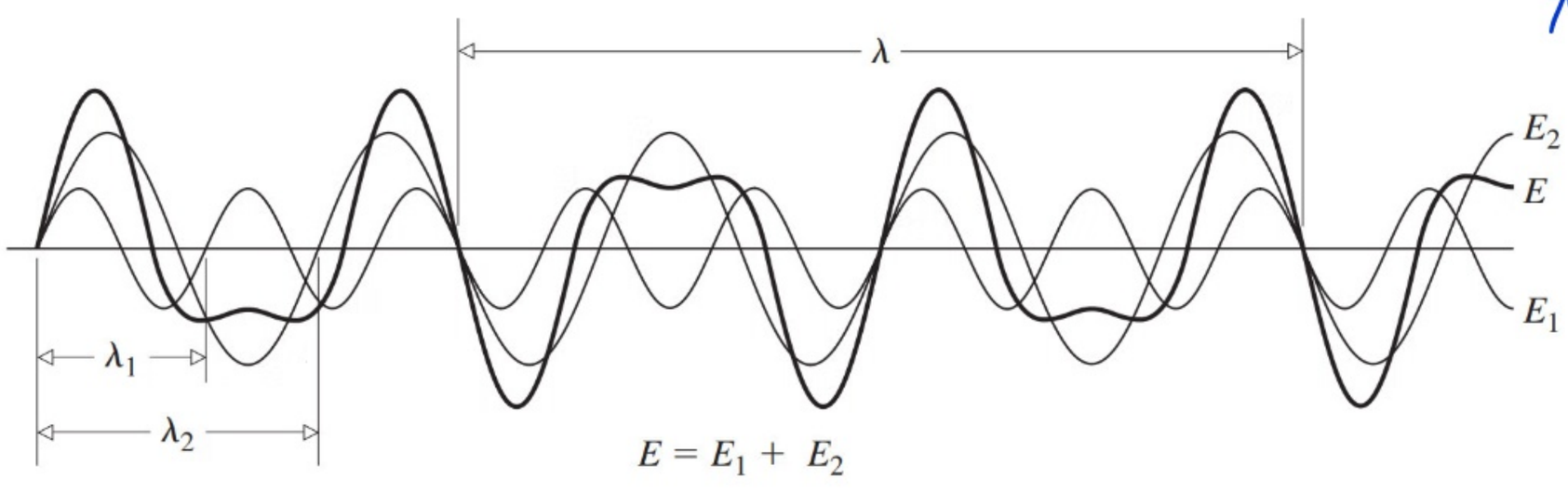
Fourier Series, Coherence

1.) Anharmonic waves

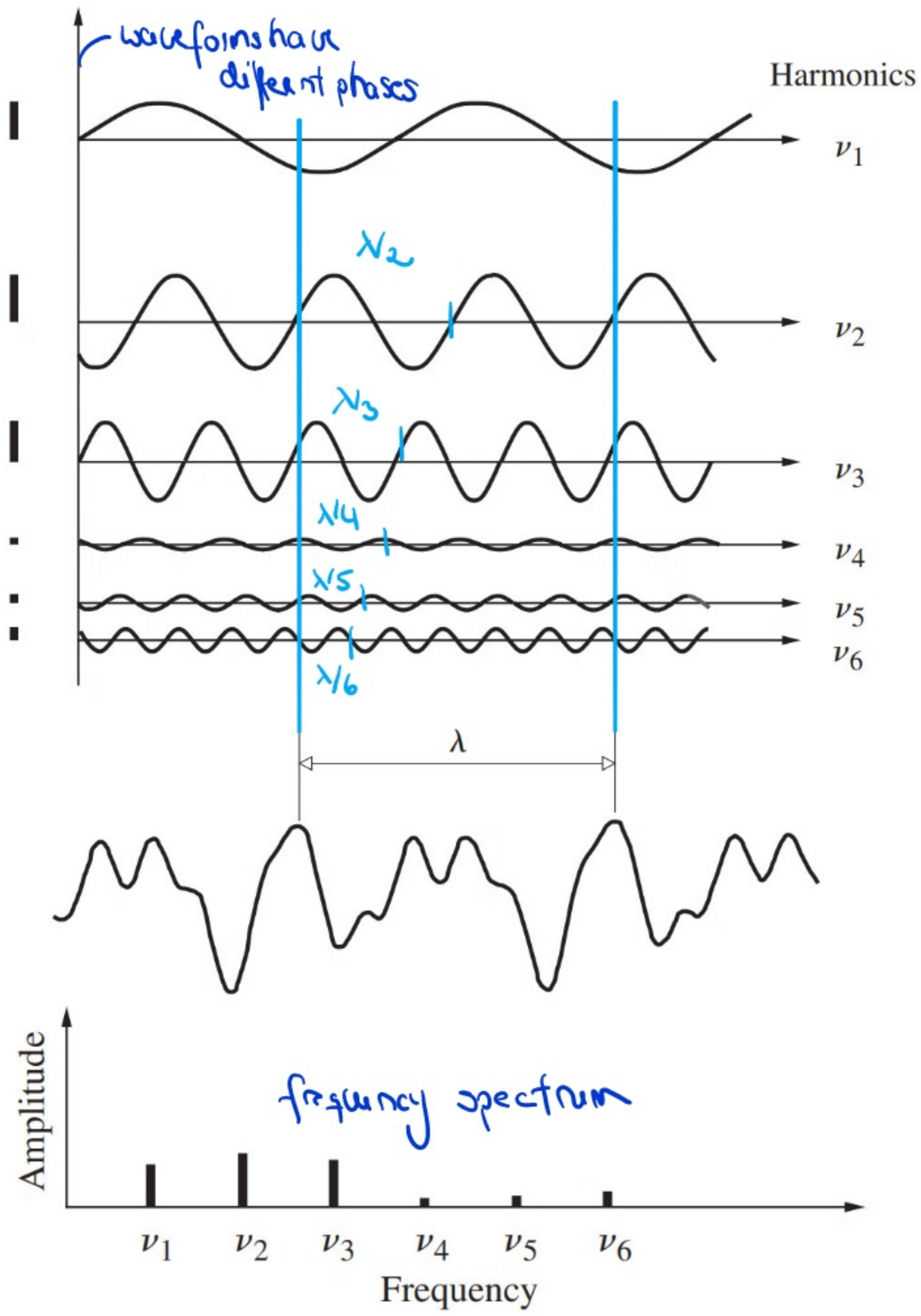
We concluded the last lecture with an analysis of the disturbance resulting from the superposition of two waves of different frequency. The addition of many waves of different frequencies forms the basis of Fourier analyses, a concept that appears everywhere in theoretical physics. In the following, we will discuss how any real wave can be constructed out of appropriately selected harmonic waves both in the time and the spatial domain.

The shape of a wave in space or signal in time is usually referred to as a waveform. Adding such waveforms of different frequencies will result in a composite that is typically no longer sinusoidal, i.e. it will be anharmonic. As illustrated below, however, the resulting pattern will eventually repeat itself and thus be periodic with a spatial period λ.

often many cycles are required before the pattern becomes visible



In the special case, that the wavelengths of subsequent waveforms fit a whole number of times into the longest wavelength λ, the periodicity of the resultant will also be λ, which is thus also referred to as the **fundamental wavelength**. To illustrate how much of each frequency needs to be added to arrive at the final resultant it is convenient to use a **frequency spectrum**, as shown in the bottom plot on the right.

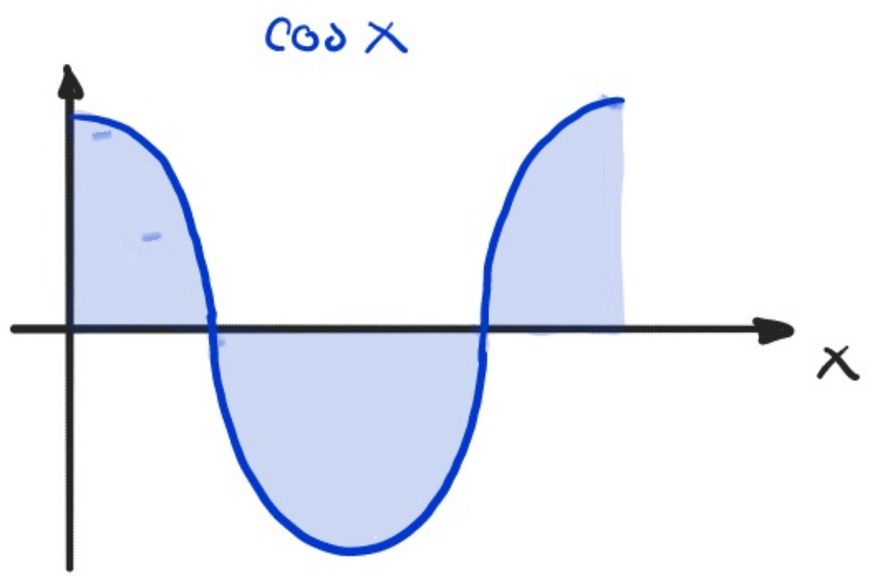


This idea is summarised in **Fourier's theorem** which states that any function $f(x)$ that has a spatial period λ can be synthesised by a **sum of harmonic functions** whose wavelengths are integral submultiples of λ (i.e. $\lambda/2$, $\lambda/3$, etc). Mathematically, we can describe a periodic function $f(x) = f(x+\lambda)$, where $\lambda = 2\pi/k$, in the following form

$$f(x) = \frac{A_0}{2} + \sum_{m=1}^{\infty} A_m \cos(kmx) + \sum_{m=1}^{\infty} B_m \sin(mkx). \qquad (*)$$

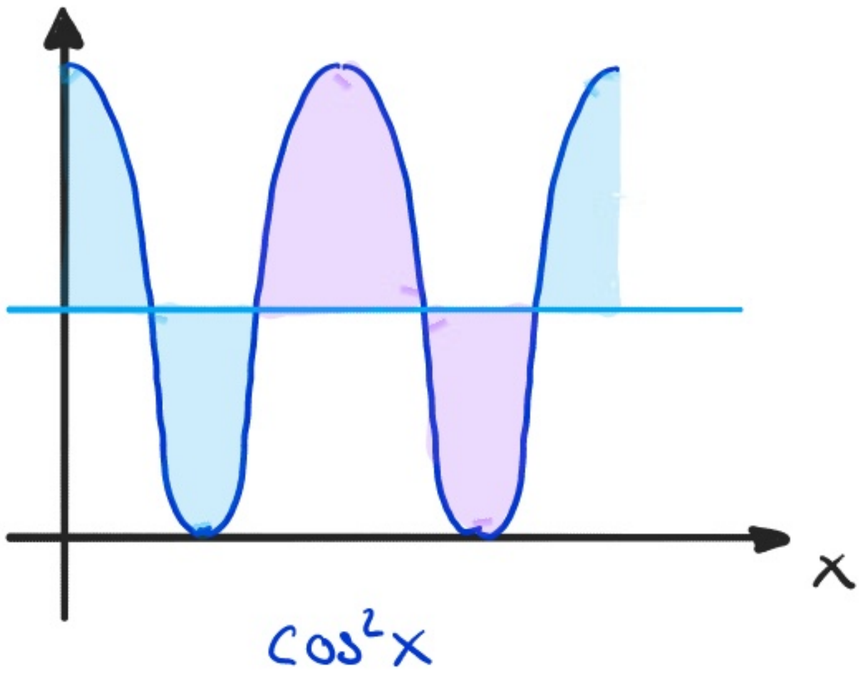
What remains to be determined is the value of each of the constant coefficients A_0, A_m and B_n. To do so, note that the area under the $f(x)$ curve over a distance of, e.g. L, must be equal to the sum of all areas under separate plots of each of the terms on the right-hand side over the same distance L. If $L = \lambda$, then the positive contributions above and the negative contributions below the horizontal axis exactly cancel for all trigonometric contributions and only the constant factor $A_0/2$ remains. Assuming that we have an expression for $f(x)$, we can thus write

$$A_0 = \frac{2}{\lambda} \int_0^\lambda f(x)\,dx \,.$$



The other components can be extracted in a similar manner by accounting for the geometry / orthogonality of the trigonometric functions. Multiplying both sides of Eqn. (*) with $\cos(nkx)$ and integrating from $x = 0$ to $x = \lambda$, we find

$$\int_0^\lambda \cos(nkx) f(x)\,dx = \int_0^\lambda \frac{A_0}{2} \cos(nkx)\,dx + \sum_{m=1}^{\infty} A_m \cdot \underbrace{\int_0^\lambda \cos(nkx)\cos(mkx)\,dx}_{= \lambda/2 \cdot \delta_{nm}} + \sum_{m=1}^{\infty} B_m \underbrace{\int_0^\lambda \cos(nkx)\sin(mkx)\,dx}_{=0} .$$



We thus find

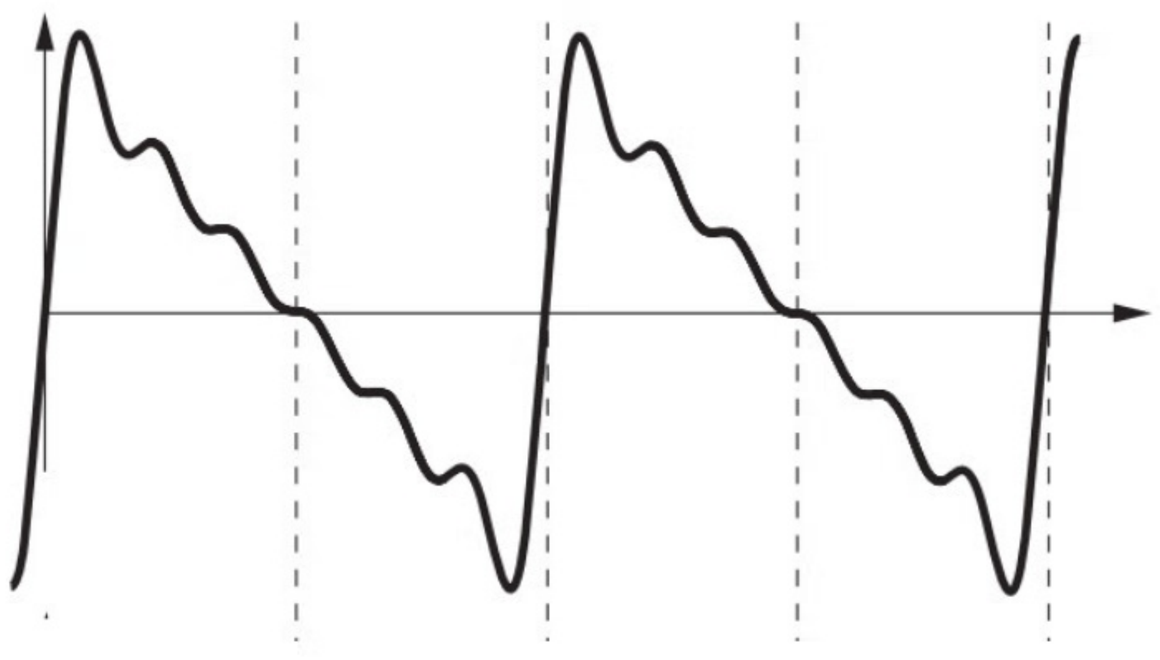
$$A_m = \frac{2}{\lambda} \int_0^{\lambda} f(x) \cos(mkx)\, dx.$$

Similarly, multiplying Eqn. (*) with $\sin(nkx)$ and integrating from $x = 0$ to $x = \lambda$, we obtain a relationship for the coefficients B_m:

$$B_m = \frac{2}{\lambda} \int_0^{\lambda} f(x) \sin(mkx)\, dx.$$

Note that it is very useful to recognise symmetries in your waveform $f(x)$. If for example $f(x)$ is an even function with $f(-x) = f(x)$, the Fourier series will only contain cosine terms, so $B_m = 0 \; \forall m$. Similarly, if $f(x)$ is an odd function with $f(-x) = -f(x)$ than the series expansion will only contain sine terms and $A_m = 0 \; \forall m$. An example is the serrated 'saw tooth' function, which is odd and only requires sinusoids.

representation for six harmonics



adding in terms with higher and higher frequency and smaller amplitudes will result in smoother and smoother functions

While the previous discussion was concerned with stationary waves, we can readily expand the formalism to periodic travelling waves $f(x \pm vt)$:

$$f(x \pm vt) = \frac{A_0}{2} + \sum_{m=1}^{\infty} A_m \cos(mk[x \pm vt]) + \sum_{m=1}^{\infty} B_m \sin(mk[x \pm vt]),$$

---

where the coefficients are equivalent to those given above.

As discussed in previous lectures, it is often more convenient to work in complex notation. Expanding out the sine and cosine in Eqn. (*) (or similarly in the travelling wave case by substituting $x \to x \pm vt$), we find

$$f(x) = \frac{A_0}{2} + \sum_{m=1}^{\infty} A_m \left(\frac{1}{2} e^{imkx} + \frac{1}{2} e^{-imkx} \right) + \sum_{m=1}^{\infty} B_m \left(\frac{1}{2i} e^{imkx} - \frac{1}{2i} e^{-imkx} \right)$$

$$= \frac{A_0}{2} + \sum_{m=1}^{\infty} \left(\frac{1}{2}(A_m - iB_m) e^{imkx} + \frac{1}{2}(A_m + iB_m) e^{-imkx} \right).$$

We can rewrite this as

$$f(x) = \sum_{m=-\infty}^{\infty} c_m e^{-imkx} \quad (**)$$

(sign is again a convention)

with $c_0 = A_0/2$, $c_m = \begin{cases} 1/2(A_m + iB_m) & \text{for } m > 0, \\ 1/2(A_m - iB_m) & \text{for } m < 0. \end{cases}$

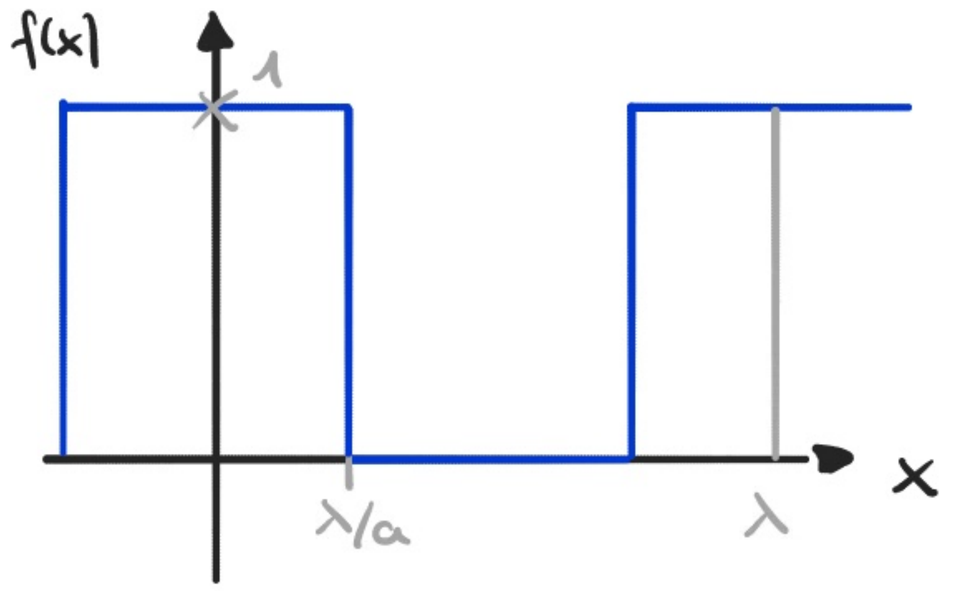
We can equivalently obtain the coefficients by multiplying Eqn. (**) with e^{inkx} and integrating from $x=0$ to $x=\lambda$:

$$c_n = \frac{1}{\lambda}\int_0^{\lambda} f(x)\, e^{inkx}\, dx \; .$$

Note that the limits of integration could also be $-\lambda/2$ and $+\lambda/2$ or any other interval in x that spans one wavelength.

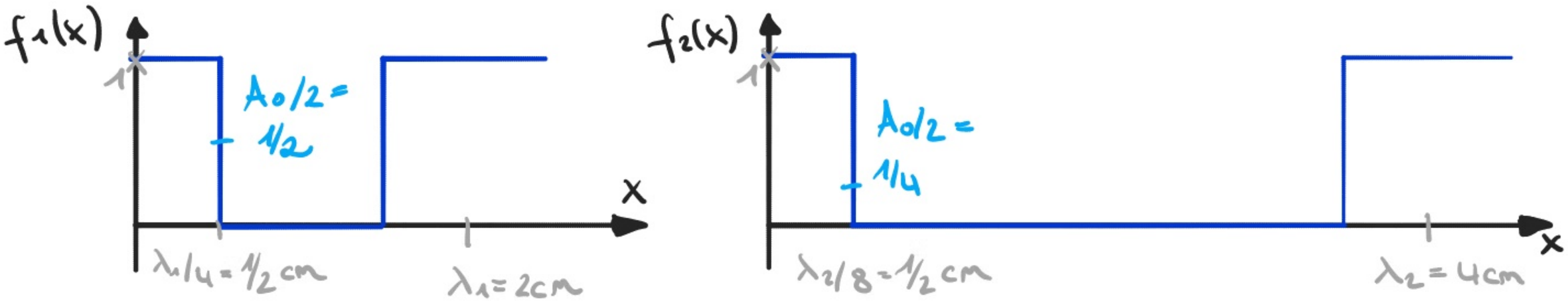
Finally note that as one increases λ, the spacing in frequency of the harmonics in Eqn. (*) or (**) gets smaller and more harmonics are used to approximate $f(x)$. The overall shape in k-space (where $k \equiv mk$) thus stays the same, but becomes more finely spaced. We can illustrate this by looking at the square wave. A rectangular wave pattern is represented by the Fourier series (see figure for the definition of a)

$$f(x) = \frac{2}{a} + \sum_{m=1}^{\infty} \frac{4}{a} \cdot \operatorname{sinc}\left(\frac{2\pi m}{a}\right) \cos(mkx).$$



The width of the rectangular peak is $2\lambda/a$ and can be any fraction of the total wavelength λ. For a square wave we require $2\lambda/a = 1$.

To investigate the behaviour of the Fourier series for different λ, we choose $a_1 = 4$ with $\lambda_1 = 2cm$ and $a_2 = 8$ with $\lambda_2 = 4cm$

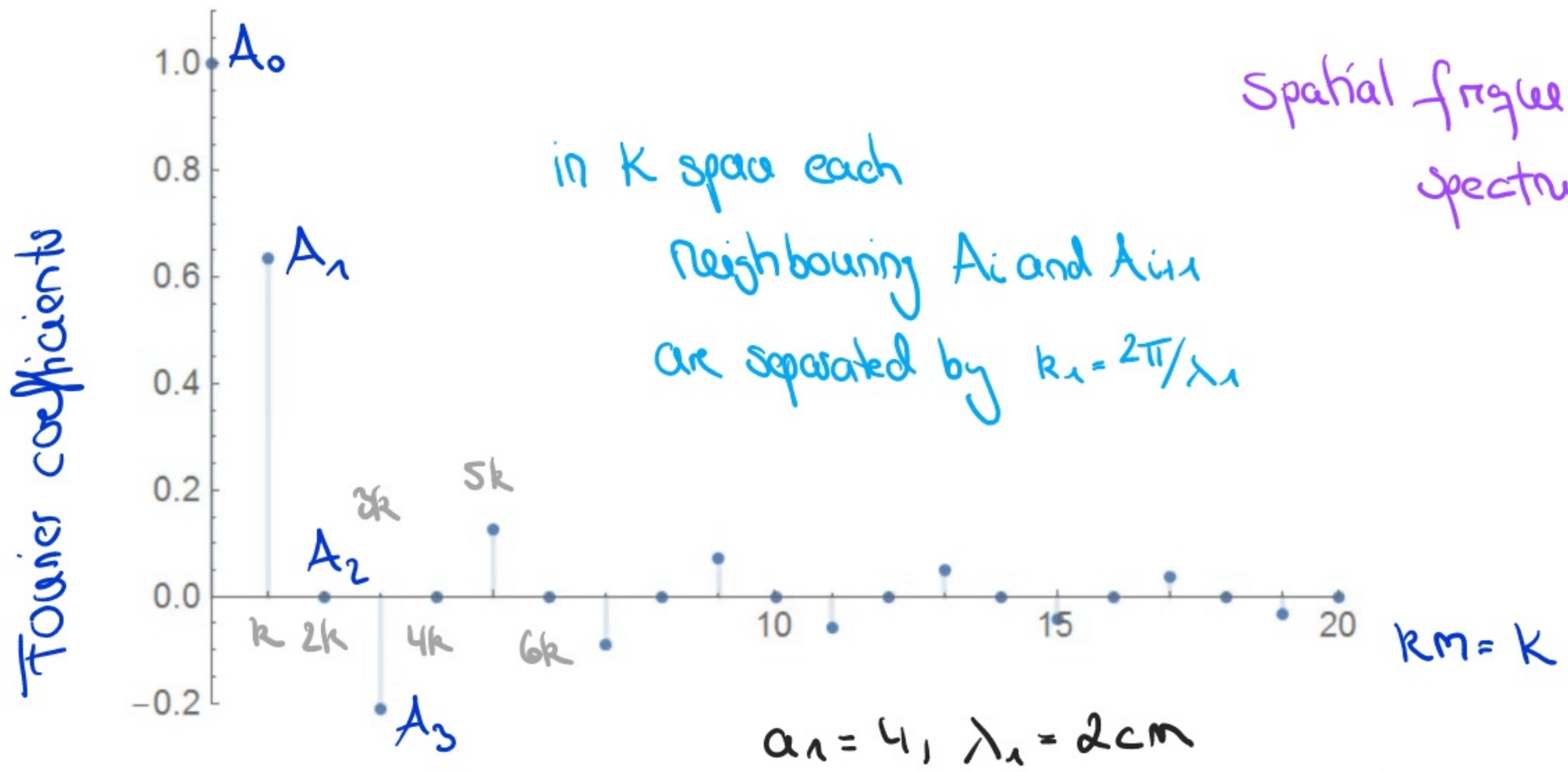


with

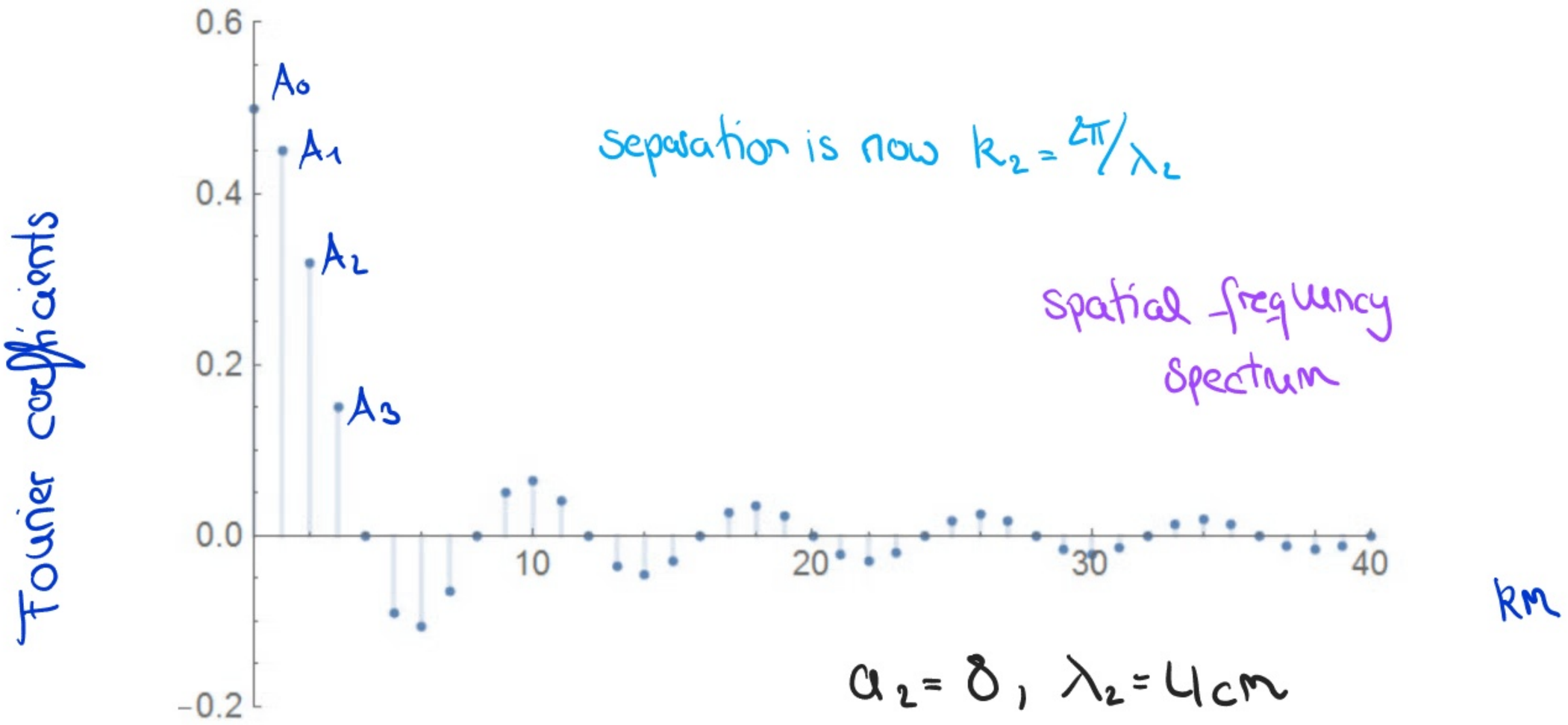
$$f_1(x) = 1/2 + \sum_{m=1}^{\infty} \mathrm{sinc}(m\pi/2)\cos(mk_1x),$$

$$f_2(x) = 1/4 + \sum_{m=1}^{\infty} \frac{1}{2}\mathrm{sinc}(m\pi/4)\cos(mk_2x).$$

We can calculate the coefficients for $f_1(x)$ and plot them as a function of $K (= mk)$; i.e. for $K=0$: $A_0 = 1$, $K=k$: $A_1 = \mathrm{sinc}(\pi/2) = 2/\pi$, $K=2k$: $A_2 = \mathrm{sinc}(\pi) = 0$, $K=3k$: $A_3 = -2\pi/3$, ...



Similarly for the second set of parameters, we obtain

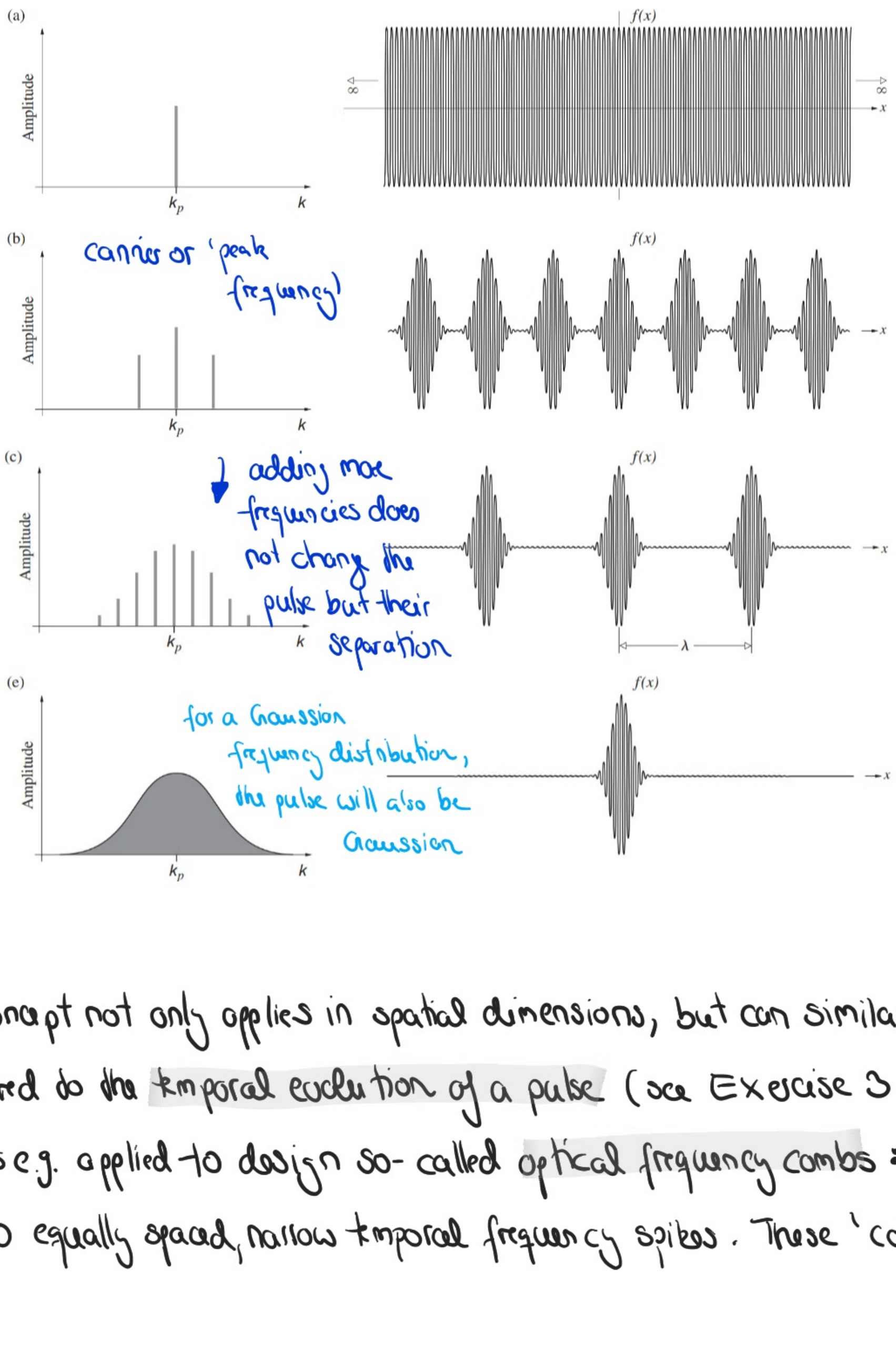


We observe that the overall shape is retained but traced out by more Ai as a result of $\lambda_2 > \lambda_1$ so $k_2 < k_1$. In the limit where $k \to 0$, the discrete sum approaches an integral; i.e. the spectral lines are infinitely close.

2.) Non-periodic waves

All real waves are pulses (albeit rather long ones sometimes), so it is crucial that we know how to analyse non-periodic functions (specifically important for Optics and QM). In Lecture 10, we saw how adding two harmonics of different frequency produces beating of the carrier frequency: at some point the two contributions fall out of phase, creating a minimum of the envelop, before coming back in phase again to produce a maximum.

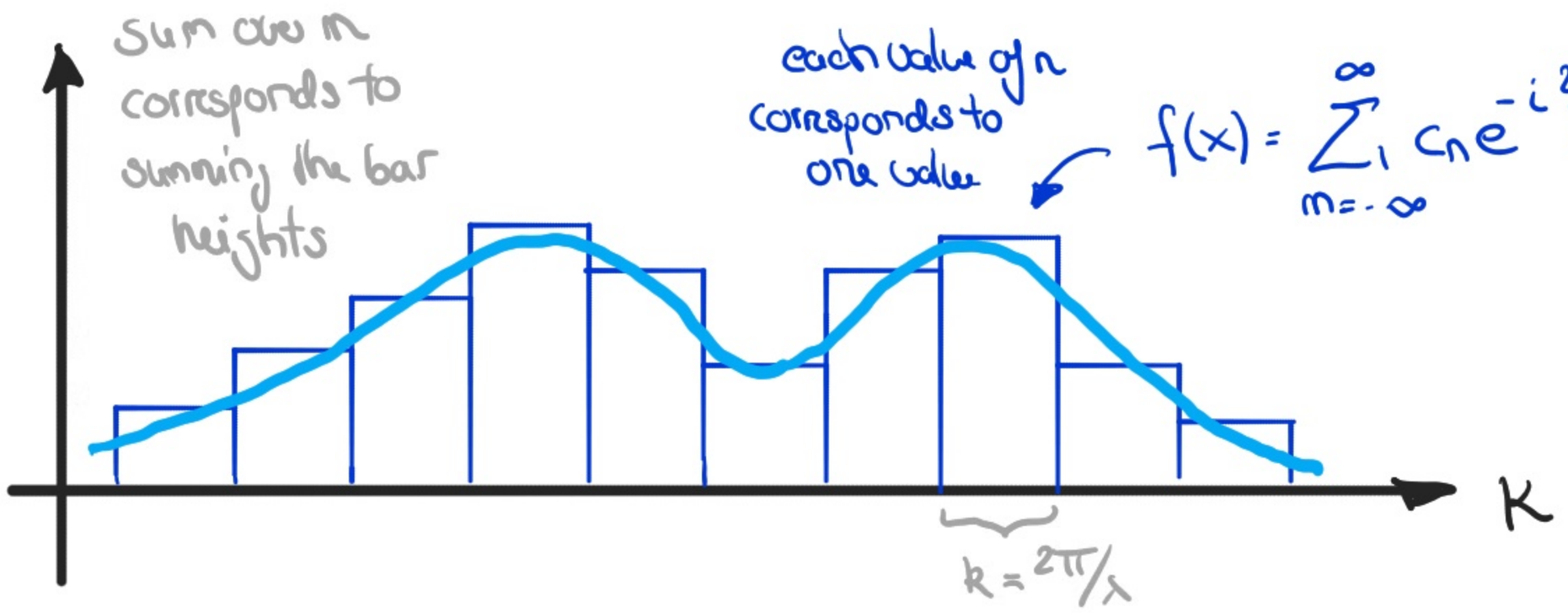
Adding more frequency components would require a greater distance in space before all components are back in phase again; i.e. by adding more frequencies symmetrically around the carrier frequency, individual pulses are separated further. This is illustrated below



This concept not only applies in spatial dimensions, but can similarly be transferred do the temporal evolution of a pulse (see Exercise 3 in PS#3), which is e.g. applied to design so-called optical frequency combs: 10 000 - 100 000 equally spaced, narrow temporal frequency spikes. These 'colored'

teeth of the comb can be used very much like a ruler to measure the frequency of light with extraordinary precision, a concept that was awarded the Nobel Prize of Physics in 2005.

To mathematically describe a single pulse, we can return to the concept of a square wave. Pushing $\lambda \to \infty$, the function $f(x)$ will no longer be periodic and we are left with a single pulse. The limit $\lambda \to \infty$ corresponds to $k \to 0$, which as explained above marks the transition from a Fourier series to a Fourier integral. This is illustrated again below:



As $\lambda \to \infty$, the sum approaches an integral, i.e. since $dk = 2\pi dn/\lambda$

$$\sum_n \to \int_{-\infty}^{\infty} dn = \int_{-\infty}^{\infty} \left(\frac{\lambda}{2\pi}\right) dk \,.$$

k becomes a continuous variable

We thus arrive at the following form for $f(x)$, when replacing $\lambda c_n \to F(k)$ in the limit $\lambda \to \infty$:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(k)\, e^{-ikx}\, dk,$$

- for a single pulse / wavepacket

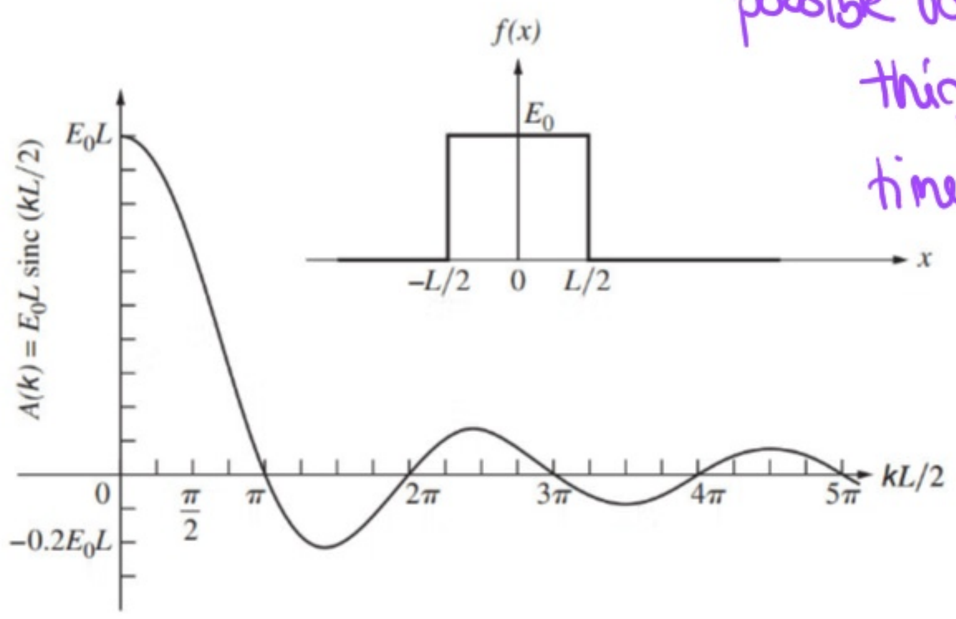
$$F(k) = \int_{-\infty}^{\infty} f(x)\, e^{ikx}\, dx.$$

spectral components of the wavepacket

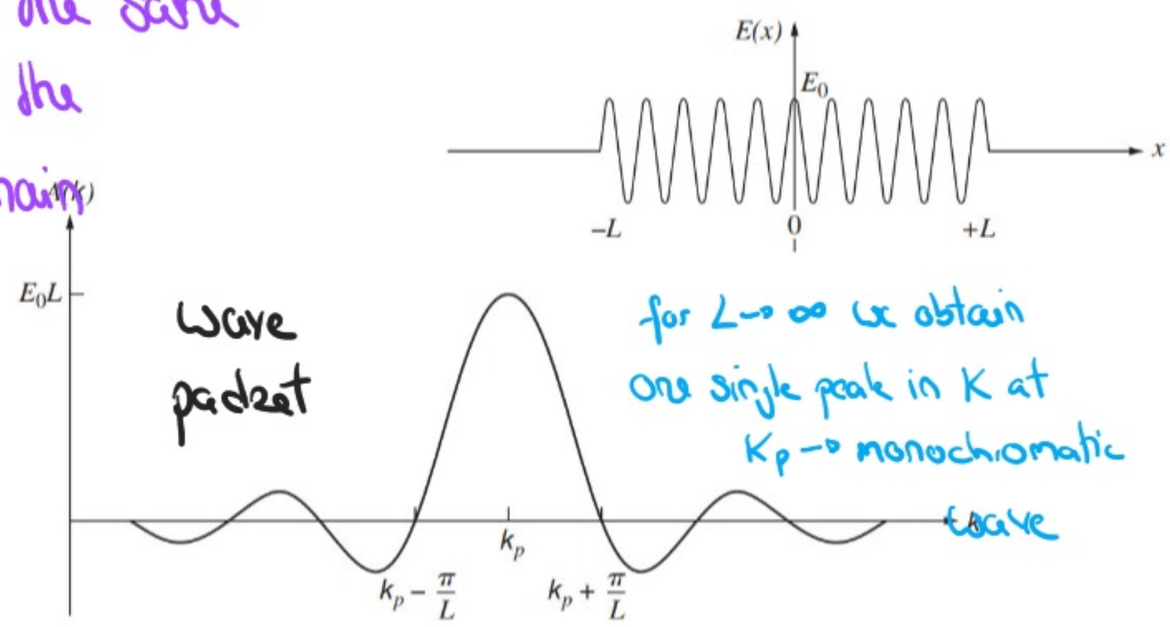
Note that the integrals contain different signs in the exponentials and the integrations are performed over different variables. In principle, we could have also defined the Fourier integral in terms of sines and cosines. One final note of caution: there are many different conventions on where to put the factor $1/2\pi$ and which signs are present in the exponents, so you might see different definitions of Fourier integrals in the literature. Examples for the Fourier-transforms of the square wave and a cosine ('wave-train'):

square pulse & transform (only even contributions)

possible to do the same thing in the time domain



cosine 'wave-train' with carrier frequency k_p (only even contributions)



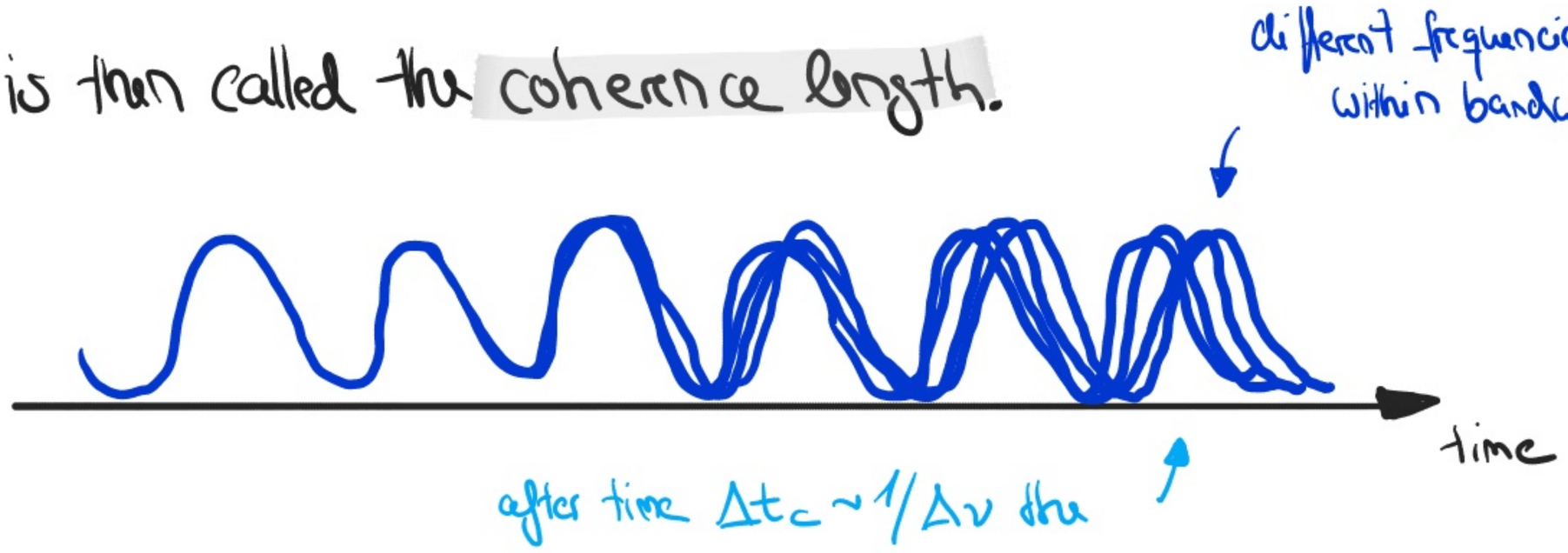
Note that $|A(k)|^2$ is referred to as the power spectrum, provides detailed information about the distribution of energy in each of the spectral components.

Plotting the transform of the cosine-pulse suggests that the wave packet can be characterised by a width $\Delta K = 2\pi/L$, which is also referred to as the frequency bandwidth. As the width of the pulse in x-space is given by $\Delta x = 2L$, we can relate $\Delta x \cdot \Delta K = 4\pi$. Hence, $\Delta x \propto 1/\Delta K$ suggesting that a very short pulse in x-space has a very broad frequency spectrum, and vice versa. Note that the same thing holds for a pulse of width $\Delta t = 2T$ in the time domain, where $\Delta\omega = 2\pi/T$ and thus, $\Delta\omega\,\Delta t = 4\pi$ or equivalently $\Delta\nu \sim 1/\Delta t$.

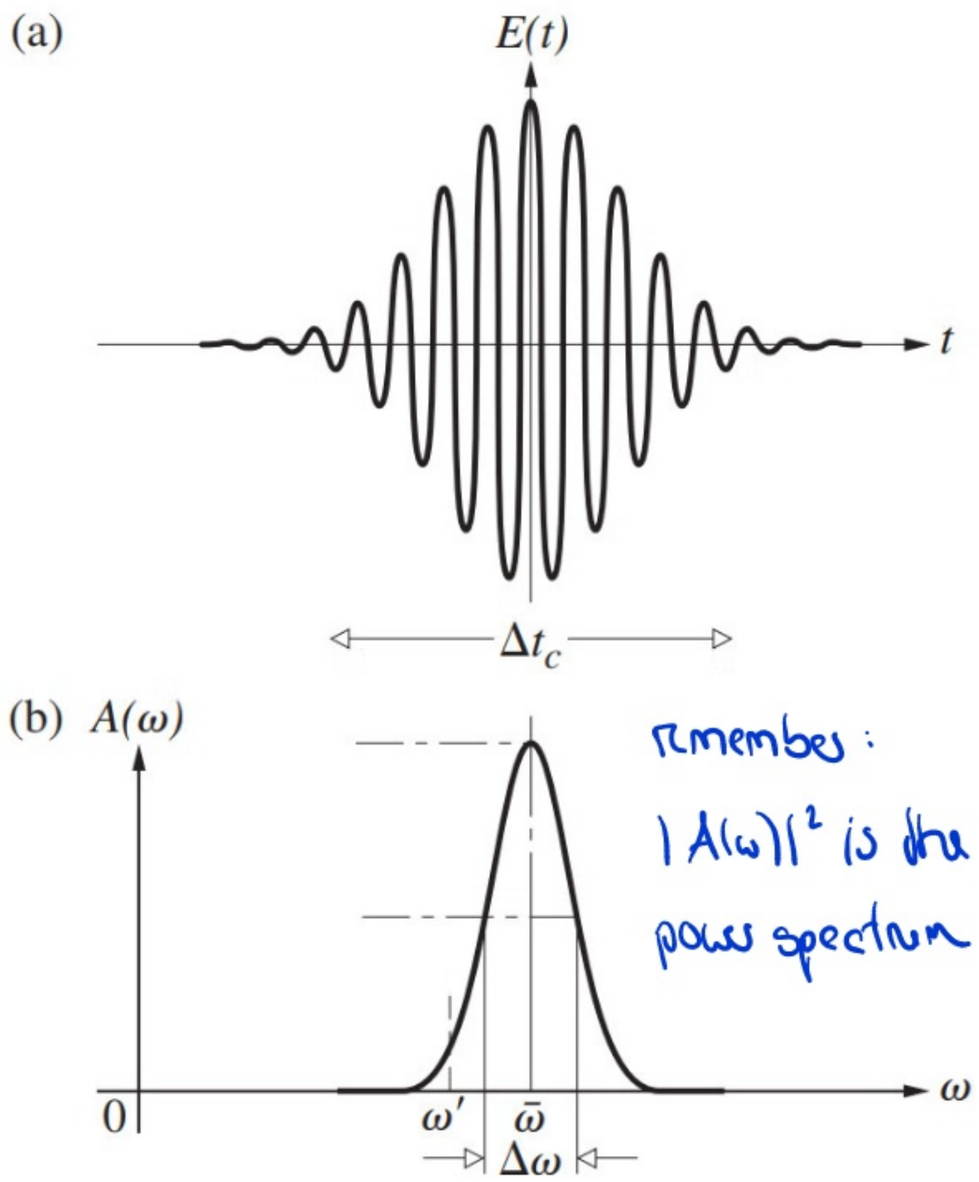
3.) Coherence length

If we consider light emitted by a source that is not a laser, analysing the corresponding spectrum typically reveals the energy to be contained within a narrow frequency band. Such spectral lines are never infinitely thin but have a bandwidth $\Delta\nu_0$ (due to the underlying electron transition having a duration of the order of $\tau \sim$ ns and hence $\Delta\nu_0 \sim 1/\tau$). As the atoms are further in random thermal motion, $\Delta\nu_0$ is typically also broadened by the Doppler effect and collisions between different atoms, resulting in a quasi-monochromatic source with spectral width $\Delta\nu$. After a time $\Delta t_c \sim 1/\Delta\nu$, the so-called coherence time, the phase of the wave will thus become uncertain (see next page). The distance $\Delta l_c = c \cdot \Delta t_c = c/\Delta\nu$ over which the phase is well de-

defined is than called the coherence length.



We would now like to know what the wave-train of such partially coherent light will look like. As observations of the frequency distribution of quasi-monochromatic sources reveal a Gaussian patten, we can deduce that the net amplitude of the electric field is also Gaussian. Suppose that each of the individual photon wavepackets that constitute the final beam are a harmonic function modulated by a Gaussian, then the Fouries transform will be also Gaussian as illustrated below. The relative phases between different photon wavepackets are undefined so there is no defined phase relationship between a specific frequency ω' in different pulses and the resulting shape of the final wave train will differ from that of the individual pulses. Adding these short pulses of width

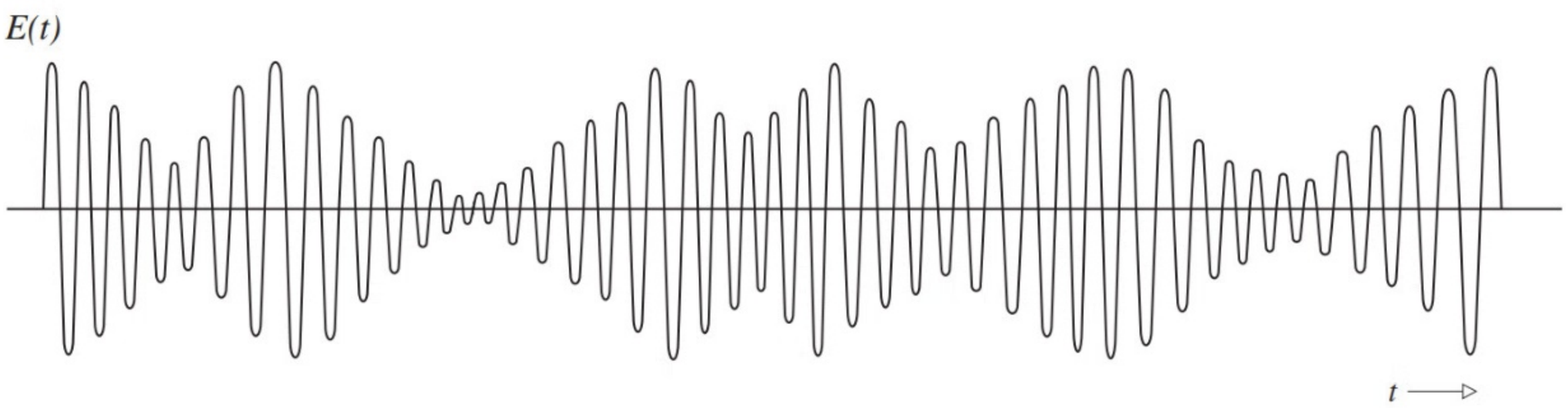


remember:
$|A(\omega)|^2$ is the power spectrum

Δt_c with undefined relative phase and random relative timing will result in a overall wavetrain that is almost sinusoidal but has an uncertainty in frequency. The composit light wave will thus be of the form

$$E(t) = E_0(t) \cos\left[\varepsilon(t) - 2\pi\bar{\nu}t\right],$$

where $\bar{\nu} = \bar{\omega}/2\pi$ denotes the central frequency of the sinusoid / carrier, and $E_0(t)$ and $\varepsilon(t)$ denote the fluctuations in amplitude and frequency respectively. This can be illustrated as



predicting phase only possible within Δt_c

Note that partially coherent light with a short coherence length can nonetheless be very useful in interferometric instruments such as optical coherence tomography.

Phys 434 - Lecture 12

Intro to Polarisation:

So far, we have treated wave superpositions without discussing the vec-
tor nature of the electric field. In effect, we have been dealing with only one component of $\vec{E}$, which is adequate for the case of linear polarisation, where the orientation of the field is constant, while the amplitude and sign can vary. In this case, $\vec{E}$ and $\vec{R}$ lie in a fixed plane, also known as the plane of vibration. We are now going to generalise this concept.

1.) Polarisation states

Consider a superposition of two monochromatic, orthogonal optical disturbances, moving through the same region of space, travelling in the same direction. We can write them as

$$\vec{E}_x(z,t) = E_{ox}\cos(kz - \omega t)\hat{i},$$ ($\hat{i}$: unit vector in x-direction)

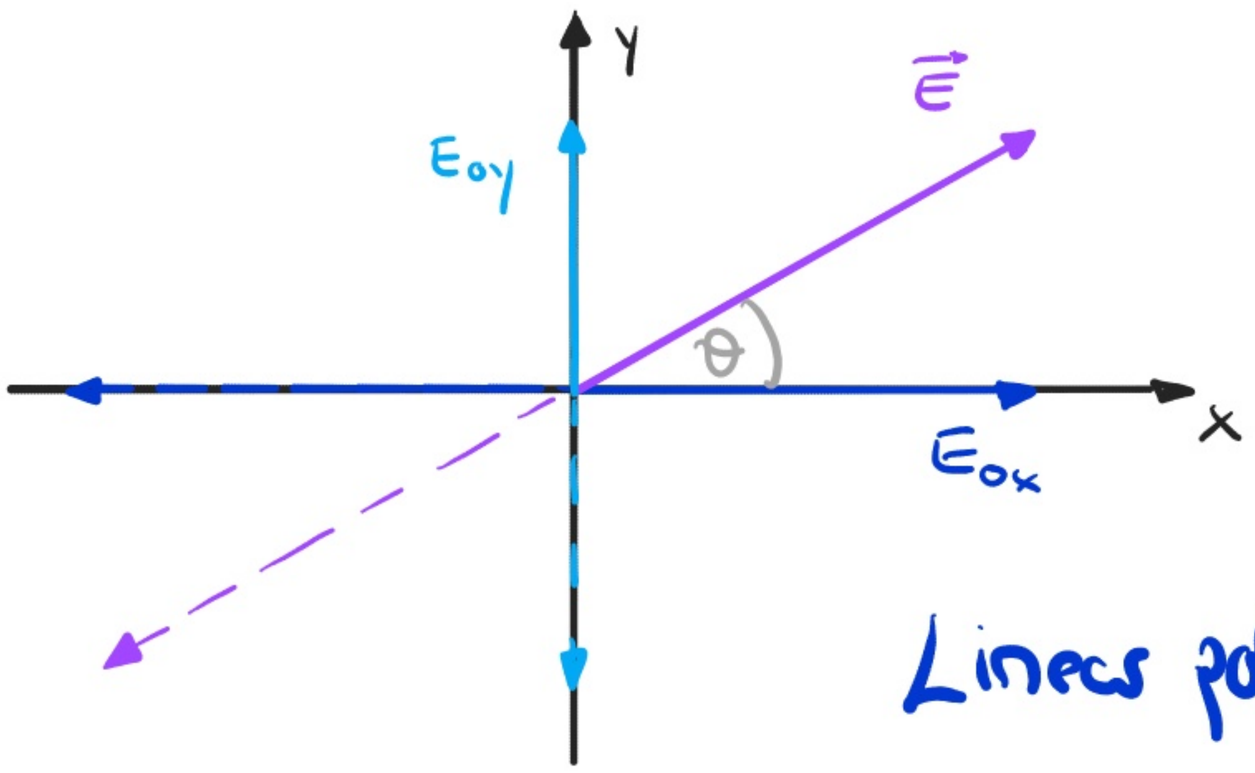
$$\vec{E}_y(z,t) = E_{oy}\cos(kz - \omega t + \varepsilon)\hat{j},$$ ($\hat{j}$: unit vector in y-direction)

where ε is the relative phase between the vectors, i.e. it takes some time ε/ω for both to attain the same phase (E_y lags E_x if $\varepsilon > 0$).

We now want do know what the resultant vector $\vec{E} = \vec{E}_x + \vec{E}_y$ looks like. To determine the behaviour, we focus on a fixed point $z = z_0$ and investigate what happens for different ε.

(i) $\varepsilon = 0, 2\pi$: $\vec{E} = (E_{0x}\hat{i} + E_{0y}\hat{j})\cos(kz_0 - \omega t)$

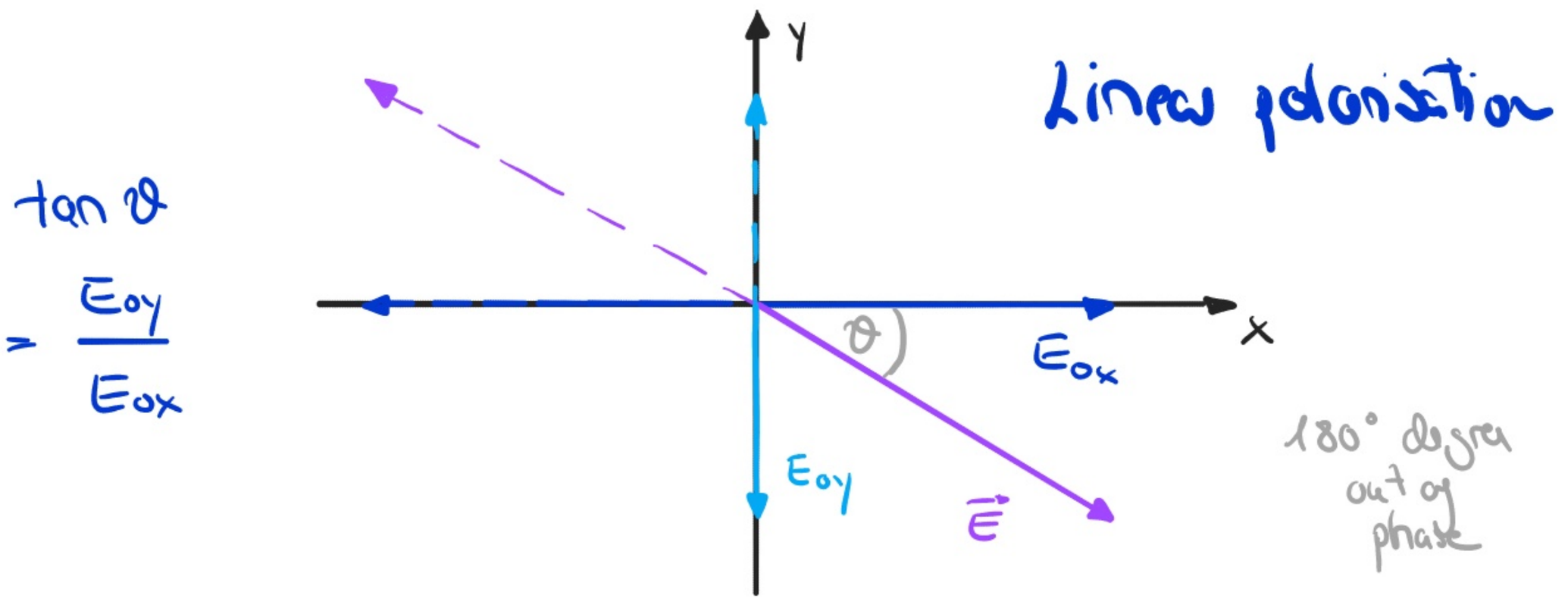
the angle θ depends on the amplitudes E_{ox} and E_{oy}



in phase

Linear polarisation

(ii) $\varepsilon = \pi, -\pi$: $\vec{E} = (E_{0x}\hat{i} - E_{oy}\hat{j})\cos(kz_0 - \omega t)$

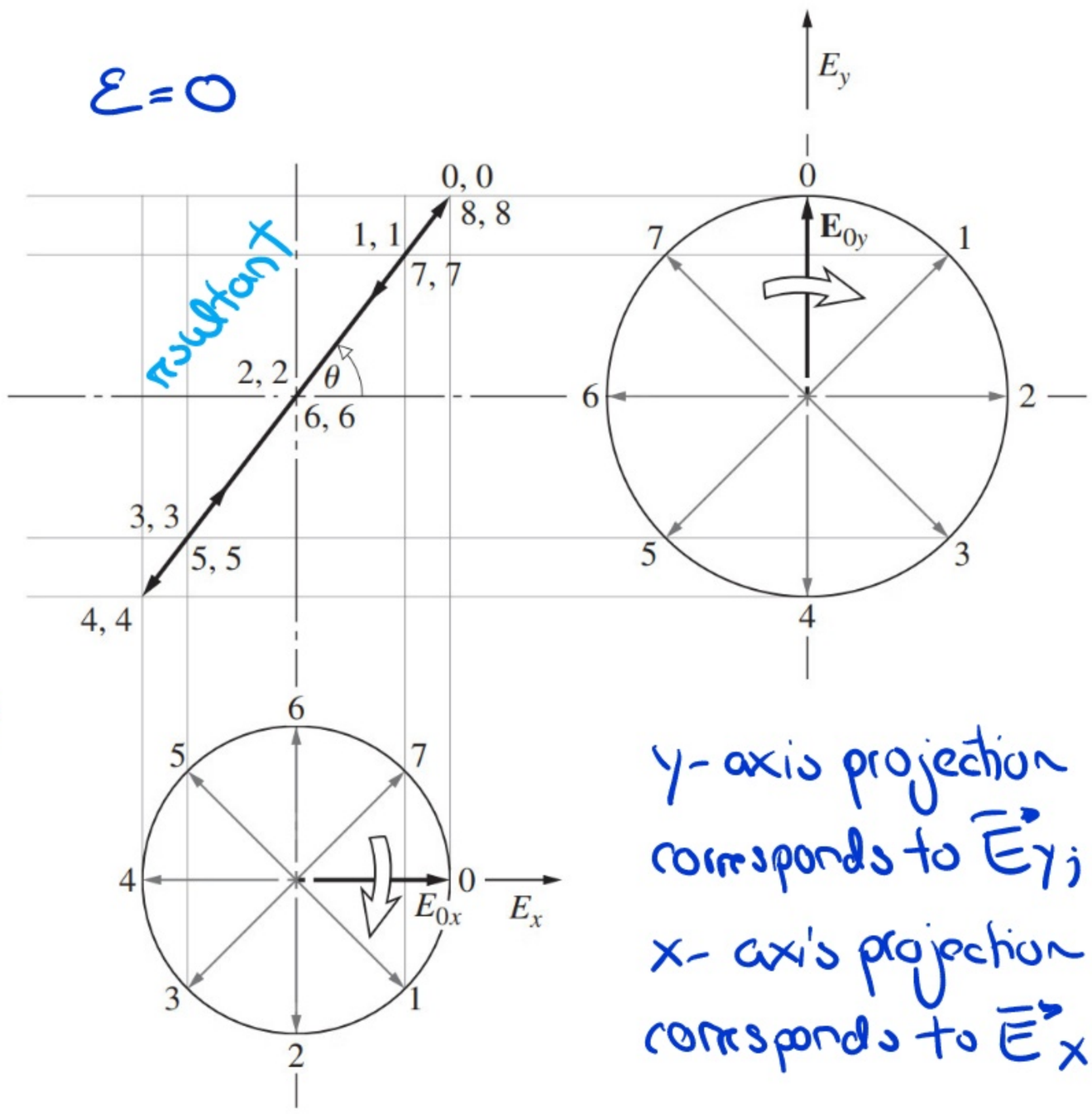


Linear polarisation

180° degree out of phase

In both cases, the resultant light field has still one direction, but the plane of vibration has rotated. The summation can be very easily

illustrated using phasor notation. The in-phase case can be illustrated as



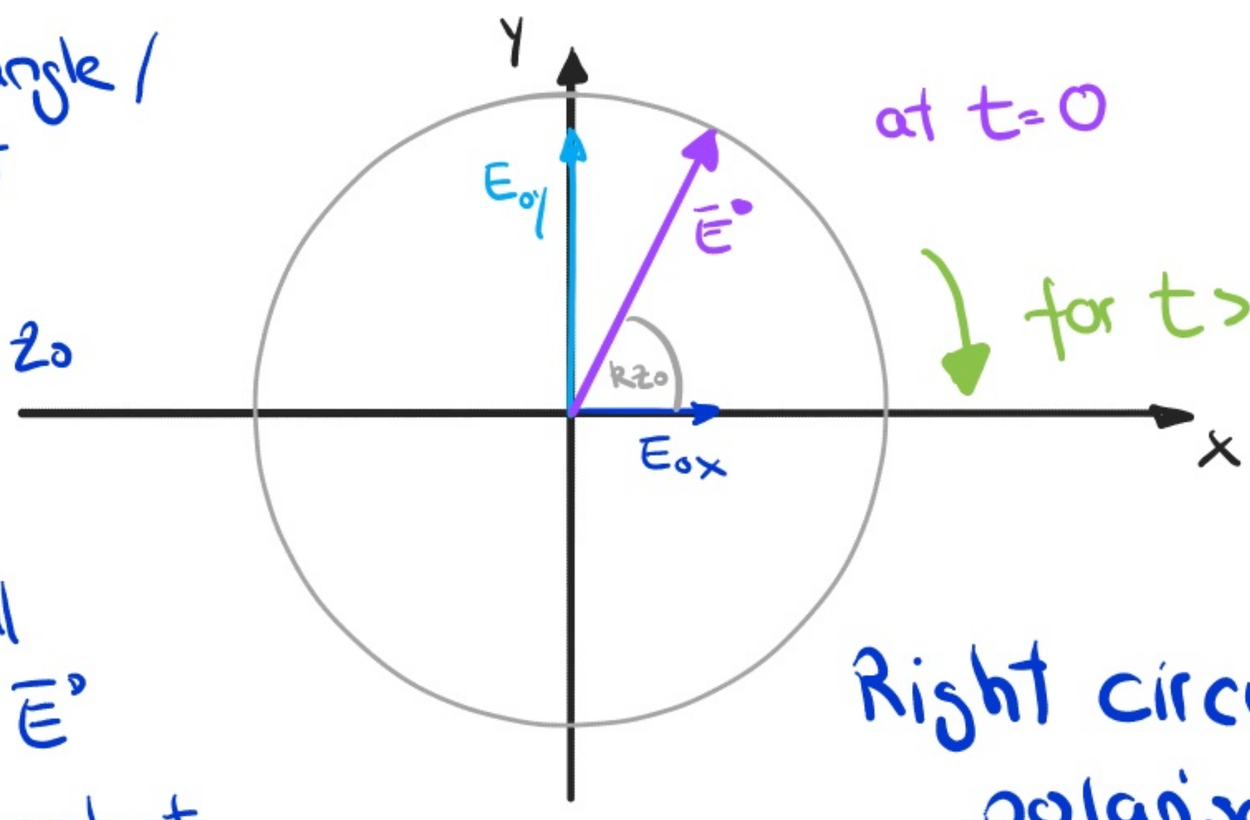
both phasors are rotating at the same velocity

y-axis projection corresponds to $\vec{E}_y$;
x-axis's projection corresponds to $\vec{E}_x$

(iii) $\varepsilon = -\pi/2$: $\vec{E}_y = E_{oy} \sin(kz_o - \omega t)\hat{j}$

if $E_o \equiv E_{ox} = E_{oy}$, $\vec{E} = E_o[\cos(kz_o - \omega t)\hat{i} + \sin(kz_o - \omega t)\hat{j}]$

the initial angle / phase off set depends on the position z_o



at t=0

for t>0

for a fixed position, as time progresses $\vec{E}$ will rotate in a clockwise circle as seen from an observer looking back at the source

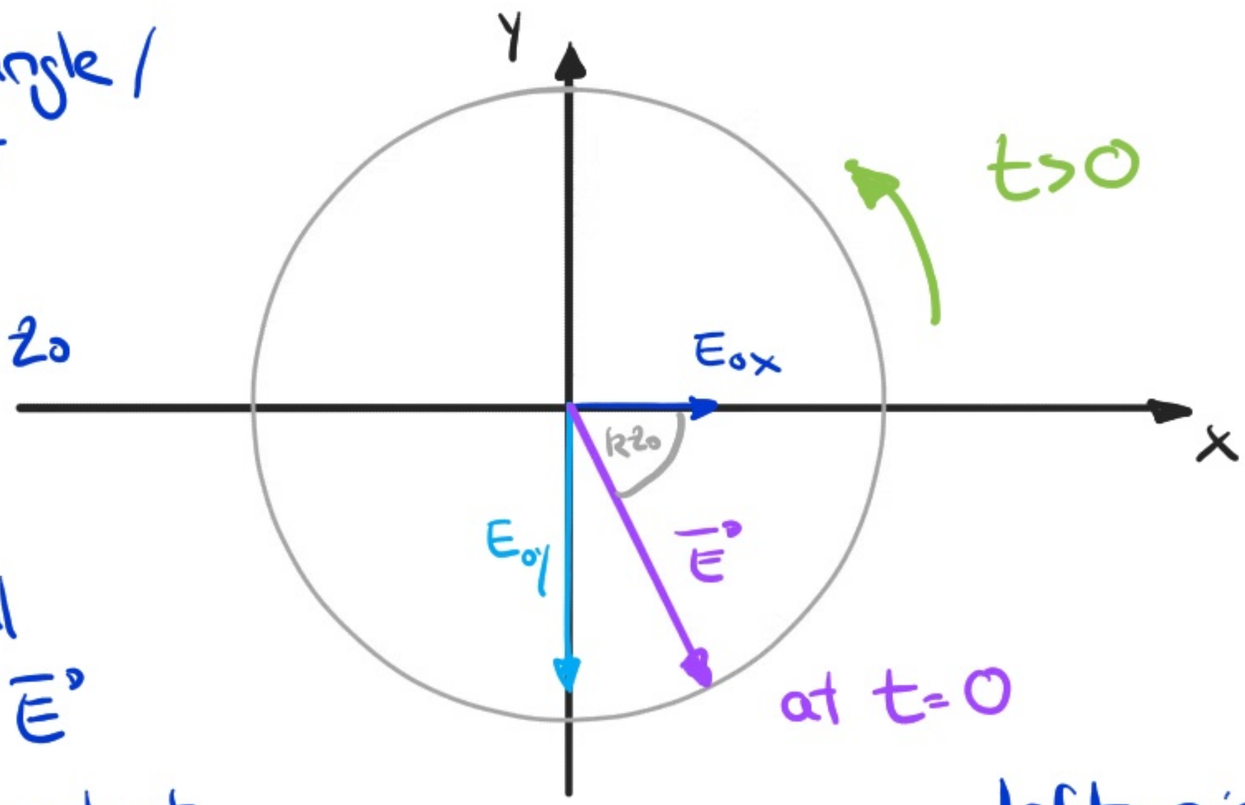
the overall length of $\vec{E}$ will be constant

Right circular polarisation

(iv) $\varepsilon = +\pi/2$: $\vec{E}_y = -E_{oy}\sin(kz_o - \omega t)\hat{j}$

if $E_o \equiv E_{ox} = E_{oy}$, $\vec{E} = E_o[\cos(kz_o - \omega t)\hat{i} - \sin(kz_o - \omega t)\hat{j}]$

the initial angle / phase off set depends on the position z_o



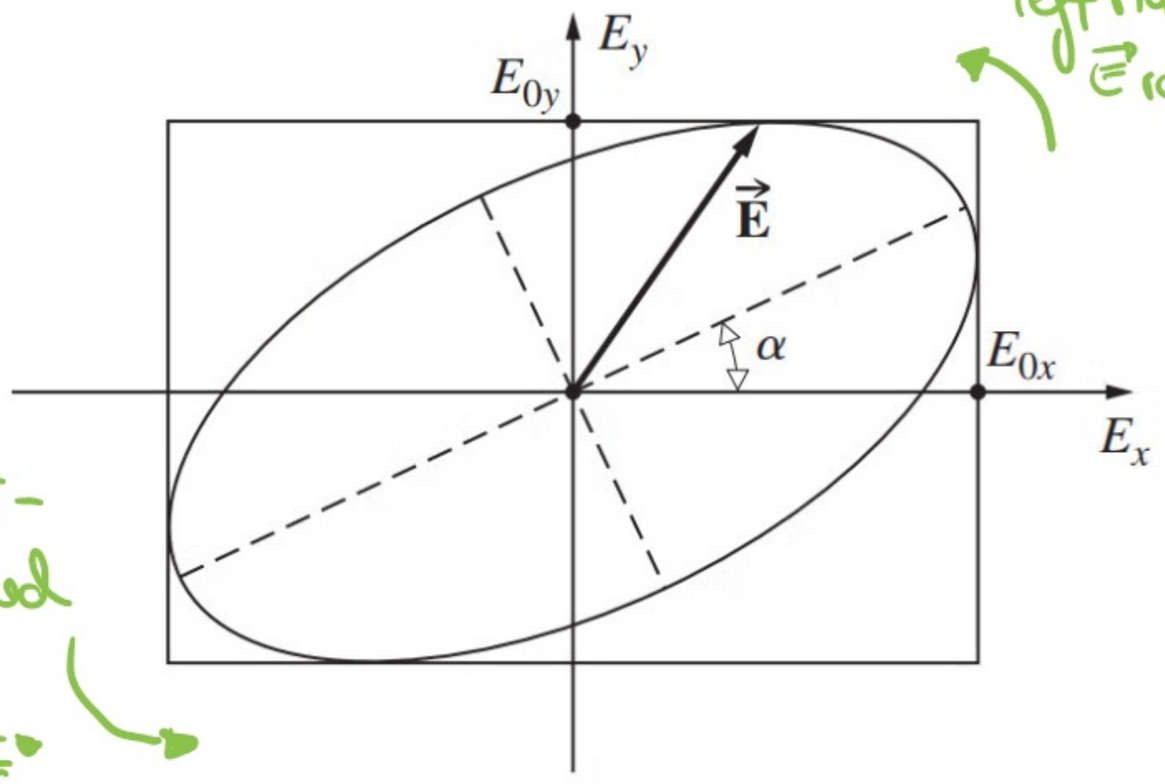
for a fixed position, as time progresses $\vec{E}$ will rotate in a counterclockwise circle as seen from an observer looking back at the source

the overall length of $\vec{E}$ will be constant

left circular polarisation

A note of caution: The handedness of the $\vec{E}$ field is sometimes defined as seen from the source, not the observer as discussed above. In this case, left and right circular polarisation would be exactly inverted.

As far as the mathematical description is concerned, all four cases (i) - (iv) can be seen as special cases of elliptically polarised light. In this most general case, $E_{oy} \neq E_{ox}$ and ε can take any value, resulting in an electric field $\vec{E}$ that traces out an elliptical path as a function of time (for a fixed position z_o). With a bit of algebra it is possible to combine the equations for $\vec{E}_x$ and $\vec{E}_y$ to obtain the ellipse parameters, e.g. the angle α, the ellipse is making with the $\hat{i}\hat{j}$-coordinate system:



left-handed $\vec{E}$ rotation

right-handed $\vec{E}$ rotation

$$\tan 2\alpha = \frac{2E_{0x}E_{0y}}{E_{0x}^2 - E_{0y}^2} \cos\varepsilon .$$

by changing ε, we can rotate the ellipse

2.) Jones vector

The above discussion clearly indicates that a circular polarisation state can be achieved by superposition of two perpendicular linear polarisations. Conversely, a linearly polarised state can be obtained by superposing two left and right circularly polarised waves. These relationships can be efficiently represented via a vector representation of the polarisation state.

If we write $\vec{E}^\circ = \mathrm{Re}\left[E_{ox}\, e^{i(kz-\omega t+\varphi_x)}\hat{\imath} + E_{oy}\, e^{i(kz-\omega t+\varphi_y)}\hat{\jmath}\right]$, with the respective phases φ_x and φ_y, we can define the polarisation state by a normalised vector, the so-called Jones vector

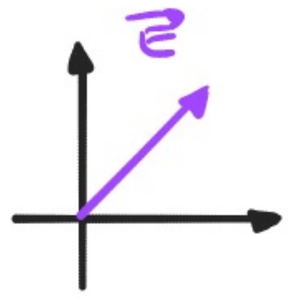
$$\vec{E}_J^\circ = \left(E_{ox}^2 + E_{oy}^2\right)^{-1/2} \begin{pmatrix} E_{ox}\, e^{i\varphi_x} \\ E_{oy}\, e^{i\varphi_y} \end{pmatrix} .$$

the normalisation prefactor doesn't alter the polarisation state

the overall (global) phase of $\vec{E}^\circ$ itself has no meaning

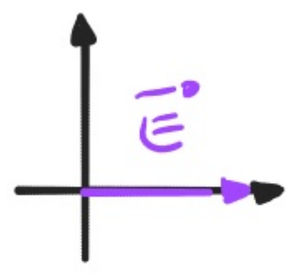
We can for example calculate

→ 45° linear polarisation
$E_{0x} = E_{0y}$ and e.g. $\varphi_x = \varphi_y = 0$



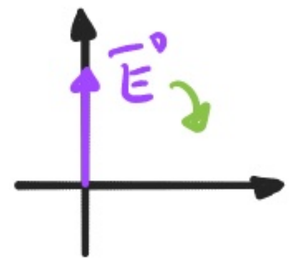
$$\vec{E}_J = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

→ horizontal linear polarisation
$E_{0y} = 0$ and e.g. $\varphi_x = 0$



$$\vec{E}_{J,H} \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

⇒ right circular polarisation
$E_{0x} = E_{0y}$ and e.g. $\varphi_x = 0$
$\varphi_y = -\pi/2$



$$\vec{E}_{J,R} \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

Using this notation it is straight forward to calculate superpositions of different polarised states and deduce the resulting polarisation. Note that $\vec{E}_{J,H}$ together with $\vec{E}_{J,V} \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ as well as $\vec{E}_{J,R}$ and $\vec{E}_{J,L} \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}$ form an orthonormal basis for all polarisation states, because their Jones vectors are orthogonal.

3.) Polarisation and angular momentum

In previous lectures, we have seen that an EM wave impinging on a medium can transfer energy and thus also linear momentum to that medium. If a circularly polarised beam of light enters a medium, we would similarly expect the rotating electric field to set electrons into circular motion. More precisely, the rotating $\vec{E}$ field will exert a torque τ onto the electrons, which

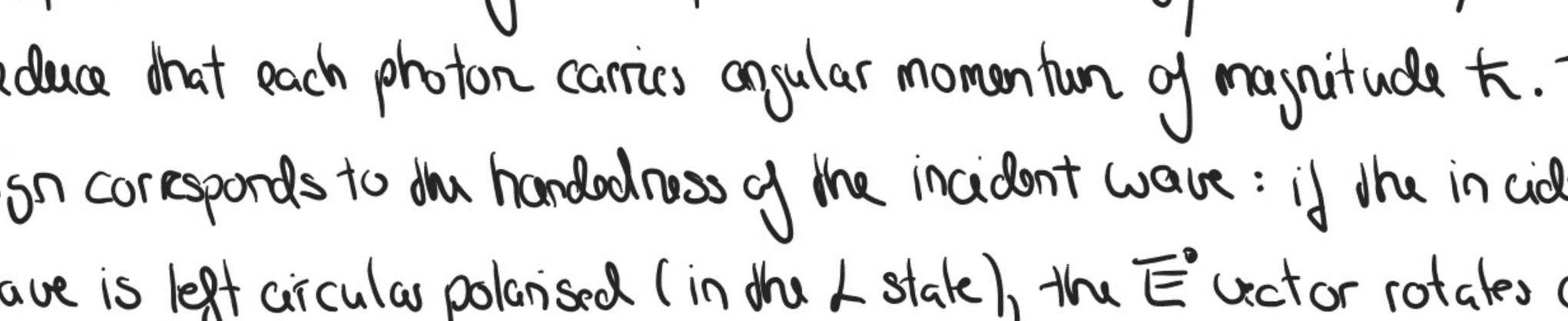
results in the transfer of power (energy per time) to the material, i.e.

$$P = dE/dt = \omega \tau = \omega \, dL/dt,$$

↑ EM wave drives circular motion of some frequency

where we have used the fact that a torque is equal to the rate of change of the angular momentum L over time. This implies that any electron absorbing the (quantity of) energy E from an incident circular wave will also absorb an amount of angular momentum given by $L = E/\omega$.

According to QM, an EM wave transfers energy in quantised packets i.e. photons. As each of those photons has the energy $E = \hbar\omega$, so we can deduce that each photon carries angular momentum of magnitude $\hbar$. The sign corresponds to the handedness of the incident wave: if the incident wave is left circular polarised (in the L state), the $\vec{E}$ vector rotates counter clockwise, in which case we define the angular momentum vector to be parallel to $\vec{k}$ (see below). Similarly for a right polarised wave (the R state), the $\vec{E}$ vector rotates clockwise and thus $\vec{L} \uparrow\downarrow \vec{k}$:



In this picture, a linearly polarised beam is a superposition of $+\hbar$ and $-\hbar$

angular momentum states, with each photon exhibiting either $+\hbar$ or $-\hbar$ once it is angular momentum / spin is measured. As a whole, the linearly polarised beam will impart no angular momentum on the target.

4.) Unpolarised light and stokes parameters

'Unpolarised' (also called 'natural' or 'randomly polarised') light refers to the situation where the polarisation fluctuates randomly on time scales much faster than any measurement time, because of incoherent light fluctuating in frequency. Note that the opposite of this would be a monochromatic plane wave, which can be depicted as an infinite wave train (composed of two orthogonal components) and is thus perfectly polarised. 'Real' light is typically neither completely polarised or completely unpolarised, but has fluctuations in polarisation that are somewhat regular. Such light is typically referred to as partially polarised.

The Jones formalism introduced earlier cannot be used to describe unpolarised or partially polarised light, as the two basis vectors are no longer sufficient to describe the imperfectly polarised state. This is very similar to QM, where vectors can no longer be used to describe mixed quantum states and instead matrices have to be invoked. The equivalent formalism in Optics is referred to as Stokes parameters.

We can write partially polarised light as the superposition of two quasi monochromatic (i.e. partially coherent) light sources. As discussed in Lecture 11, the respective optical fields are

$$\vec{E}_x(t) = E_{ox}(t)\cos\left(\bar{k}z - \bar{\omega}t + \varepsilon_x(t)\right)\hat{i}, \quad \text{fluctuating amplitude and phase}$$

$$\vec{E}_y(t) = E_{oy}(t)\cos\left(\bar{k}z - \bar{\omega}t + \varepsilon_y(t)\right)\hat{j}.$$

The four Stokes parameters are then defined as

$$S_0 = \langle E_{ox}^2 \rangle + \langle E_{oy}^2 \rangle, \quad \leftarrow \text{standard irradiance}$$

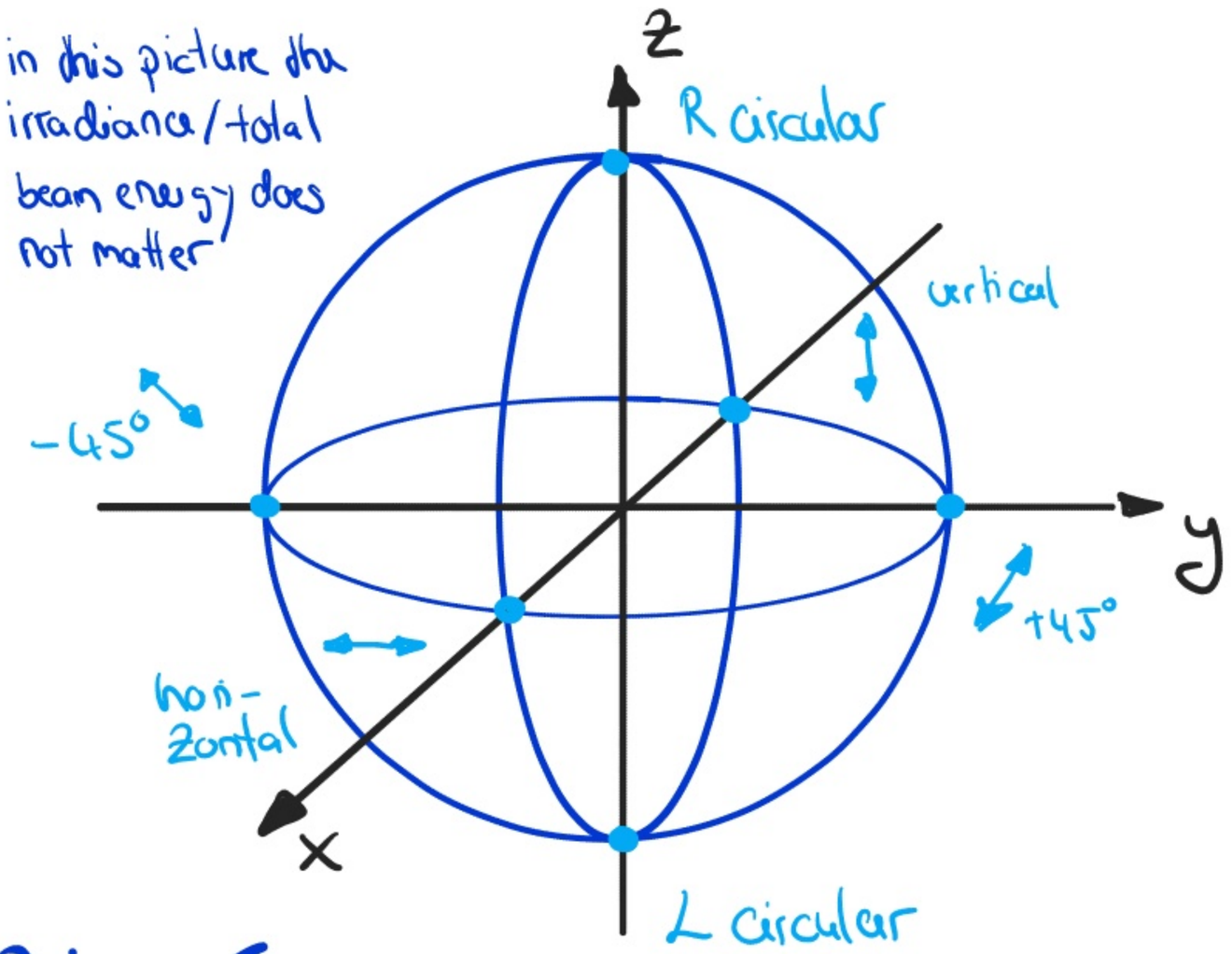
$$S_1 = \langle E_{ox}^2 \rangle - \langle E_{oy}^2 \rangle,$$

$$S_2 = \langle 2 E_{ox} E_{oy} \cos(\varepsilon_y - \varepsilon_x) \rangle,$$

$$S_3 = -\langle 2 E_{ox} E_{oy} \sin(\varepsilon_y - \varepsilon_x) \rangle,$$

where $\langle\ \rangle$ indicates the time average. Commonly, these parameters are normalised by S_0 so that $S_0 \to 1$, $S_1 \to S_1/S_0$, $S_2 \to S_2/S_0$ and $S_3 \to S_3/S_0$. The polarisation state is thus represented as a vector of the form (S_0, S_1, S_2, S_3). Note that for partially polarised light $V \equiv (S_1^2 + S_2^2 + S_3^2)^{1/2}/S_0$ defines the degree of polarisation, whereas unpolarised light is characterised by $\langle E_{ox}^2 \rangle = \langle E_{oy}^2 \rangle$ as neither component averages to zero, so the corresponding Stokes vector is $(1,0,0,0)$. Finally, the above equations show that perfectly polarised light satisfies $S_0 = S_1 + S_2 + S_3$. For those states, the vector (S_1, S_2, S_3) traces out

the surface of a sphere, the so-called Poincaré sphere. We can illustrate this in Cartesian coordinates as



Poincaré sphere

$$+45°\ \text{pol.} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

$$\text{vertical pol.} = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}$$

$$L\ \text{pol.} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$$

Un- and partially polarised states lie within the sphere

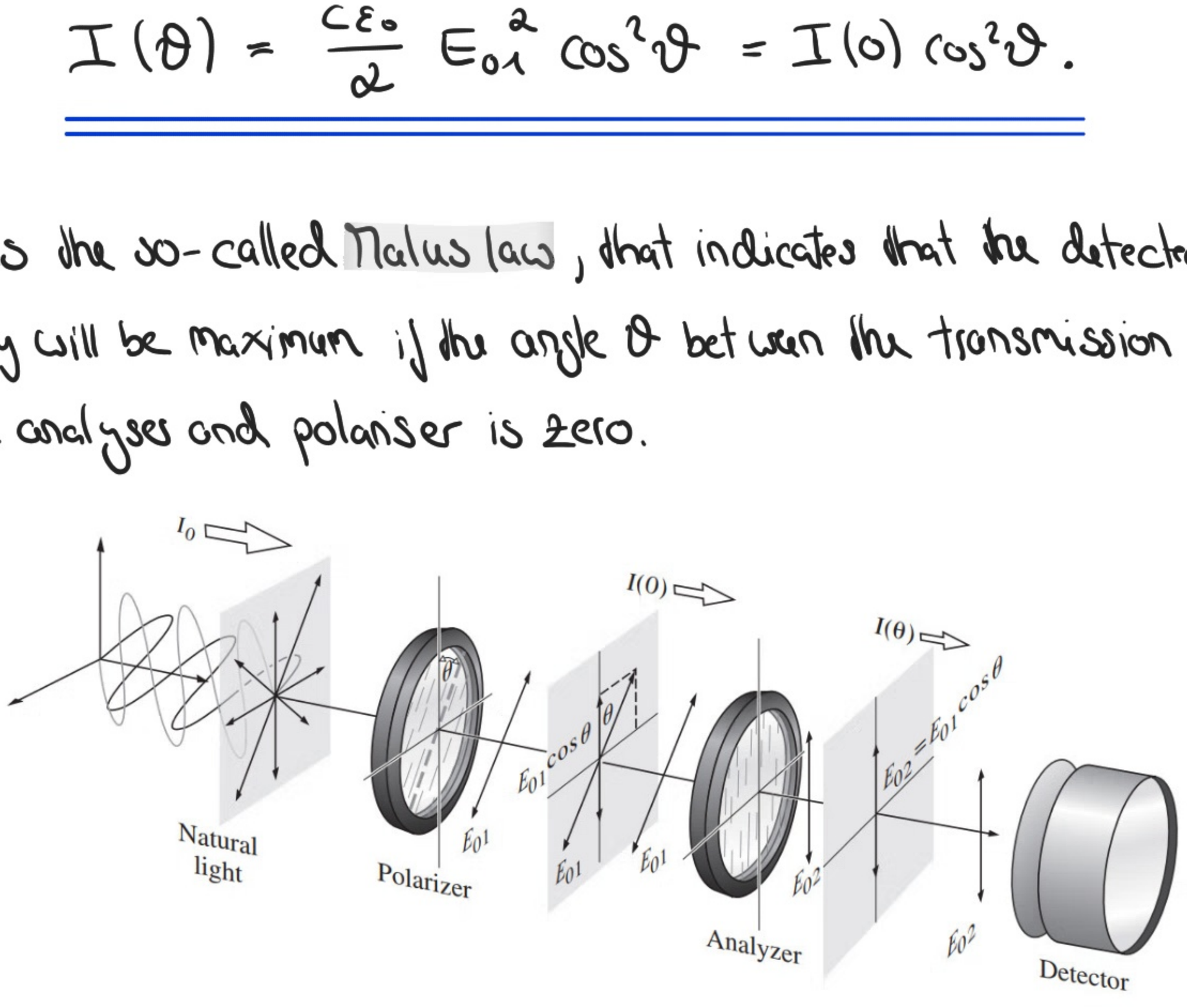
5.) Controlling polarisation states

Now that we know what polarisation is and how we can represent this, we need to consider methods that allow us to generate or manipulate the polarisation state of light. Polarisers, e.g., are optical devices whose input is natural light while the out put is one specific polarisation state (the rest is either reflected or absorbed). One aspect that all the different polarisers have in common is that they introduce some sort of asymmetry that affects different polarisation states in a different manner. Before

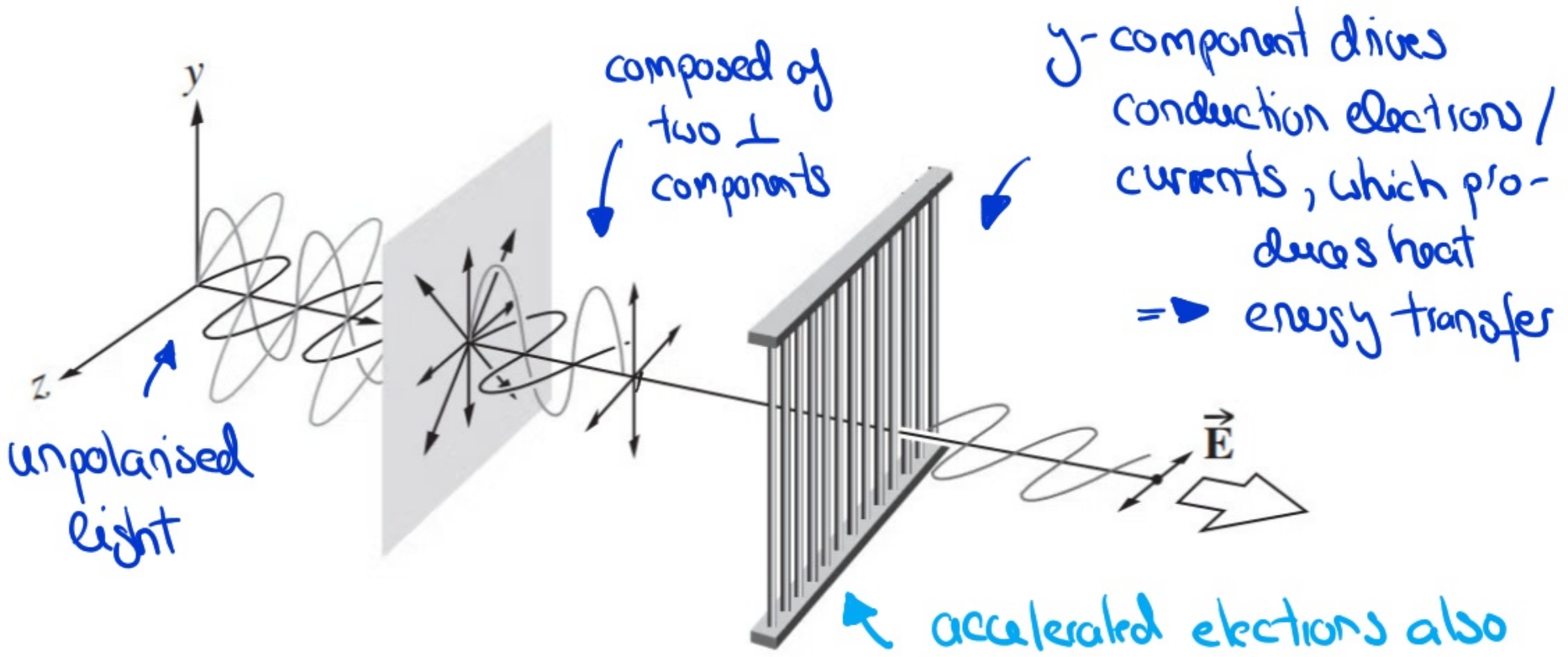
addressing different examples, we need to discuss how we can determine that a device is actually a linear polarises. Assume that natural light hits an ideal linear polarises and will become linearly polarised along the transmission axis of the device. Only the field component parallel to this axis will pass through the polariser unaffected. Rotating the transmission axis will however not change our measurement of the beam intensity, as all $\vec{E}$ field directions are equally present. To study the polarisation of the transmitted light, we thus use an identical analyser, whose transmission axis is vertical and only transmitts a specific fraction of the light proportional to $\cos\vartheta$ as illustrated below. The irradiance at the detector will thus be

$$I(\vartheta) = \frac{c\varepsilon_0}{2} E_{01}^2 \cos^2\vartheta = I(0)\cos^2\vartheta .$$

This is the so-called Malus law, that indicates that the detected intensity will be maximum if the angle ϑ between the transmission axes of the analyses and polariser is zero.

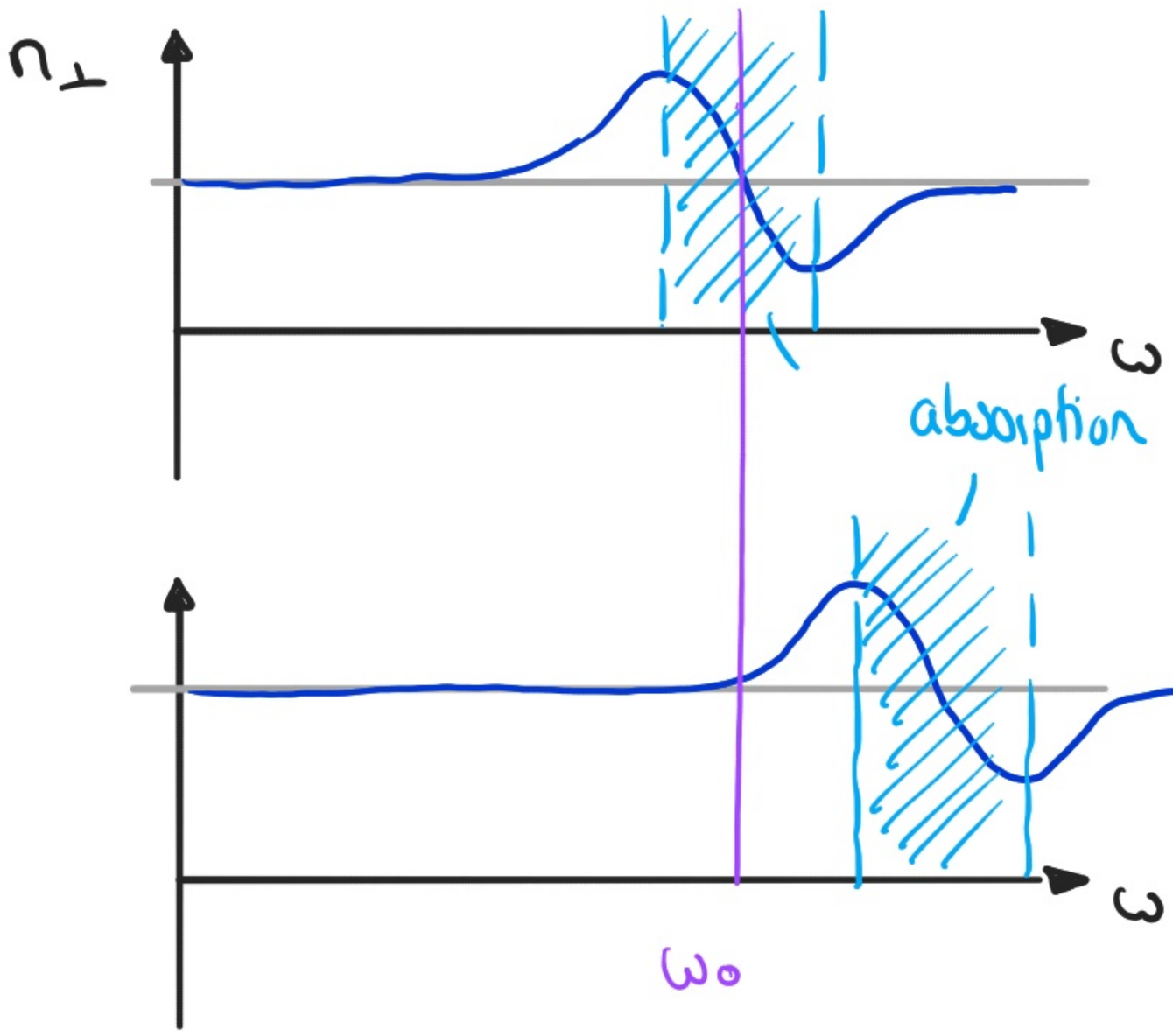


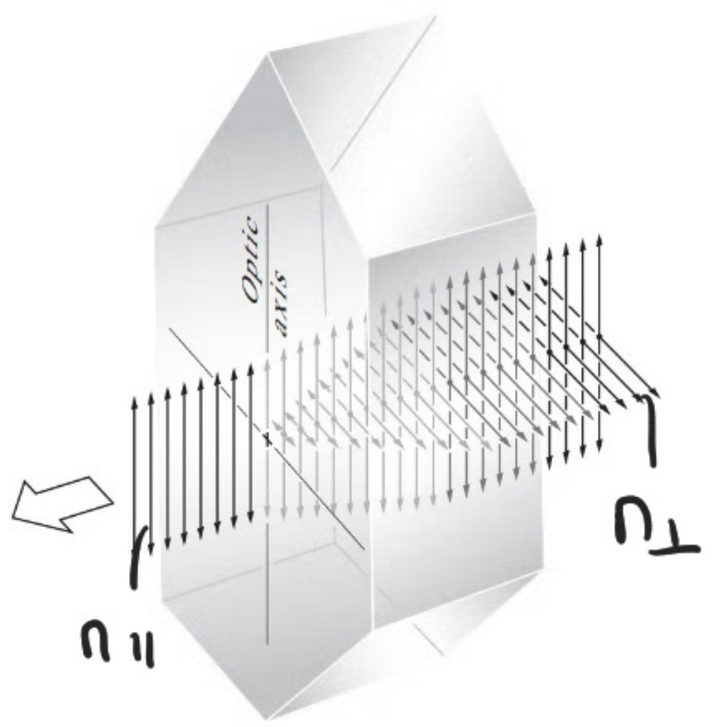
The simplest form of a polariser relies on dichroism, which refers to the selective absorption of one of the two orthogonal linear polarisation states. The medium is itself highly anisotropic and the simplest device of this sort is essentially a wire grid as explained below



The above sketch illustrates that the transmission axis is perpendicular to the orientation of the wires. While the above concept works very well in the long wavelength limit (e.g. with microwaves), it is difficult to use for light, and common modern polarisers rely on long molecules stretched along a specific axis (as done in so-called polaroids used e.g. in LCD displays or sunglasses) or crystalline materials that respond differently to electric fields along different crystalline axes. Such dichroic crystals typically have an internal anisotropy in their crystal lattices causing the electrons to respond depending on the direction of the elec-

tric field, i.e. the polarisation. The material has an internal optical axis and the light perpendicular to this axis is strongly absorbed. In a very naive picture, one could explain this by envisaging two different Lorentz models with different resonance frequencies depending on whether the $\vec{E}$ field was parallel or perpendicular to the optical axis :





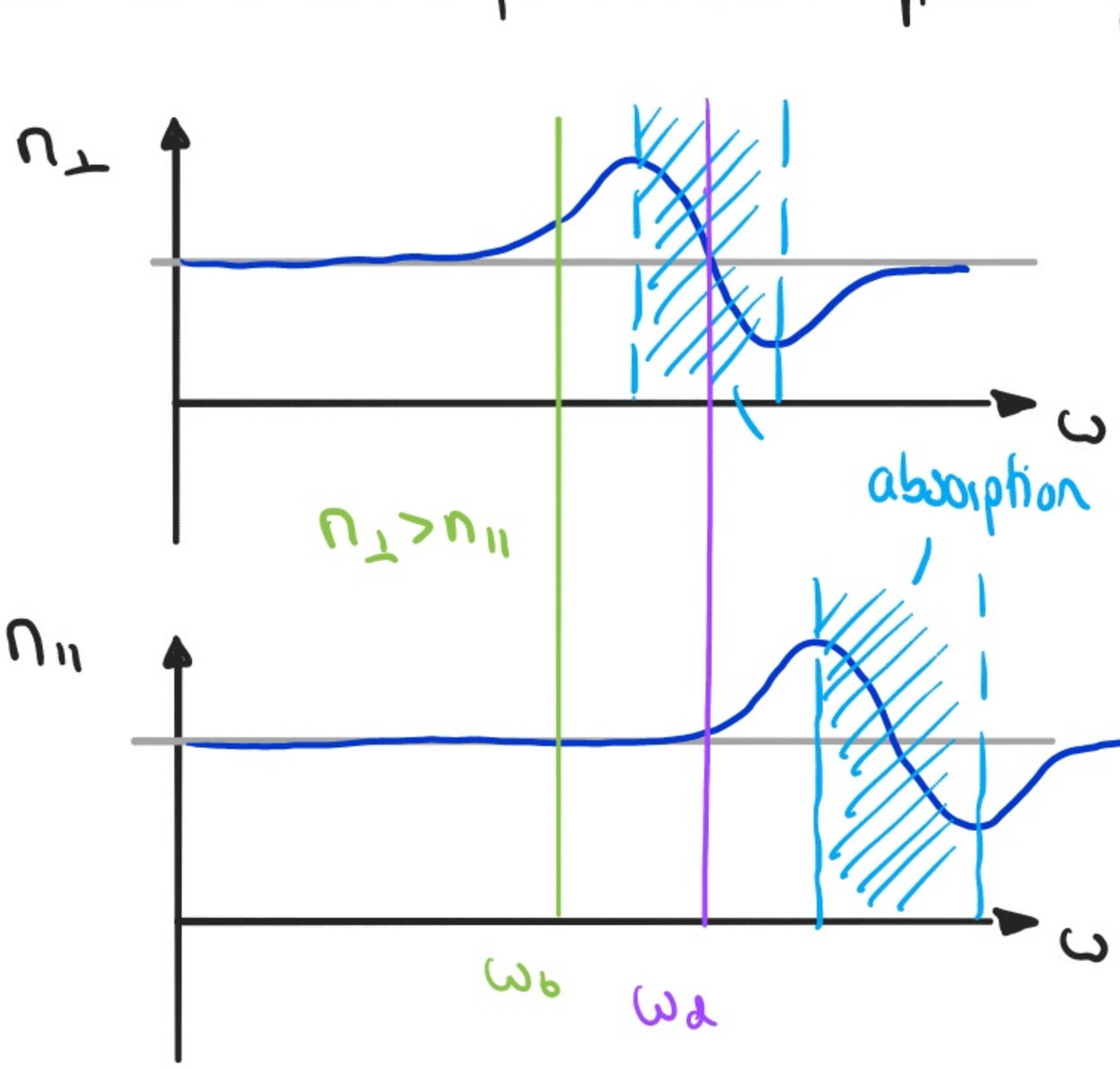
At the frequency ω_0, the parallel component is transmitted, while the perpendicular one is strongly absorbed. Note further that this behaviour is strongly frequency dependent and dichroic behaviour is typically observed for only a small range of frequencies.

Phys 434 – Lecture 14

Birefringence, Scattering, Retarders

1.) Birefringence

At the end of last lecture, we invoked two different Lorentz models to explain the behaviour of dichroic crystals, i.e. at a frequency, where light ⊥ polarised to the optical axis is strongly absorbed, while the ∥ polarisation is not. Consider the same picture for a different frequency ω_b – the two polarisations will experience two different refractive indices, $n_\perp > n_\parallel$.



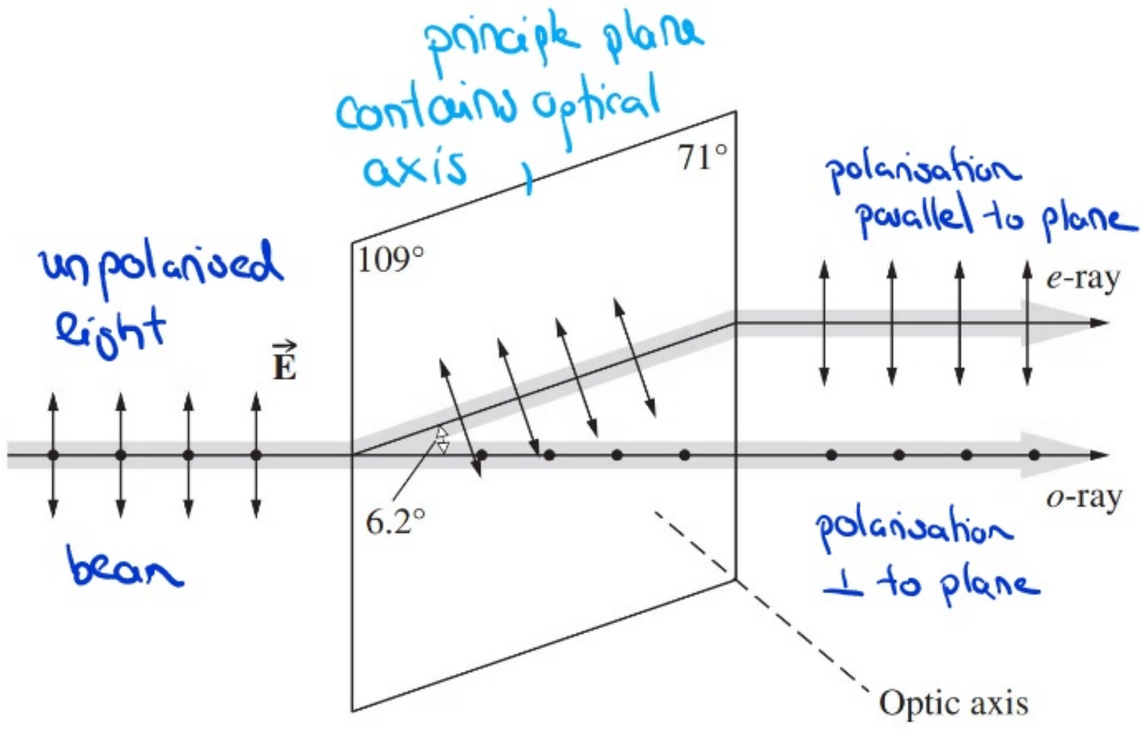
This is the underlying idea for the phenomenon of birefringence.

Many crystalline substances exhibit this anisotropic behaviour in their optical properties as a result of an internal anisotropy, i.e. electronic bin-

(essentially the underlying atomic structure)

ding forces are anisotropic, providing the microscopic justification for the two Lorentz models. Dichroic crystals are a special subcategory of this. For many birefringent materials, however, the resonance frequencies are above the optical range, so that absorption is negligible but many other phenomena are present. We will discuss some of these in the following.

Looking for example through a calcite crystal, we will see a double image; i.e. sending a narrow beam of natural light through the crystal, it will split and emerge as two parallel beams. Rotating the crystal, one of the rays will be stationary while the other one will move in a circle following the crystal motion. For historical reasons, the former one is referred to as the ordinary ray and the latter as the extraordinary ray. Examining the rays through an analyser shows that the light is linearly polarised and the rays' respective polarisations are orthogonal to each other.



To understand where this is coming from consider Maxwell's equation in

a medium, where no free charges or currents are present. We have

$$\nabla \cdot \vec{D} = 0, \qquad \nabla \times \vec{E} = -\partial \vec{B}/\partial t,$$

$$\nabla \cdot \vec{B} = 0, \qquad \nabla \times \vec{H} = \partial \vec{D}/\partial t.$$

Now consider a plane wave propagating in such a medium, so that

$$\vec{E} = \vec{E}_0 e^{i(\vec{k}\vec{r} - \omega t)} \quad \& \quad \vec{B} = \vec{B}_0 e^{i(\vec{k}\vec{r} - \omega t)}.$$

anisotropic medium

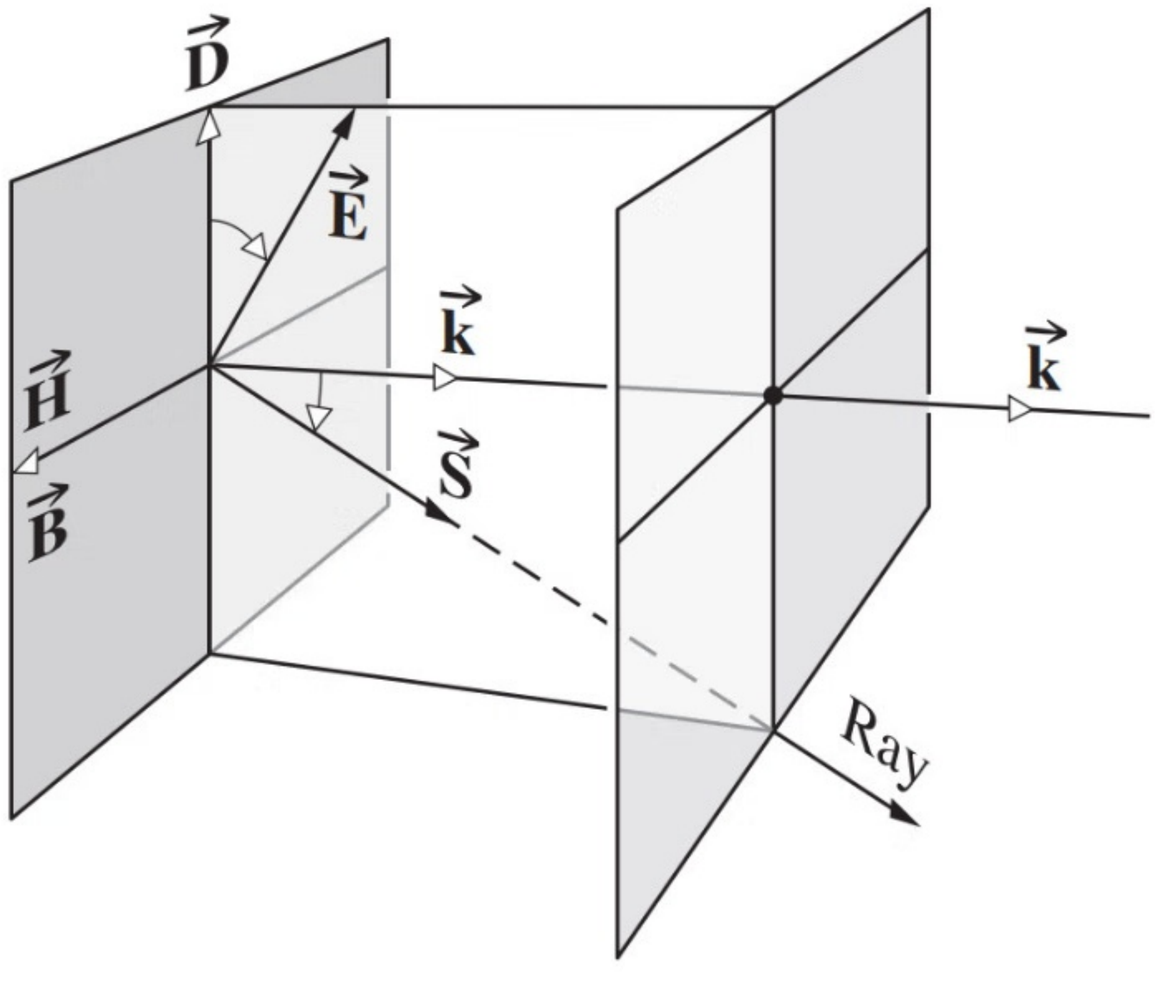
Neglecting magnetic response, we have $\vec{H} = \vec{B}/\mu_0$ and $\vec{D} = \underline{\underline{\varepsilon}}\vec{E}$. For the most general case, $\underline{\underline{\varepsilon}}$ is the so-called permittivity tensor. It will change the direction of $\vec{D}$ with respect to $\vec{E}$ but leave $\vec{k}$ and ω unaffected. Substituting the plane-wave solutions into Maxwell's equations, we replace $\vec{\nabla} \to i\vec{k}$ and $\partial t \to -i\omega$ to find

$$\vec{k} \cdot \vec{D} = 0, \qquad \vec{k} \times \vec{E} = \omega \mu_0 \vec{H},$$

$$\vec{k} \cdot \vec{H} = 0, \qquad \vec{k} \times \vec{H} = -\omega \vec{D}.$$

This tells us that even for an anisotropic medium, where $\underline{\underline{\varepsilon}}$ is not a scalar but a tensor, $\vec{H} \perp \{\vec{k}, \vec{D} \& \vec{E}\}$ and $\vec{k} \perp \vec{D}$, but the electric field $\vec{E}$ does no longer have to be $\perp$ to $\vec{k}$. As we neglect any magnetic response $\vec{H} \parallel \vec{B}$, so the fields vibrating within the medium are $\vec{D}$ and $\vec{H}$ (or $\vec{B}$) and not as before $\vec{E}$ and $\vec{B}$.

The direction of the ray corresponds to the direction of energy transfer i.e. the Poynting flux $\vec{S} = \frac{1}{2}\vec{E} \times \vec{H}$, which now is not necessarily parallel to the wave vector $\vec{k}$. The geometry of all vectors is illustrated below



To understand the response of the two distinct rays, consider a choice of coordinate system where $\underline{\underline{\varepsilon}}$ is diagonal so that

$$D_x = \varepsilon_{xx} E_x = \varepsilon_o E_x,$$
$$D_y = \varepsilon_{yy} E_y = \varepsilon_o E_y,$$
$$D_z = \varepsilon_{zz} E_z = \varepsilon_e E_z,$$

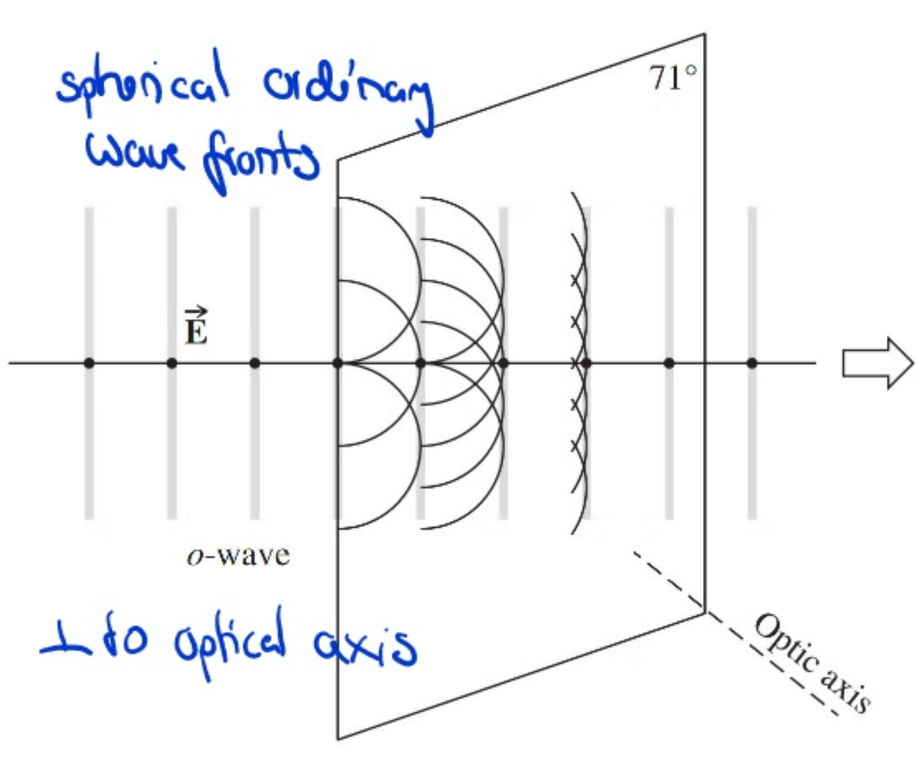
letter 'o'

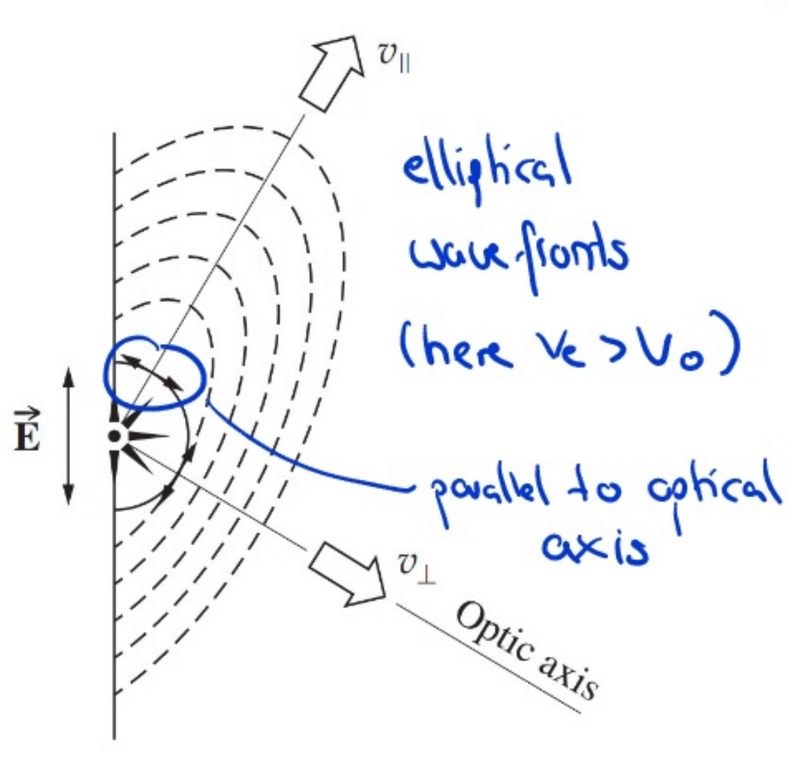
- for a uniaxial crystal, one special direction, i.e. the optical axis

i.e. the optical axis is aligned with the z-direction and $\varepsilon_o = n_o^2 \varepsilon_0$ and $\varepsilon_e = n_e^2 \varepsilon_0$ (capturing the behaviour of the ordinary and extra-ordinary ray). The details of the response will depend on the angle the

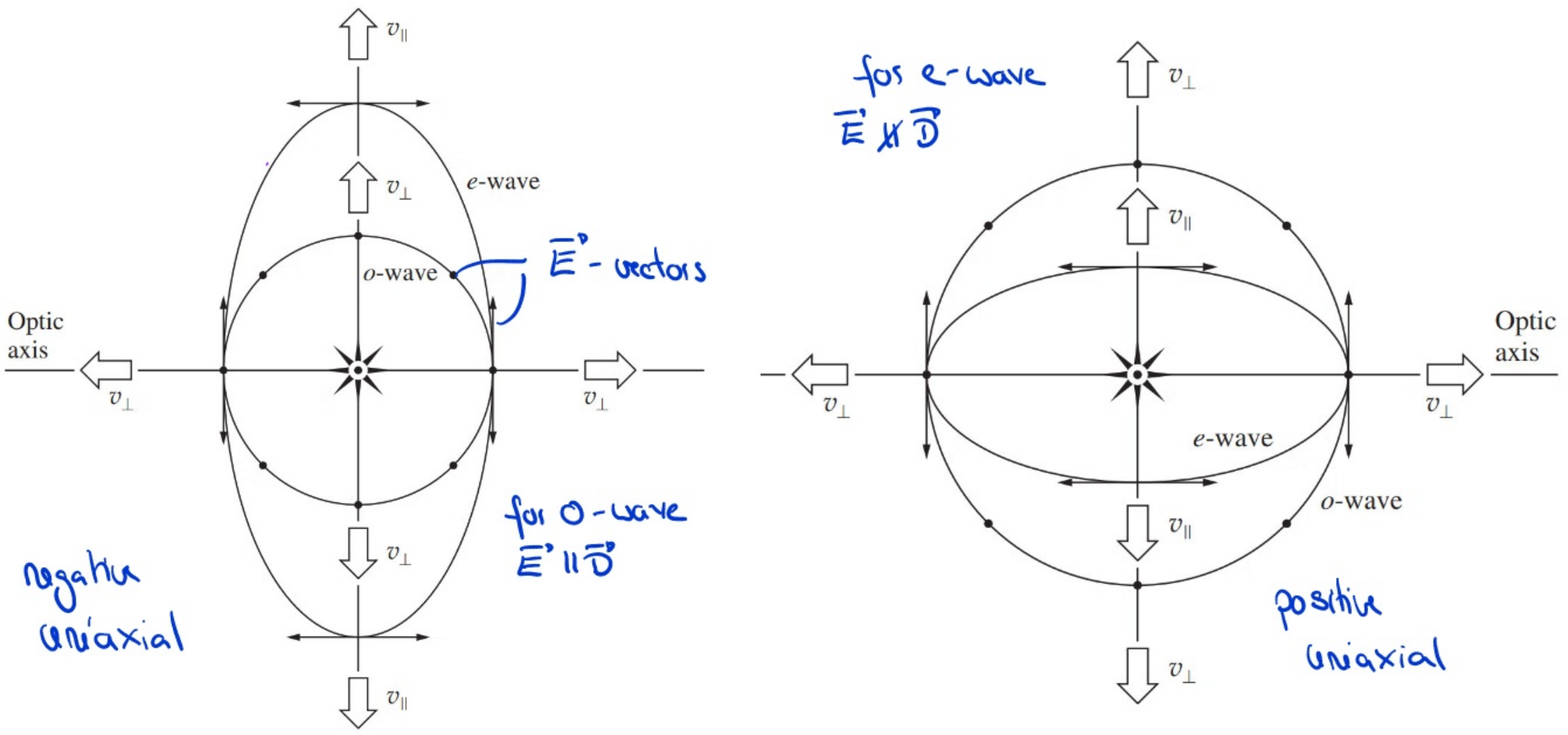
optical axis makes with respect to the incident ray, i.e. the propagation direction $\vec{k}$ (see PS #4), but we can readily observe that the rays will propagate with two different velocities, i.e. $v_o = c/n_o$ and $v_e = c/n_e$.

Remembering that v_e is the velocity of the wave component ∥ to the optical axis and v_o the velocity of the wave component ⊥ to the optical axis, we can invoke Huygens' principle to understand the behaviour of a birefringent crystal. As the wave hits the surface, the electrons are driven into oscillation and emit secondary wavelets. These will superimpose to form the refracted wave. For the $\vec{E}$ field component that is perpendicular to the optical axis, only the ordinary refractive index n_o matters and the secondary wavelets will propagate in all directions with the same velocity v_o as they would in an isotropic medium. This 'normal' behaviour constitutes the ordinary ray. However, for the polarisation component that is parallel to the optical axis the respective velocity will differ and the resulting wavefronts will no longer be spherical but instead elliptical. This diverts the propagation direction from $\vec{k}$ and constitutes the 'extraordinary' ray.





The form of the resulting ellipsoid, will depend on whether $v_e > v_o$ (ellipsoid elongated normal to the optical axis) or $v_e < v_o$ (elongated along the optical axis). More precisely, in the former case we require $n_e < n_o$ (i.e. as is the case for a negative uniaxial crystal like calcite), while the latter corresponds to $n_e > n_o$ (as one has for a positive uniaxial crystal like quartz or ice). The respective ellipsoids can be illustrated as



As the ordinary ray propagates like it would in an isotropic medium and travels at the same speed in every direction, it satisfies snell's law. The extraordinary ray however doesn't travel through a birefringent crystal at the same speed in all directions, so it doesn't greatly obey Snell's law.

To conclude the discussion on birefringence, we would like to know what the dispersion relation relating ω and $\vec{k}$ is inside a birefringent crystal

-for an arbitrary direction of propagation. Recall that in an isotropic medium we derive $\omega = ck/n$ by deriving a wave equation by taking the curl of Faraday's law. Repeating the same procedure for the anisotropic medium, we arrive at

$\neq 0$ because of $\underline{\underline{\varepsilon}}$

$$\nabla \times (\nabla \times \vec{E}) = \nabla (\nabla \cdot \vec{E}) - \nabla^2 \vec{E} = -\partial_t \nabla \times \vec{B},$$

$$\Rightarrow \quad \vec{k} \times (\vec{k} \times \vec{E}) = \vec{k} (\vec{k} \vec{E}) - \vec{k}^2 \vec{E}$$
$$= -\omega^2 \mu_0 \vec{D} = -\omega^2 \mu_0 \underline{\underline{\varepsilon}} \vec{E} .$$

This equation defines the dispersion relation $\omega(\vec{k})$. It will depend on the direction of $\vec{k}$ relative to the optical axis as well as the two permissivities of the birefringent crystal. Assuming that $\vec{k}$ and the optical axis span an angle ϑ, it can be shown (see 9.3.4) that the dispersion relation can be written as $\omega = c|\vec{k}| / n(\vartheta)$ with

$$\underbrace{1/n^2(\vartheta) = \cos^2\vartheta / n_e^2 + \sin^2\vartheta / n_o^2}_{\text{extraordinary ray}} \quad \text{or} \quad \underbrace{n(\vartheta) = n_o}_{\text{ordinary ray}} .$$

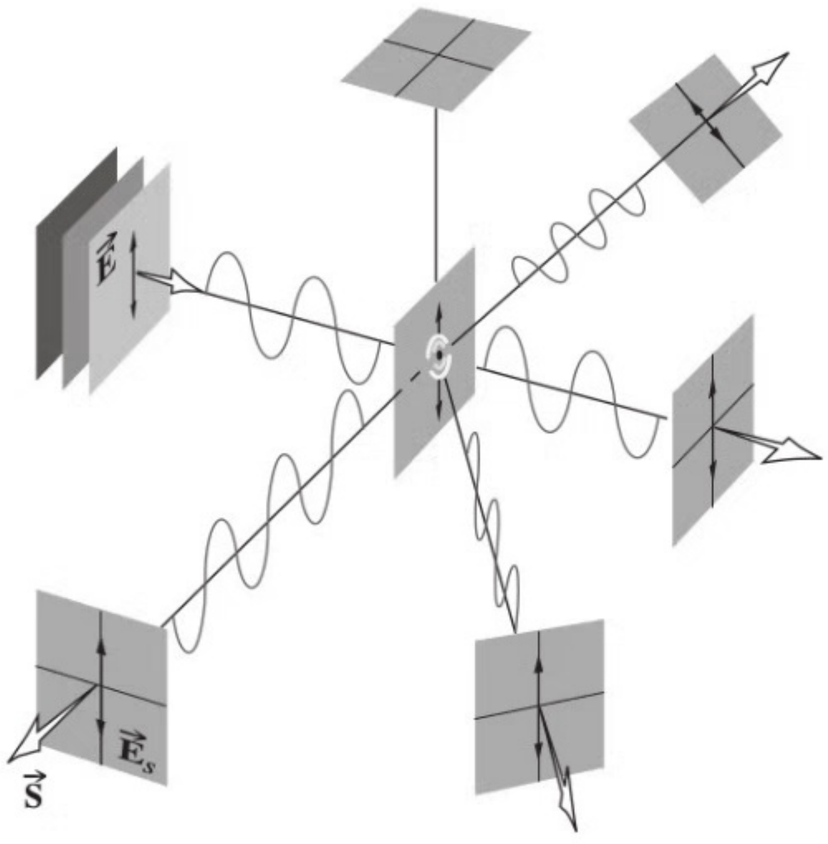
While this effect can be employed to build polarisers, birefringence and dichroism are not the only ways to polarise light:

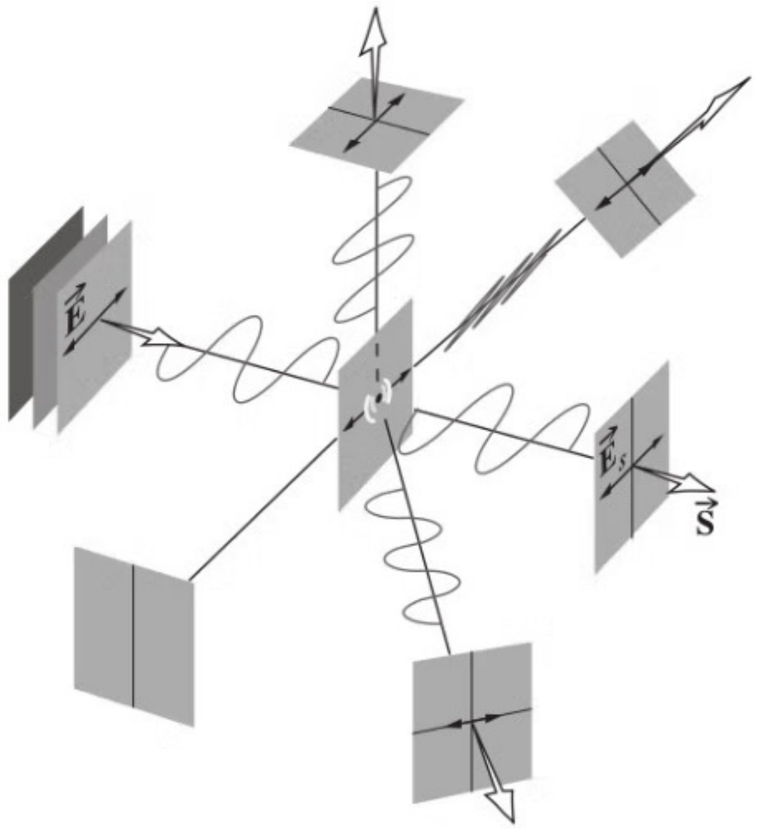
2.) Polarisation by Scattering

$d \ll \lambda$

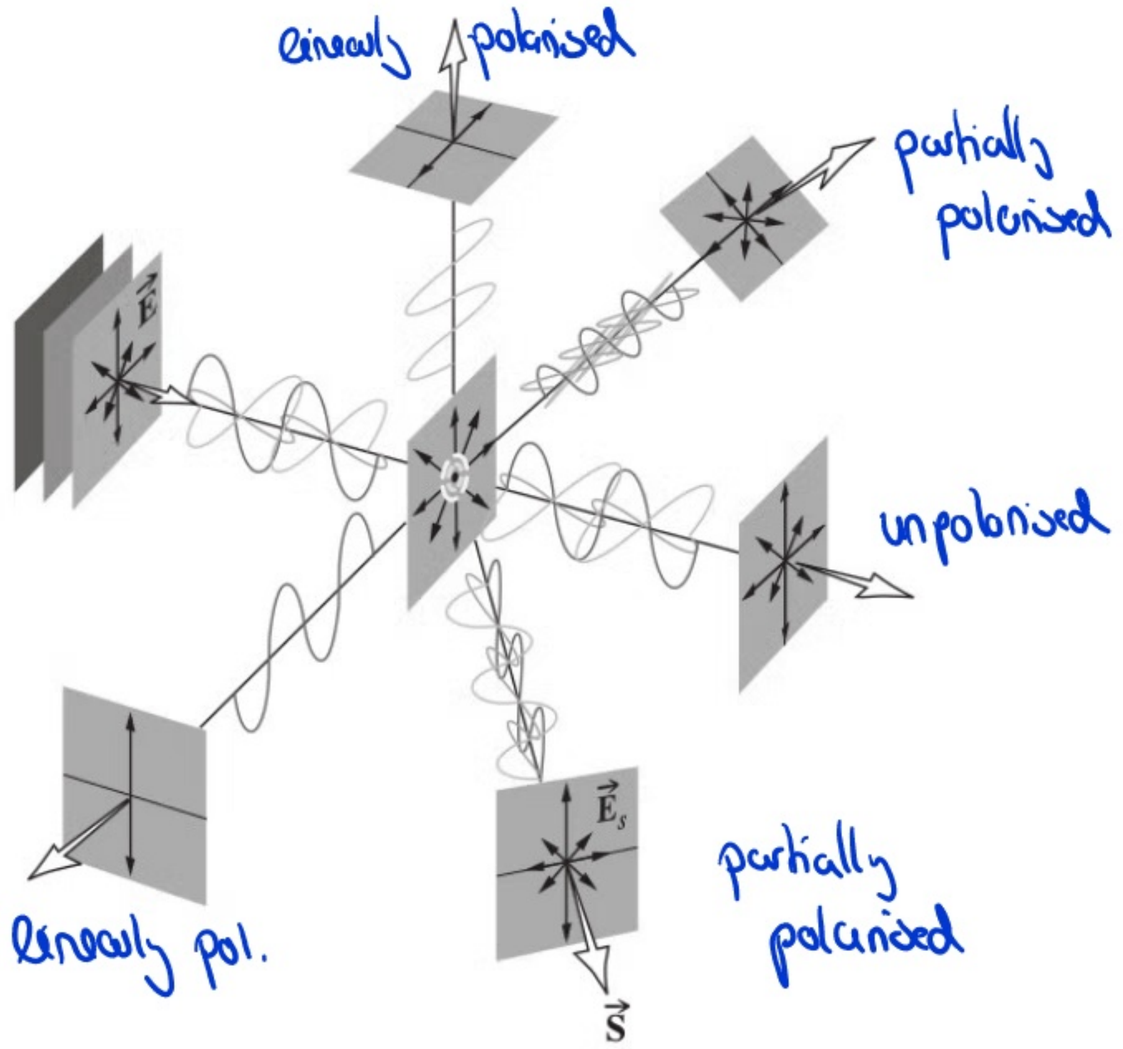
We already discussed that sunlight streaming into the atmosphere from one direction is scattered in all directions by the air molecules. While the red end of the spectrum is mainly unaffected, the blue / high-frequency end is strongly scattered making the sky appear blue.

Consider a linearly polarised plane wave incident on an air molecule. As illustrated below, the electric fields of the scattered radiation are parallel to the incident $\vec{E}^0$ vector and hence perpendicular do the direction of propagation. As the oscillator does not emit in the direction of its axis, there won't be a field component. Considering instead an unpolarised incident wave (which can be decomposed into two orthogonal, incoherent states), the light will be unpolarised in forward direction, becoming increasingly more polarised as the angle increases. For the directions $\perp$ to the primary beam, the light is completely linearly polarised.





- for unpolarised light: superposition of the two cases



Looking at a small part of the sky, one can verify that the sun light is indeed partially polarised. The partial comes from the depolarising effects of multiple scatterings and anisotropies in the air molecules.

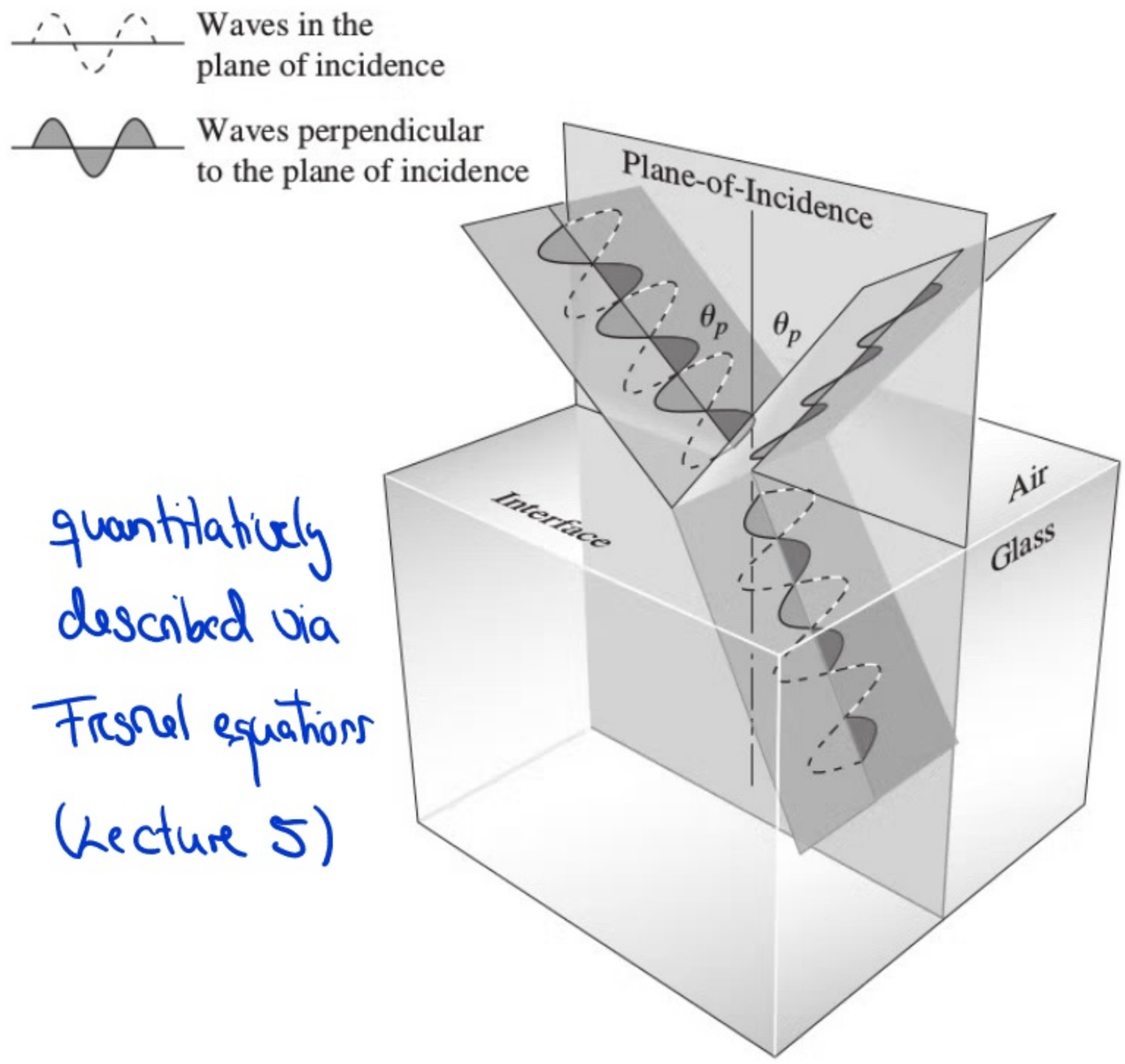
3.) Polarisation by reflection

One of the most common sources of polarisation is reflection from dielectric media. As discussed in Lecture 4, the reflection coefficients for light polarised parallel or perpendicular to the plane of incidence differ, i.e. $R_{\parallel} \neq R_{\perp}$. Specifically, for rays incident at the Brewster angle θ_p, we saw that the reflected light wave polarised $\parallel$ to the incidence plane would vanish entirely. The reflected light ray is thus perfectly linearly polarised $\perp$ to the plane of incidence, while the transmitted ray is partially polarised $\parallel$ to the plane of incidence as illustrated below:

This occurs at $\theta_p + \theta_t = 90°$, in which case:

$$\tan \theta_p = n_t / n_i .$$

Brewster's law



quantitatively described via Fresnel equations (Lecture 5)

The problem in using this concept is that the reflected beam, although completely polarised, is weak and the majority of the energy is put into the transmitted beam. To construct an effective polariser, one could however stack many thin plates and add the intensity of multiple reflected beams.

4.) Retarders

Polarisers discussed so far are able to set the polarisation to a defined state, but they are not able to transform coherently between different states. Optical elements that are able to achieve this are called retarders. They change the state of polarisation by introducing a relative phase between the two linear polarisation directions, i.e. we want a device that shifts

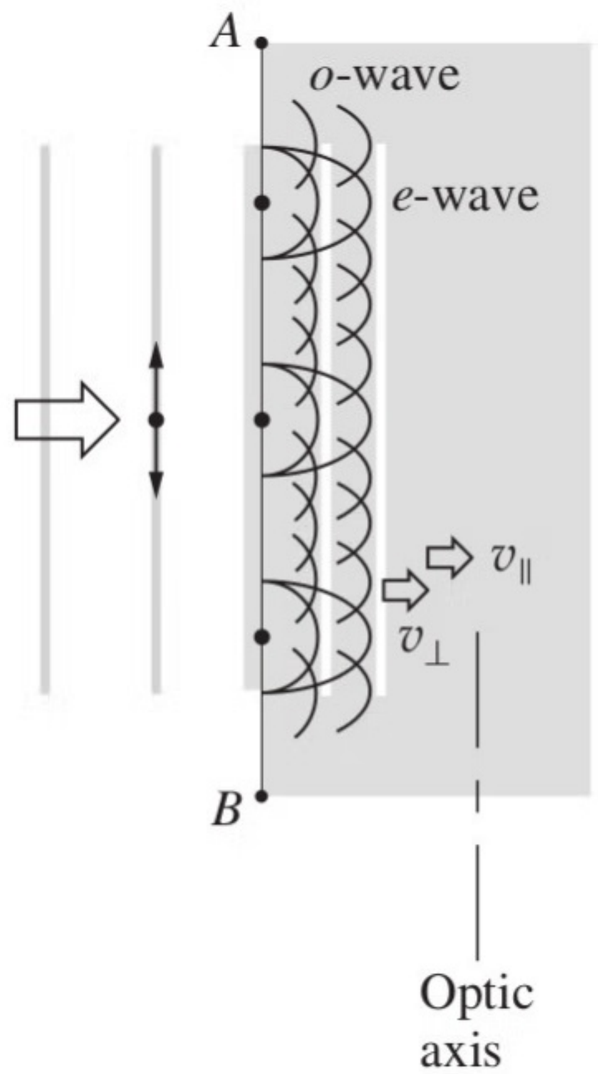
the relative phase ε between two plane waves (see Lecture 12) by a fixed amount. This can be achieved by means of birefringence. Suppose that a material has an optical axis so that $n_e < n_o$. The respective velocities of the extraordinary and ordinary rays thus satisfy $v_e > v_o$. Light travelling through thickness d of this material will thus pick up a different phase depending on its polarisation. Now consider the material to be cut in a way that the optical axis is parallel to the two interfaces (i.e. $\perp$ to the direction of propagation) as illustrated below. The two resulting wave fronts of the e & o wave will be parallel but move at different speeds and hence leave the material with different phases. From the relative optical path length difference ΔOPL, we can deduce the phase difference $\Delta\varphi$:

$$\Delta OPL = d\,|n_o - n_e|\,,$$

$\lambda_0 \hat{=}$ vacuum wavelength

$$\rightarrow\quad \Delta\varphi = \frac{2\pi}{\lambda_0}\,\Delta OPL$$

$$= \frac{2\pi}{\lambda_0}\, d\,|n_o - n_e|\,.$$



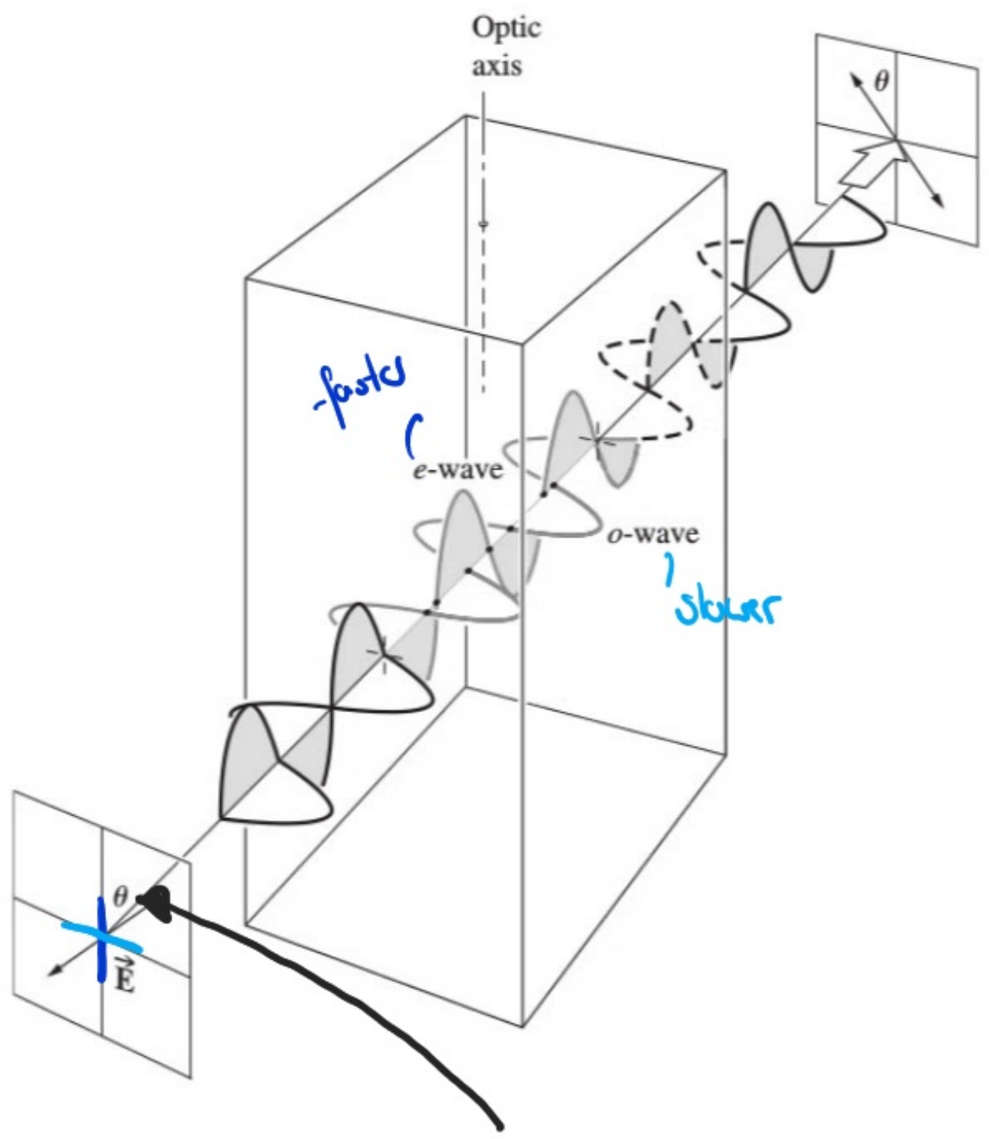
The effect of such a 'wave plate' on the polarisation state of the incoming light will thus depend on both the angle of the optical axis and the thickness of the plate. E.g. if $\Delta\varphi = 2\pi$, the relative retarda-

tion is one full wavelength and the two rays are back in phase, so that there is no observable effect on the incident beam and changes are only dependent on the wavelength λ_0, i.e. of chromatic nature. Alternatively,

- for $\Delta\varphi = \pi\ (180°)$, so-called **half-wave plates**:

rotate the polarisation by 180° of longitude on the Poincaré sphere; e.g. for 45° polarisation the polarisation is flipped to -45°; i.e. the **polarisation axis** is essentially **rotated**

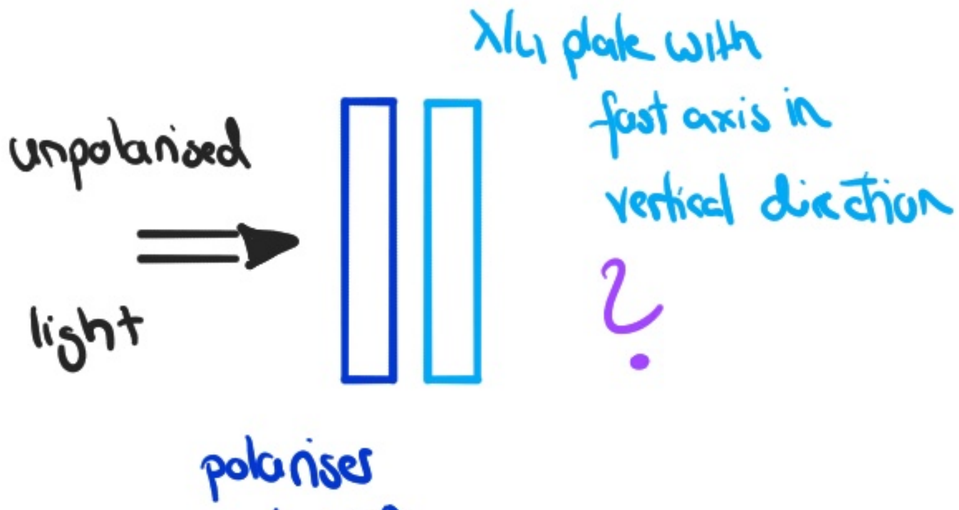
flip the axis of line-arly polarised / elliptical light; change the handedness of circular / elliptical light



- for $\Delta\varphi = \pi/2\ (90°)$, so-called **quater-wave plates**:

convert **linear polarisation** to **circular polarisation** (if $\vartheta = 45°$) or elliptical polarisation (for other values of ϑ); **handedness** follows from the direction it would take to rotate the initial $\vec{E}$ vector towards the slow axis ($\perp$ to optical axis) following the smallest angle.

Example:



- this is essentially a circular polariser !!

output state
≙ right circular polarised light

After the polariser, we can characterise the state as $1/\sqrt{2}\begin{pmatrix}1\\1\end{pmatrix}$.
As the vertical direction propagates faster, the horizontal component will pick up a phase of $\pi/2$:

$$1/\sqrt{2}\begin{pmatrix}e^{i\pi/2}\\1\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}i\\1\end{pmatrix}$$

$$= \frac{1}{\sqrt{2}}\, i \begin{pmatrix}1\\-i\end{pmatrix}$$

global Jones vector phase

One final note: In practice, it is often very difficult to make plates that have a true width on the order of $\lambda/2$ or $\lambda/4$, because they would have to be very thin. However, for monochromatic light, $(m+1/2)\lambda$ or $(m+1/4)\lambda$ plates, where m is an integer, would have the same effect on the incident beam. Compared to the zeroth order plates, such multi-order retarders are much easier to manufacture and thus cheaper.

Phys 434 - Lecture 15

Polarisers, Optical Activity, Liquid crystals

1.) Polarisers

At the end of last lecture, we saw that a $\lambda/4$ plate modifies a linearly polarised state to a circularly polarised one and we could represent this via a corresponding change in the Jones vector. Generally, propagation through polarisation elements can be mathematically treated by matrix multiplications because each of the interactions is linear.

In the Jones vector representation, this is achieved by means of the so-called Jones matrices. Here an optical element is described by a 2×2 matrix that changes the incident Jones vector $\vec{E}_i$ to the output vector $\vec{E}_t$:

$$\rightarrow \quad \vec{E}_t = \underline{\underline{A}}\,\vec{E}_i = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \vec{E}_i ,$$

$$\text{or} \quad E_{tx} = a_{11} E_{ix} + a_{12} E_{iy}, \; E_{ty} = a_{21} E_{ix} + a_{22} E_{iy}.$$

In the example from last lecture, where we combine a 45° polariser and a $\lambda/4$ plate with the fast axis positioned in the vertical direction, the two respective matrices can be defined as

45° polarizer $A = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$; $\lambda/4$ plate vertical fast axis $A = e^{i\pi/4}\begin{pmatrix} 1 & 0 \\ 0 & -i \end{pmatrix}$.

The combined effect of both elements is thus determined by

$$\vec{E_t} = e^{i\pi/4}\begin{pmatrix} 1 & 0 \\ 0 & -i \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\vec{E}_i$$

$$= \frac{1}{\sqrt{2}}e^{i\pi/4}\begin{pmatrix} 1 & 1 \\ -i & -i \end{pmatrix}\vec{E}_i .$$

For any given incidence state $\vec{E}_i$, this matrix will produce a right-circularly polarized state. In general, we have $\vec{E}_t = A_n \dots A_2 A_1 \vec{E}_i$. As matrix multiplication is not commutative, i.e. $A_2 A_1 \neq A_1 A_2$, the order of your calculation matters !! Finally note that there exists also a matrix representation applicable to the Stokes parameters. Those 4×4 matrices are known as Mueller matrices. For more information on this see Hecht section 8.13.3.

2.) Optical Activity

While birefringent materials exhibit different indices of refraction when exposed to different linear polarisations (∥ and ⊥ to the optical axis), materials that are optically active have different indices of refraction for left- and right-circular polarisations. More precisely, the

left and right circular polarisation components will pick up a relative phase $\Delta\varphi$ as they propagate through optically active media (see also DEMO #2). This requires the respective medium to have a handedness itself, which has important consequences for chemical and biological substances.

To understand the implications of optical activity, we want to analyse how such media affect linearly polarised light. Recalls that any linearly polarised optical field can be written as a superposition of circular polarisations. Consider for example a horizontally polarised state:

$$\begin{pmatrix}1\\0\end{pmatrix} = 1/\sqrt{2}\left[\overbrace{1/\sqrt{2}\begin{pmatrix}1\\i\end{pmatrix}}^{L} + \overbrace{1/\sqrt{2}\begin{pmatrix}1\\-i\end{pmatrix}}^{R}\right].$$

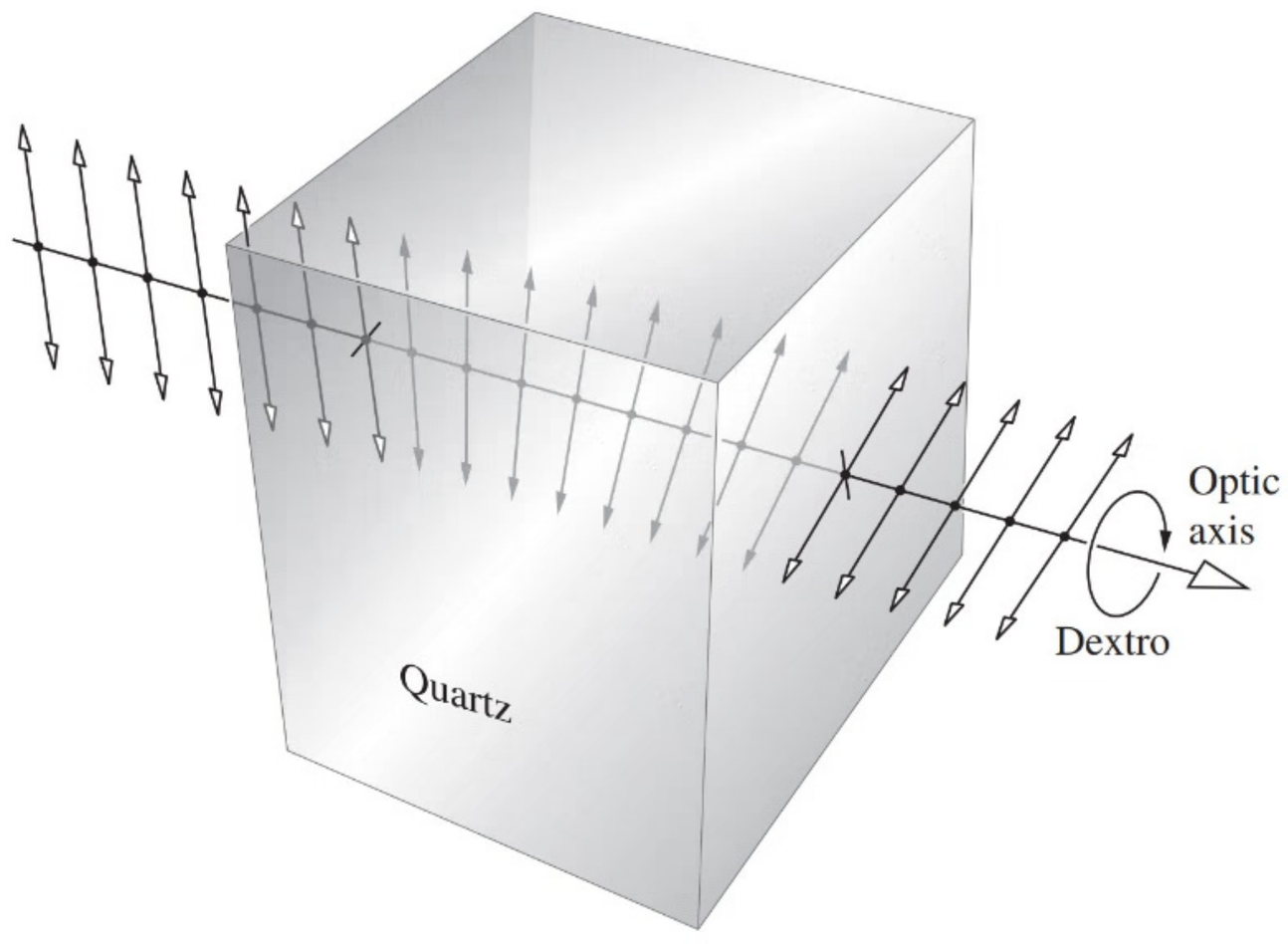
An optically active medium of width d will introduce a phase shift $\Delta\varphi = 2\pi/\lambda_0 \cdot d(n_L - n_R)$ (remember λ_0 is the vacuum wavelength). Applying this phaseshift to the L state, we obtain

$$\frac{1}{2} e^{i\Delta\varphi}\begin{pmatrix}1\\i\end{pmatrix} + \frac{1}{2}\begin{pmatrix}1\\-i\end{pmatrix} = \frac{1}{2} e^{i\Delta\varphi/2}$$

$$\times \begin{pmatrix} e^{i\Delta\varphi/2} + e^{-i\Delta\varphi/2} \\ i e^{i\Delta\varphi/2} - i e^{-i\Delta\varphi/2} \end{pmatrix} = e^{i\Delta\varphi/2}\begin{pmatrix}\cos[\Delta\varphi/2]\\ -\sin[\Delta\varphi/2]\end{pmatrix}.$$

The last expression shows that the linear polarisation state is rotated by an angle $\alpha = \pi d/\lambda_0 (n_\alpha - n_R)$. Hence as the light wave travels through an optically active medium, the polarisation vector is continuously rotated. For $n_\alpha > n_R$, we obtain a rotation in clockwise direction (looking toward the source) also referred to as d-rotatory or dextrorotatory (from Latin dextro $\hat{=}$ right). For $n_\alpha < n_R$, the rotation is counter-clockwise, which is referred to as l-rotatory or levorotatory (from Latin laevo $\hat{=}$ left).

Dextrose is called dextrose because of its optical activity



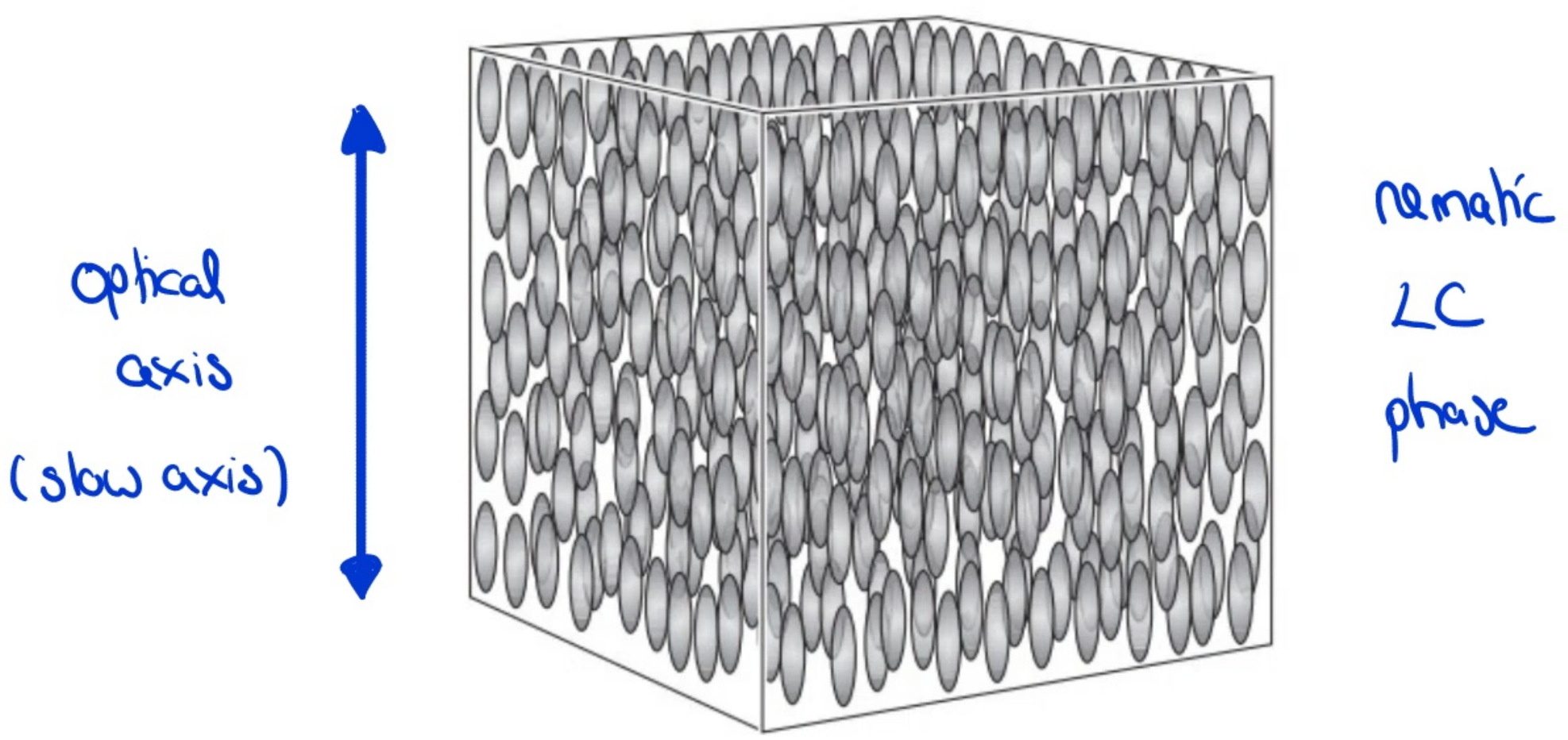
Microscopically, the optical activity originates from the underlying properties of the respective molecules. While we can easily characterise the observed effects in terms of an electromagnetic theory (as done above), the full description requires a quantum mechanical picture, which is beyond the scope of this course.

3.) Optical modulators

While the underlying concepts of optical activity are quite complicated, there is a number of physical effects that allow to externally alter the polarisation state of light. Materials can for example be made optically anisotropic by the application of mechanical stresses, which is known as photoelasticity or stress birefringence (essentially modifying the crystal structure and thus the optical axis). Similarly the presence of a magnetic field can alter the polarisation stat of light. This so-called Faraday effect rotates the plane-of-vibration of linearly polarised optical fields but the sign of the rotation angle changes depending on whether the direction of propagation is $\parallel$ to $\vec{B}$ or $\parallel$ do $-\vec{B}$. More precisely, empirically one can find for the rotation angle $\beta = V B d$, where d is the length of the medium, B the magnitude of the induction and V the so-called Verdet constant. By convention $V > 0$ corresponds to a l-rotatory medium for $\vec{k} \parallel \vec{B}$ and $V < 0$ to the d-rotatory case when $\vec{k} \parallel -\vec{B}$. The exact values of V depend on the temperatures as well as the frequency of the incident wave. The Faraday effect is an important phenomenon that helps to probe the strength of magnetic fields for example in astronomy (information about the interstellar medium) or solid-state physics (magneto-optical imaging). Finally note that polarisation can also be altered by an external $\vec{E}$ field (Kerr effect).

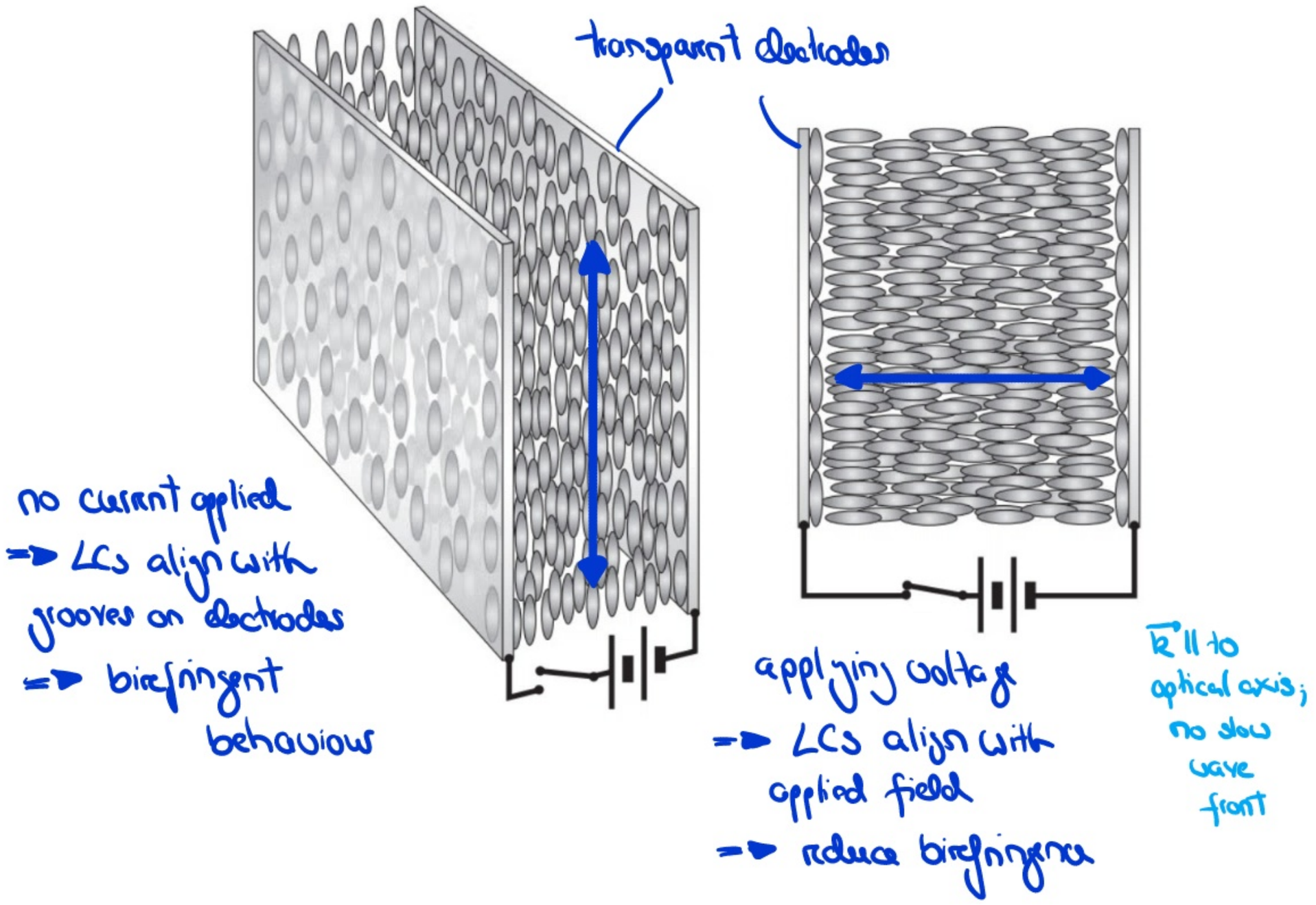
4.) Liquid crystal

Liquid crystals contain long molecules that form a new state of matter with properties between those of ordinary liquids and solids. Like in liquids the molecules can move about and thus lack positional order, but like crystals, the molecules strongly interact with each other and sustain some large-scale orientational order (in some random direction).

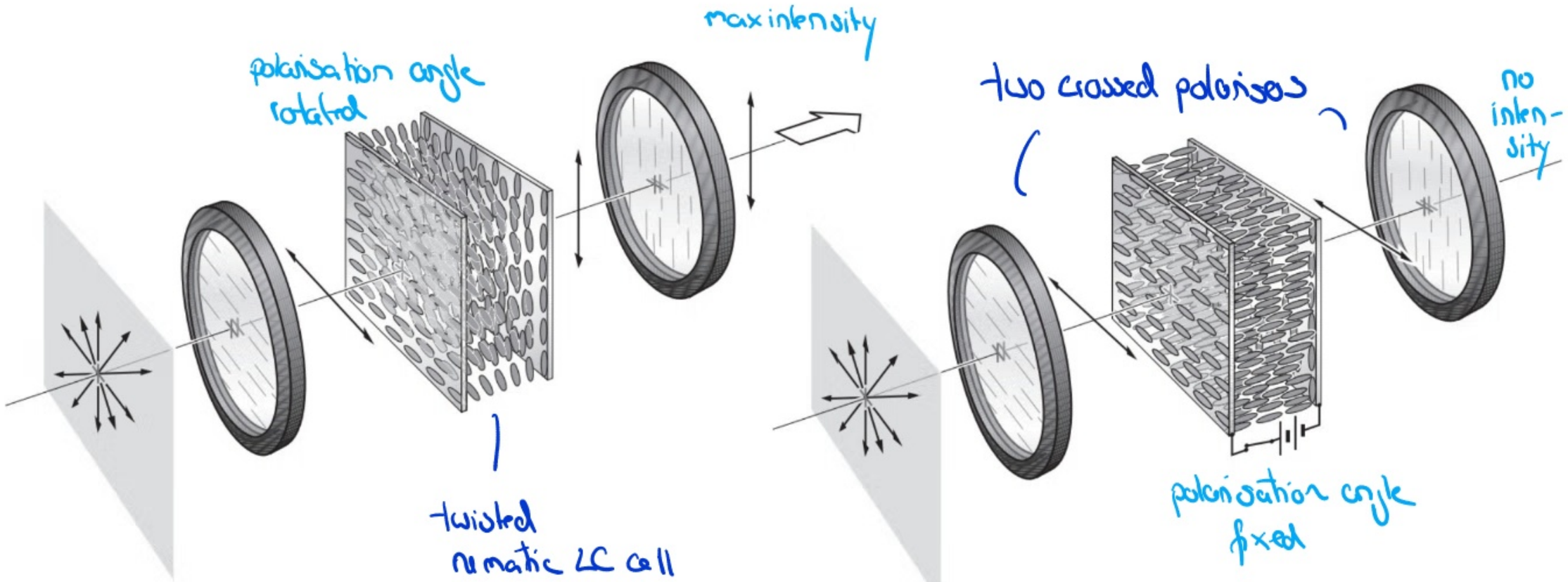


Because of this, the LC molecules behave essentially like an anisotropic dielectric that is positive uniaxial birefringent (remember $v_e < v_o$).

The long direction of the molecules determines the optical axis or slow axis (polarisation ∥ to optical axis has extraordinary behaviour). Its direction can be effectively controlled by applying an extenal electric field, e.g. by confining the liquid crystals between two transparent electrodes:



Such a cell can be effectively used as a voltage controlled phase modulator (adjust voltage do continuously vary between the two extremes). Now suppose that one of the electrodes is rotated by 90°. In absence of an external field, the cell will rotate the polarisation of the incident light as if it were an optically active medium, while it looses this ability when a field is applied. Putting this configuration between two crossed polarisers, the set up becomes a voltage-controlled switch that can either absorb a transmit light. Adding an additional mirror to the system, we can construct a very energy efficient LC display cell. It is not self-luminous and only uses ambient light. Energy is only required to apply an electric field and rotate LCs to cause absorption.


max intensity
polarisation angle rotated
twisted nematic LC cell
two crossed polarisers
no intensity
polarisation angle fixed

Phys 434 - Lecture 16

Introduction to Interference

1.) General considerations

So far, we have discussed superposition of scalar waves and polarisation properties of individual optical disturbances. To understand how polarisation affects the superposition of waves, we need do account for the vector nature of the electric field. Thus, the resultant field $\vec{E}$ at any point in space where two or more lightwaves overlap is equal do the vector sum of the individual constituents. Generally speaking optical interference corresponds to the interaction of two or more lightwaves with a resultant irradiance that deviates from the sum of the constituent irradiances.

To analyse this problem qualitatively, consider the superposition of two plane, linearly polarised waves of same frequency

polarisations of states given by $\vec{E}_{oj} = |E_{oj}| \begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix}$ ← Jones vector

$$\vec{E}_1 = \mathrm{Re}\left[\vec{E}_{o1} e^{i(\vec{k}_1\vec{r} - \omega t + \varepsilon_1)}\right],$$

$$\vec{E}_2 = \mathrm{Re}\left[\vec{E}_{o2} e^{i(\vec{k}_2\vec{r} - \omega t + \varepsilon_2)}\right],$$

where the resultant is given as $\vec{E} = \vec{E}_1 + \vec{E}_2$. As we are concerned

With the spatial and temporal changes of the resultant that can be measured by some kind of sensor, interference is best studied by looking at the irradiance $I = \varepsilon v \langle \vec{E}^2 \rangle_T$. As we are only concerned with relative irradiances, we can ignore the constant prefactor and simply calculate

$\langle \cdots \rangle_T$ ← time average

$$I = \langle \vec{E}^2 \rangle_T = \langle \vec{E} \cdot \vec{E} \rangle_T$$
$$= \langle \vec{E}_1 \cdot \vec{E}_1 + 2\,\vec{E}_1 \cdot \vec{E}_2 + \vec{E}_2 \cdot \vec{E}_2 \rangle_T$$
$$= I_1 + I_{12} + I_2 ,$$

---

where I_1 and I_2 denote the irradiances of the individual waves and I_{12} is the interference term. When evaluating the latter, we need to be careful as $\mathrm{Re}[ab] \neq \mathrm{Re}[a]\,\mathrm{Re}[b]$. Writing

$$\vec{E}_j = \tfrac{1}{2}\left(\vec{E}_{0j}\, e^{i\phi_j} + \vec{E}_{0j}^{\,*}\, e^{-i\phi_j} \right),$$

where $\phi_j = \vec{k}_j \cdot \vec{r} - \omega t + \varepsilon_j$, we can evaluate

$$I_{12} = \alpha \langle \vec{E}_1 \cdot \vec{E}_2 \rangle_T$$
$$= \tfrac{1}{2} \langle \vec{E}_{01} \cdot \vec{E}_{02}\, e^{i(\phi_1+\phi_2)} + \vec{E}_{01}^{\,*} \cdot \vec{E}_{02}^{\,*}\, e^{-i(\phi_1+\phi_2)}$$
$$+ \vec{E}_{01} \cdot \vec{E}_{02}^{\,*}\, e^{i(\phi_1-\phi_2)} + \vec{E}_{01}^{\,*} \cdot \vec{E}_{02}\, e^{-i(\phi_1-\phi_2)} \rangle_T .$$

Note that the first two terms are proportional to $e^{-2i\omega t}$ and $e^{+2i\omega t}$, respectively, and will thus vanish, when averaged over times much longer than

the period $\tau = 2\pi/\omega$. The interference term thus reduces to

$$I_{12} = \mathrm{Re}\left[\vec{E}_{01} \cdot \vec{E}_{02}^{\,*} \, e^{i(\vec{k}_1\vec{r} + \varepsilon_1 - \vec{k}_2\vec{r} - \varepsilon_2)} \right] .$$

Note that this expression not only holds for plane waves, but also for spherical wavefronts. Moreover, $\vec{E}_{01} \cdot \vec{E}_{02}^{\,*}$ denotes the inner product of the respective Jones vectors of the individual states. This implies that no interference takes place when the two overlapping waves have orthogonal polarisation states. So while their superposition will change the overall polarisation state, it does not lead to a spatial redistribution of flux density. Redistribution of energy implies, that interference does not alter the overall energy involved in the process, because

$$\int_{\text{all space}} I_{12} \, dV = 0 ,$$

but it changes how the intensity is distributed.

In the case that the polarisation states are the same and the two field amplitudes parallel, $\vec{E}_{01} \cdot \vec{E}_{02}^{\,*} = E_{01} E_{02}$, i.e. we recover the scalar treatment introduced in Lecture 10. In this case, we readily obtain

$$I_{12} = E_{01} E_{02} \cos\delta ,$$

where $\delta = (\vec{k}_1 - \vec{k}_2)\cdot\vec{r} + \varepsilon_1 - \varepsilon_2$ is the total phase difference arising from the combined path lengths and initial phases. Using

$$I_j = \alpha \langle \vec{E}_j^{\,2} \rangle_T = \tfrac{1}{2} E_{0j}^2 \,,$$

we can rewrite the interference term and obtain for the total irradiance

$$I = I_1 + I_2 + 2(I_1 I_2)^{1/2} \cos\delta \,.$$

We can observe

- $\cos\vartheta = 1$ $(\delta = 0, \pm 2\pi, \dots)$: total constructive interference (disturbances in phase, $I = I_{max}$)
- $0 < \cos\vartheta < 1$: constructive interference (disturb. out of phase, $I_1 + I_2 < I < I_{max}$)
- $\cos\vartheta = 0$ $(\delta = \pi/2)$: no interference (dist. 90° out of phase, $I = I_1 + I_2$)
- $0 > \cos\vartheta > -1$: destructive interference (dist. out of phase, $I_1 + I_2 > I > I_{min}$)
- $\cos\vartheta = -1$ $(\delta = \pm\pi, \pm 3\pi, \dots)$: total destructive interference (dist. 180° out of phase $I = I_{min}$)

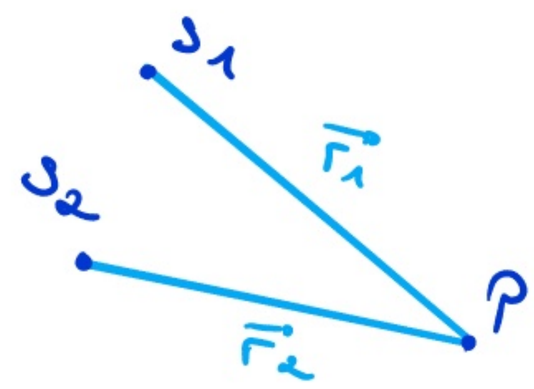
If in addition also the amplitudes are the same and thus $I_j = I_0$, then

$$I = 2 I_0 (1 + \cos\delta) = 4 I_0 \cos^2(\delta/2),$$

with $I_{min} = 0$ and $I_{max} = 4 I_0$.

Example: Considering two point sources of spherical waves with

$$\vec{E}_j(\vec{r}_j, t) = \vec{E}_{0j}(r_j)\, \mathrm{Re}\left[e^{i(k_j r - \omega t + \varepsilon_j)} \right].$$

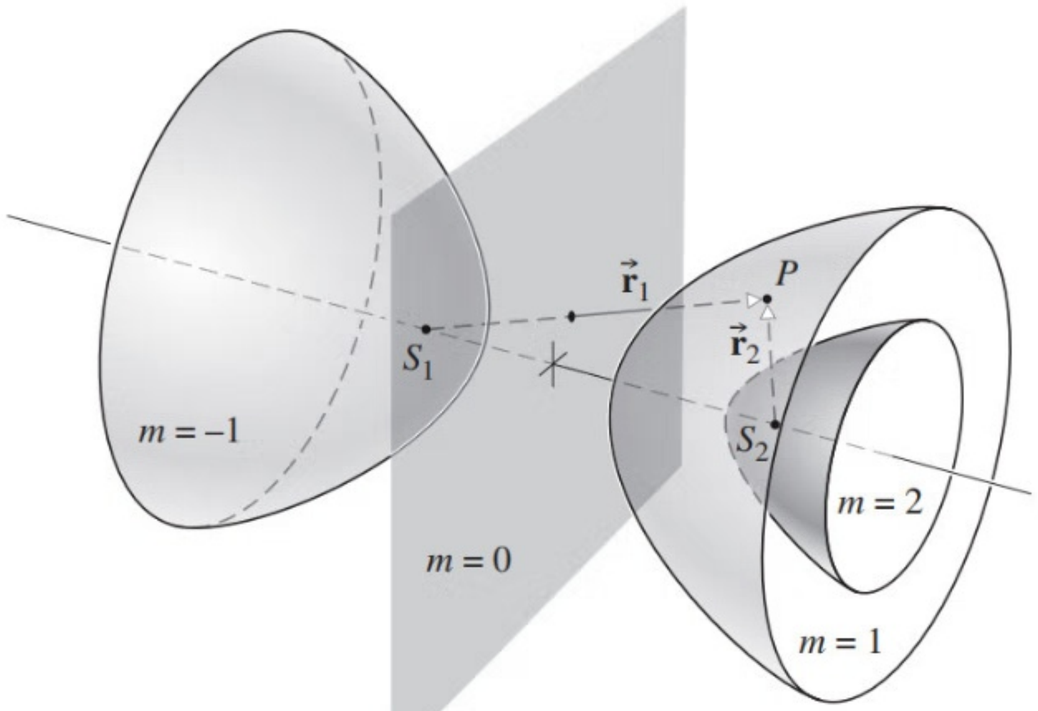


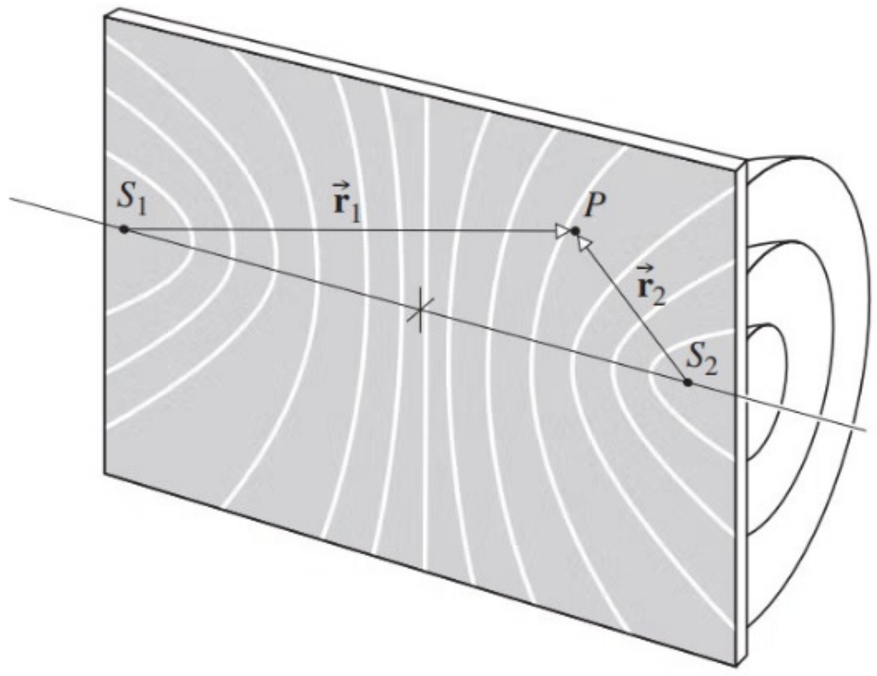
Assuming that the sources have the same polarisation, the intensity at a point P will show interference effects due to $\delta = k(r_1 - r_2) + (\varepsilon_1 - \varepsilon_2)$ (Note: If $\varepsilon_1 - \varepsilon_2$ does not fluctuate the two sources are coherent). We observe intensity maxima when $\delta = 2\pi n$ and minima when $\delta = 2\pi(n + 1/2)$ for $n \in \mathbb{Z}$, or

$$\Rightarrow \text{maxima}: \quad r_1 - r_2 = \lambda\left[n + (\varepsilon_2 - \varepsilon_1)/2\pi\right],$$

$$\rightarrow \text{minima}: \quad r_1 - r_2 = \lambda\left[n + 1/2 + (\varepsilon_2 - \varepsilon_1)/2\pi\right].$$

Both equations describe a series of concentric hyperbolic surfaces, whose foci are the source points S_1, S_2. The alternating dark and bright regions appearing on a screen are called interference fringes.

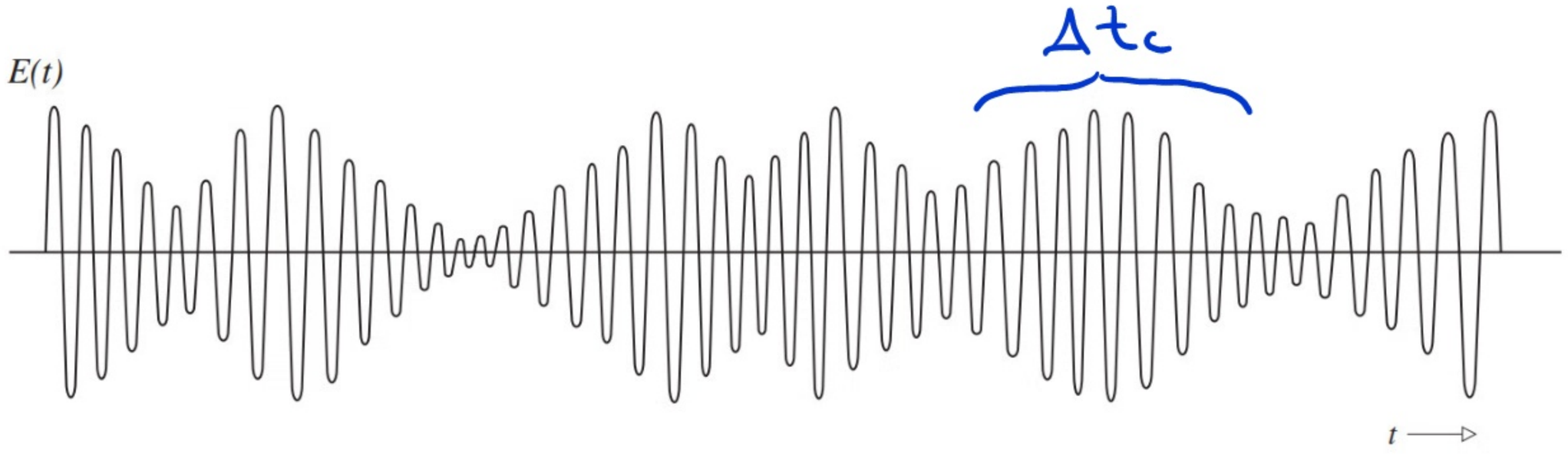




2.) Conditions for interference

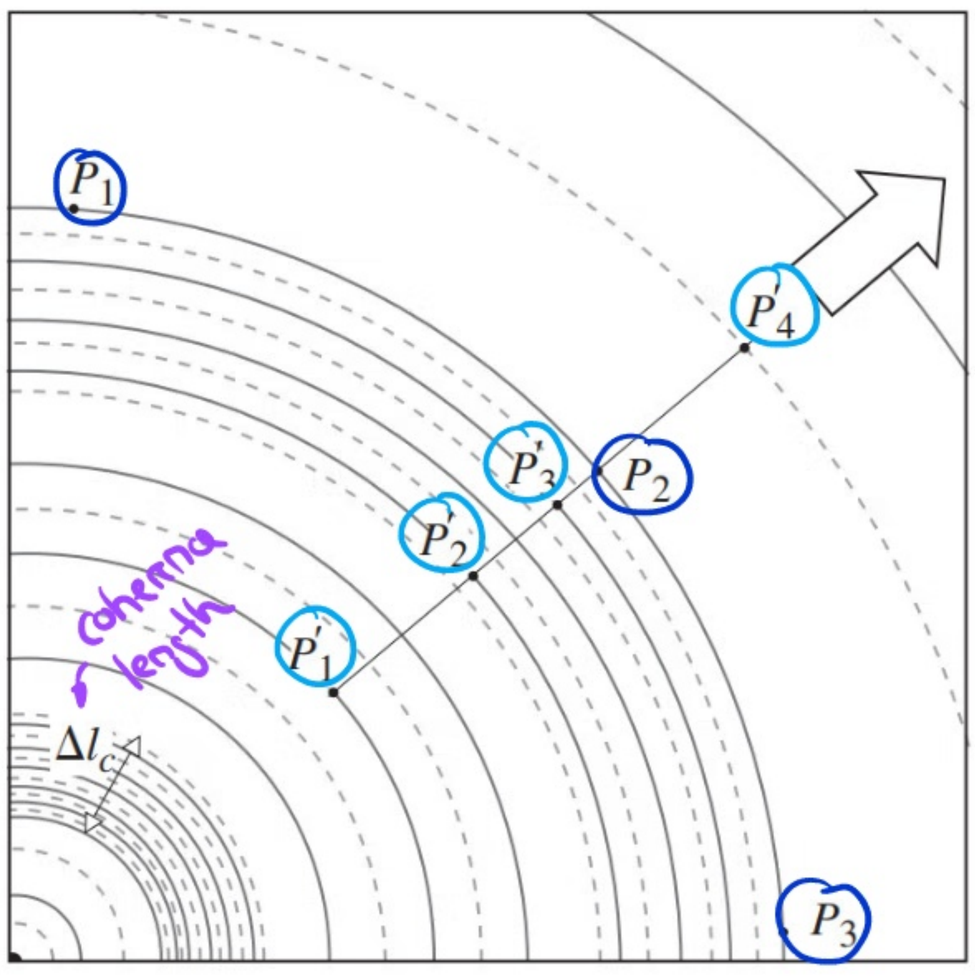
If two beams are to interfere to produce a stable pattern, they first of all must have very similar frequencies. A significant difference in frequency would result in a rapidly varying time-dependent phase difference in I_{12} and time averaging would lead to a vanishing interference term. The clearest pattern will appear if the optical fields have equal amplitudes as complete constructive / destructive interference will occur and yield maximum contrast.

Moreover, to actually observe a pattern the two sources need to emit light waves with a well-defined phase relationship, i.e. they need to be coherent. Remember that conventional quasimonochromatic sources produce light that is a mixture of photon wavetrains (see Lecture 11). The range over which the phase is well defined (and the light wave resembles a sinusoid) is called the coherence time Δt_c, related to the



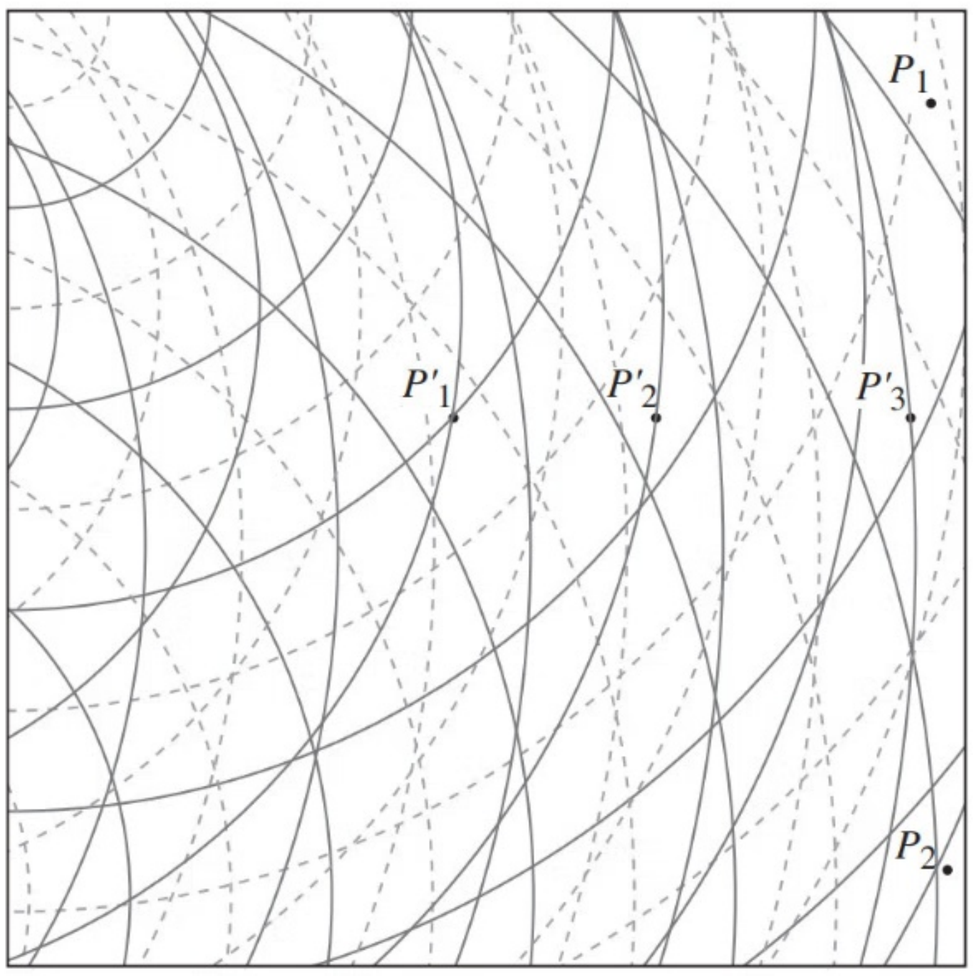
coherence length $\Delta l = c\Delta t_c$. These quantities refer to the temporal coherence of the source (essentially a measure of the source's spectral purity), but not the quantity usually called spatial coherence. Consider a point source such as a small cloud of atoms, that produces temporally incoherent light. Points P_1', P_2', P_3', P_4' illustrated below have no fixed phase relationship (knowing the phase at P_1' doesn't tell us anything about the phases of P_2', P_3', P_4'). However, because the source is small it emits spherical wavefronts and any two points on a wavefront have perfectly correlated phases (knowing phase at P_1 determines phase at P_2 and P_3).

The source exhibits perfect spatial coherence.



temporally incoherent point source exhibiting spatial coherence

Extended sources on the other hand, like the flame of a candle or the sun for example typically produce spatially and temporally incoherent light, because the total amplitude at a specific point is the result of summing out going waves from many point sources with random phases. Thus even though points P_1 and P_2 illustrated below could have the same distance from the surface of the extended source, they have no well-defined relative phase.



temporally and spatially incoherent extended light source

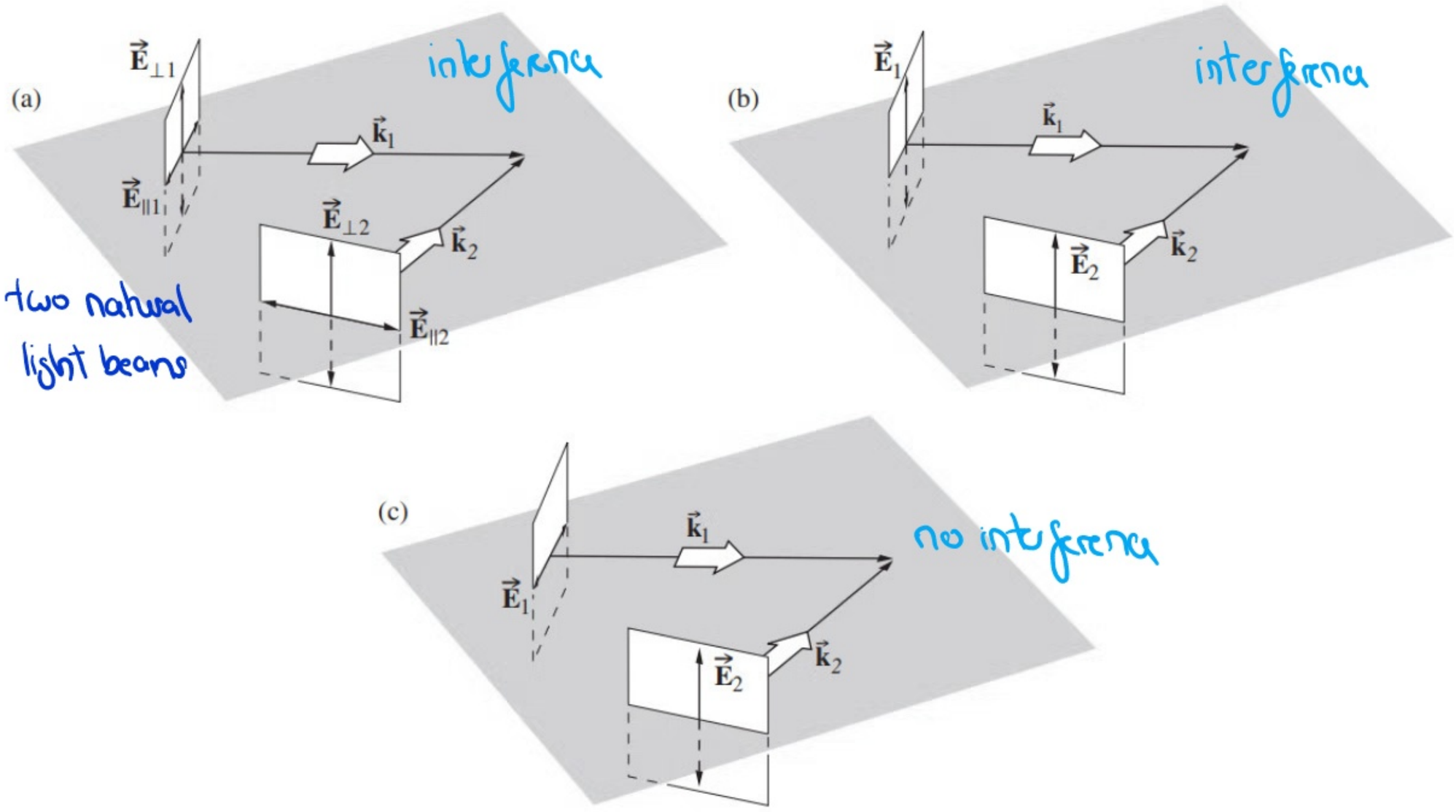
Finally, as we have already seen in the first part of this lecture, the polarisation states of different overlapping waves need to satisfy specific criteria for interference to take place because $I_{12} = Re\left[\vec{E}_{01} \cdot \vec{E}^{*}_{02} e^{\dots}\right]$. These are usually referred to as the Fresnel-Arago Laws:

1.) Orthogonal polarisations do not interfere. (no fringes because $I_{12}=0$)

2.) Unpolarised light will show the same interference pattern as two parallel, coherent polarisations.

3.) The two orthogonal constituent components of unpolarised light cannot interfere even if rotated into alignment.

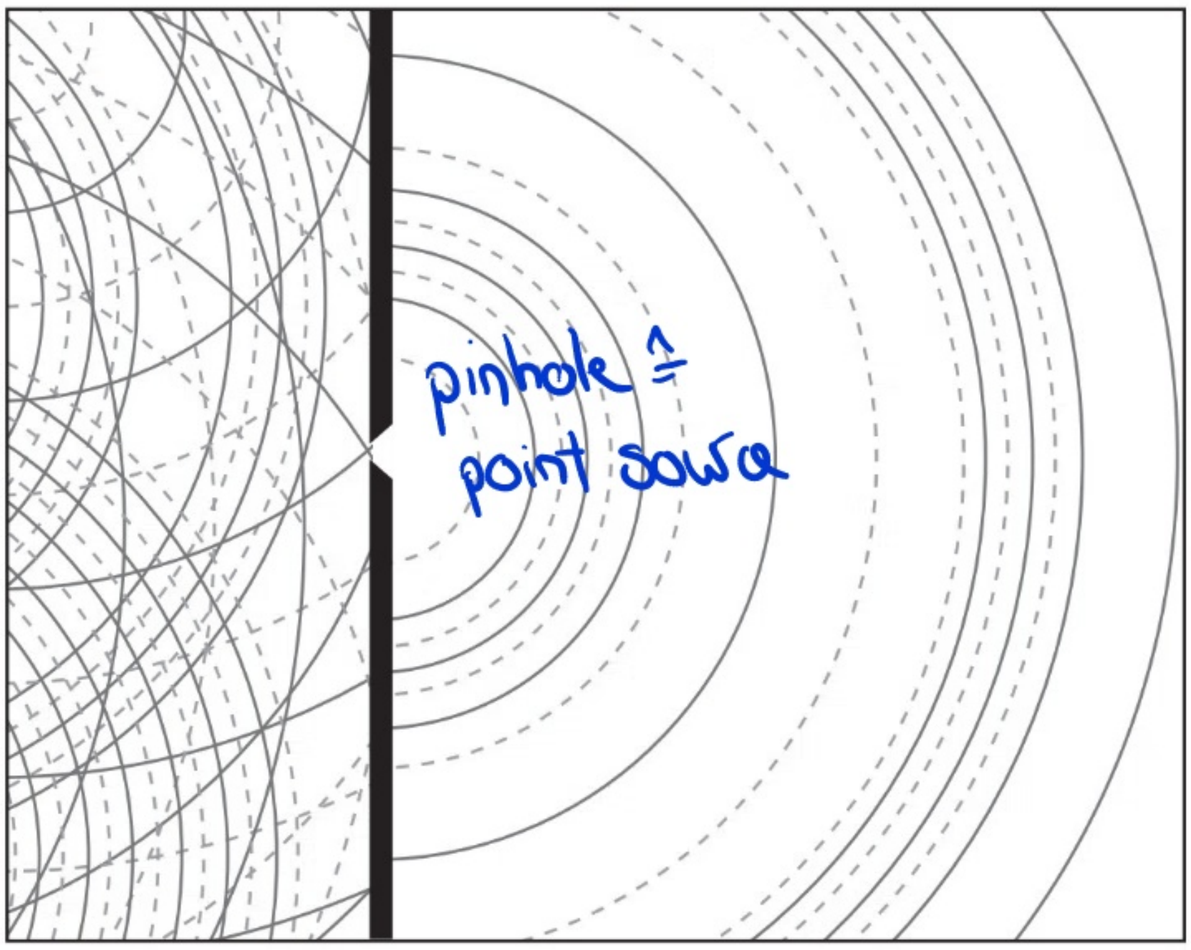
To understand the last two points remember that natural light can be decomposed into two orthogonal fields $\vec{E}_\perp$ and $\vec{E}_\parallel$ as illustrated in (a)

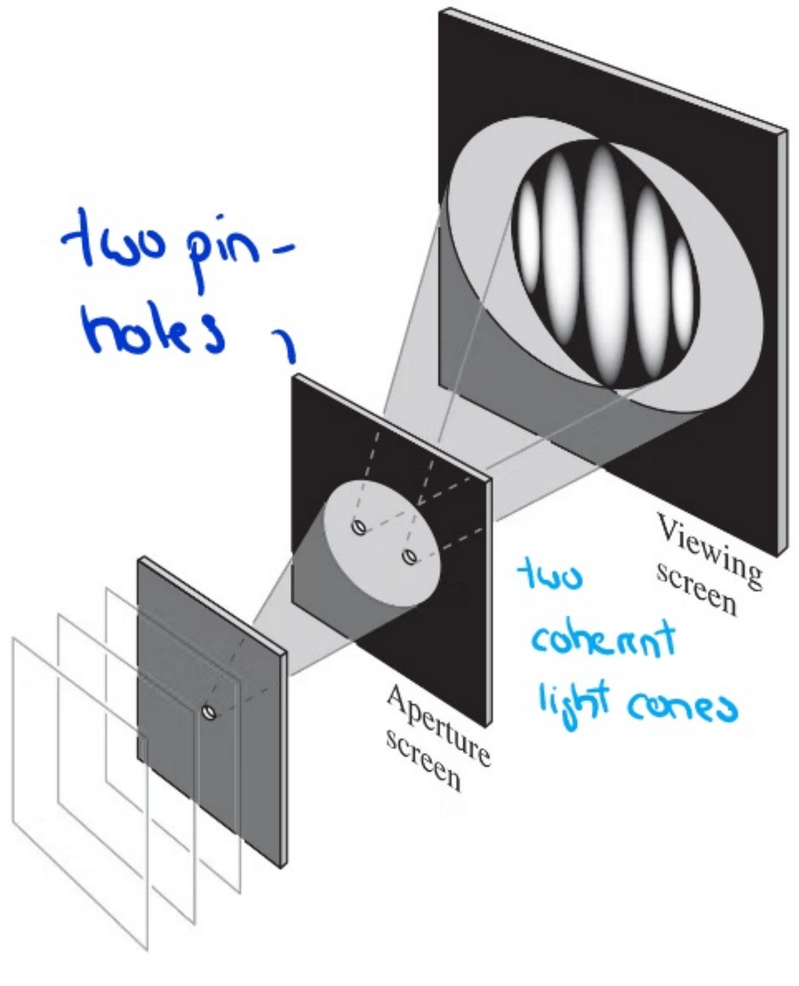


For natural light, the two orthogonal states are mutually incoherent within each of the waves. However, the $\parallel$ ($\perp$) component in beam 1 has a well-defined phase relationship to the $\parallel$ ($\perp$) component in beam 2. Hence interference fringes would be observed in (b) but not in (c), as $\parallel$ and $\perp$ components have an undefined relative phase.

3.) Wavefront-splitting interferometers

The main problem in producing sustained interference is the requirement of coherent sources. While lasers are the obvious answer to this problem it is possible to create coherent sources using a much simpler set-up which was pioneered by Thomas Young over 200 years ago. Motivated by the phenomenon of beats caused by the overlapping of sound waves he created a primary source of spatially coherent light by using a pinhole. Using that beam to subsequently illuminate two apertures he was able to extract two portions of the wavefront from the same source and perform his famous double slit experiment.



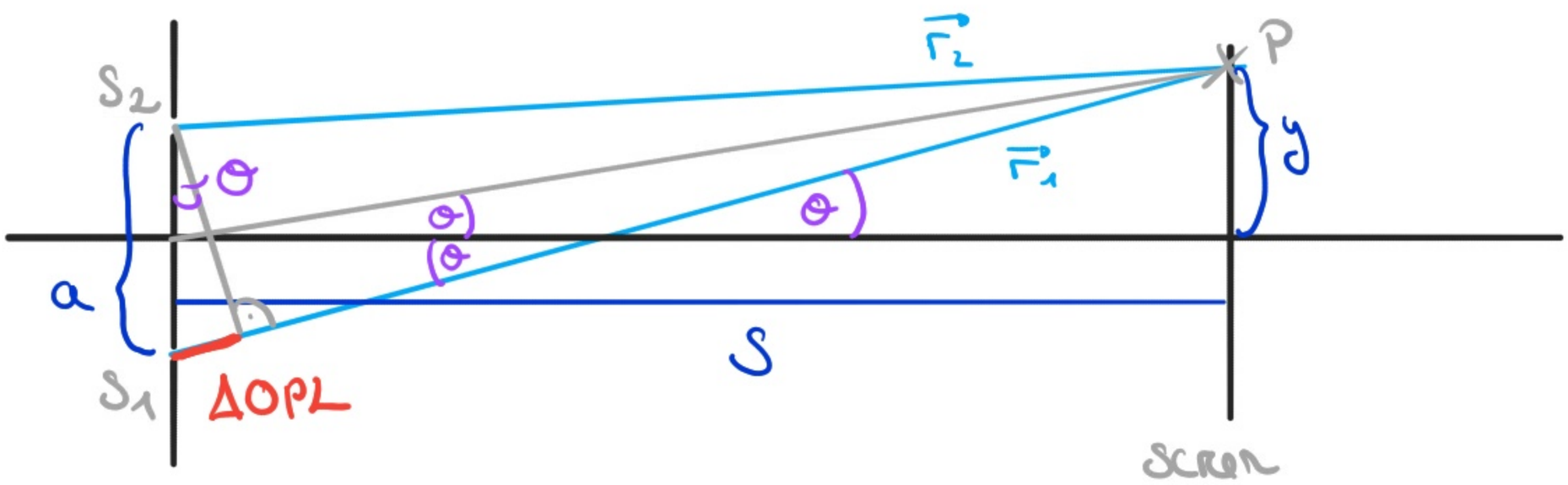


the smaller the pinhole, the more spatially coherent the light

testing interference by using two portions of a wavefront

According to Huygens' principle each point in the slits acts like the source of a spherical/cylindrical wavelet. The light cones exiting the aper-

the screen thus have well-defined (typically identical) relative phases; i.e. we obtain two coherent light sources that can interfere. While Young's original experiment used sunlight plus a pinhole as the primary light source, modern realisations typically directly illuminate the aperture screen with a plane wave from a laser. Considering the case where the screen is positioned at a distance s, much larger than the separation of the two slits a, we can determine the resulting fringe pattern by considering the optical path lengths of the two rays:



small angle approximation

since $s \gg a$

$$\Delta OPL = r_1 - r_2 = a \cdot \sin\vartheta \approx a\vartheta \approx a\, y/s .$$

We thus find for the overall phase assuming $\varepsilon_1 = \varepsilon_2$:

$$\delta = k(r_1 - r_2) \approx \frac{2\pi}{\lambda}\, a\, \frac{y}{s} .$$

$\Rightarrow$ constructive int.: $r_1 - r_2 = \lambda m$, $y_m = \frac{s}{a} m\lambda$,

$\Rightarrow$ destructive int.: $r_1 - r_2 = \lambda(m+1/2)$, $y_m = \frac{s}{a}\lambda(m+1/2)$.

Two bright/dark fringes are thus separated by $\Delta y = y_{m+1} - y_m = \lambda/a\,\lambda$. This dictates that red fringes would be broader than blue fringes due to the wavelength dependence in Δy. Assuming that the two beams further have the same irradiances $I_1 \approx I_2 = I_0$, we obtain for the total intensity

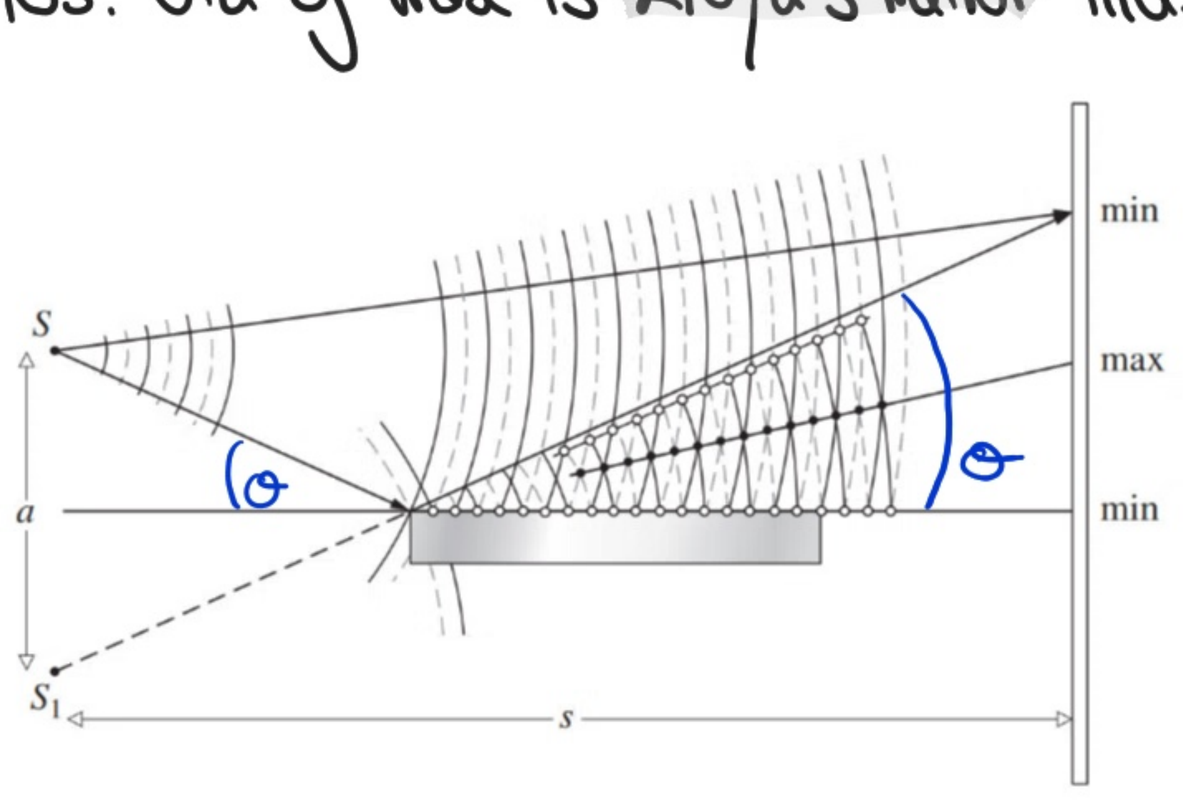
$$I(y) = 4 I_0 \cos^2(\delta/2) = 4 I_0 \cos^2\left(\frac{\pi a y}{\lambda s}\right).$$

The ideal conditions discussed above could in principle suggest

visibility $$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} = 1.$$

However in reality, temporal / spatial incoherence, background light or non-parallel polarisation can result in values $V < 1$.

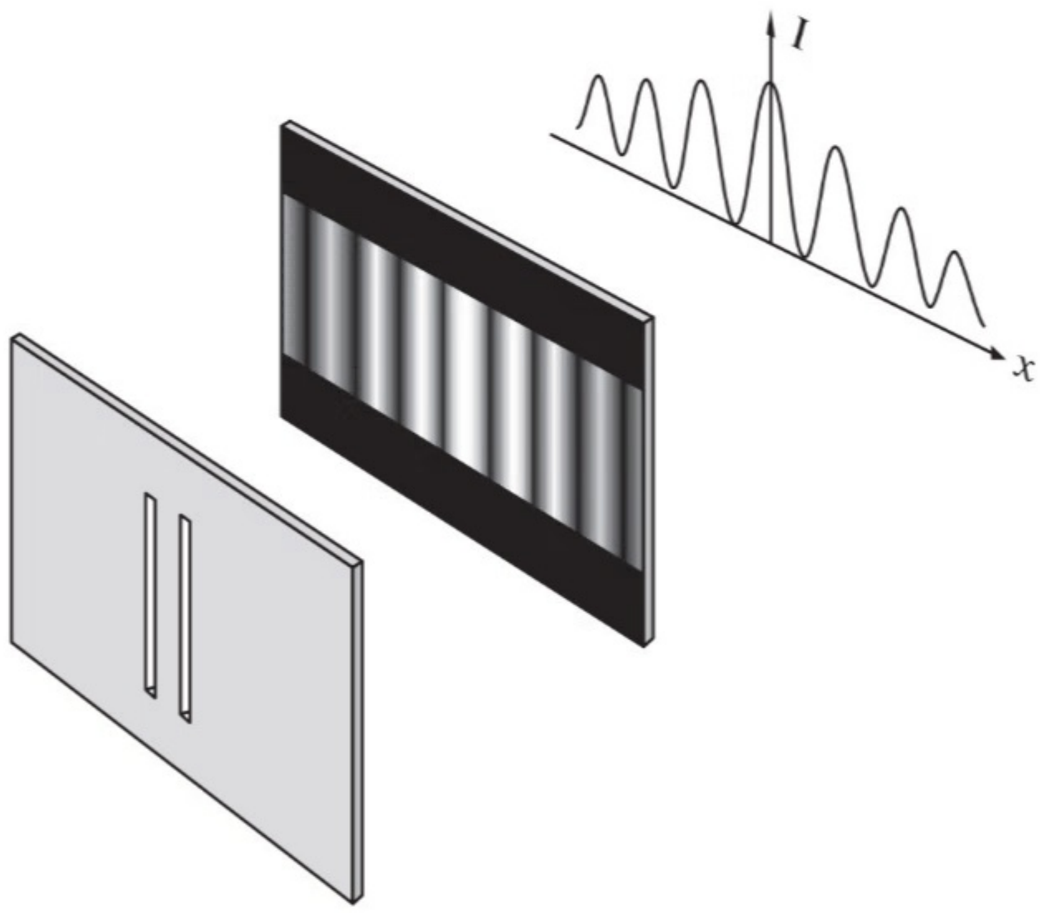
The same physical and mathematical considerations discussed so far also directly apply to a number of other wave-front splitting interferometers. One of these is Lloyd's mirror illustrated below. It consists of a flat piece of dielectric that serves as a mirror, from which a portion of a wave-front coming from a thin slit S is reflected. This



wavefront interfere with the portion proceeding from S directly to the screen. Considering the reflected wavefront to originate from the virtual source, S_2, we can directly translate our previous considerations. The key difference however is that for $s \gg a$, $\theta \ll 1$ (glancing incidence) and the reflected beam undergoes a 180° phase shift. Accounting for this gives $\delta = k(r_1 - r_2) \pm \pi$ and thus the irradiance becomes

$$I(y) = 4 I_0 \cos^2(\delta/2) = 4 I_0 \sin^2\left(\frac{\pi a y}{\lambda s}\right).$$

Note that so far, we have neglected the presence of diffraction in our discussion. We will return to this problem in future lectures but point out here that the light falling through a small aperture will not continuously irradiate the screen but follow a distinct diffraction pattern. Hence the cosine-squared fringe pattern will appear within a single-slit diffraction envelope, i.e. it will be bright in the centre and gradually reduce its intensity towards the edges.

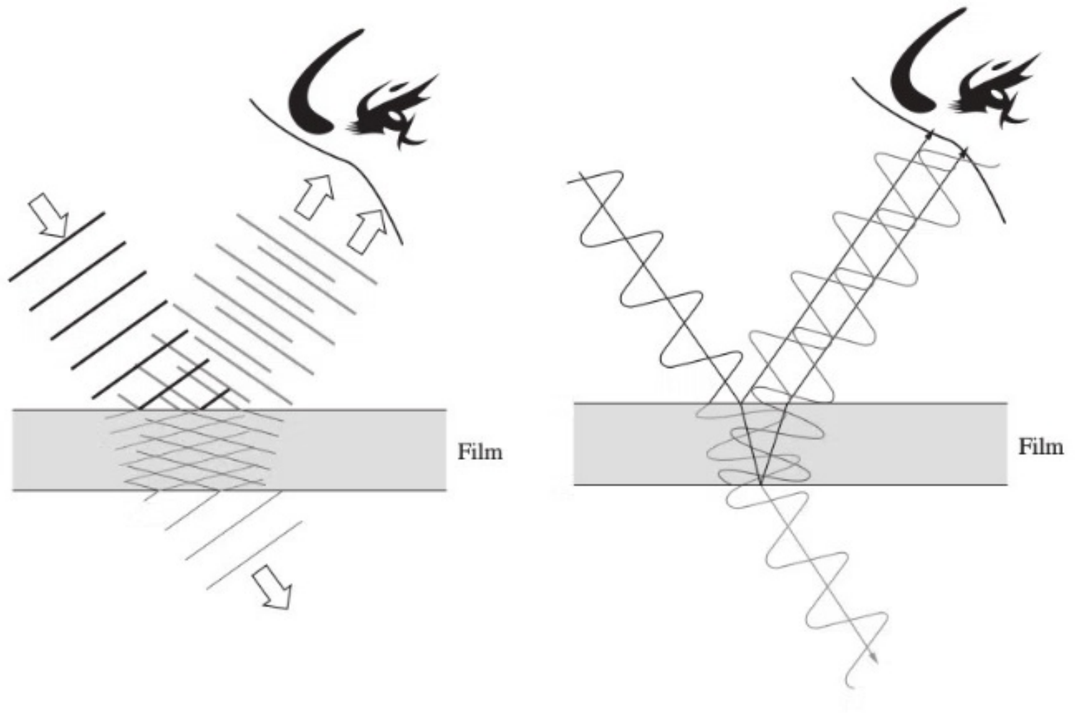


if the aperture slits are very narrow the central bands of their diffraction pattern become very broad and will overlap

Phys 434 - Lecture 17

Amplitude - splitting & Multibeam Interferometry

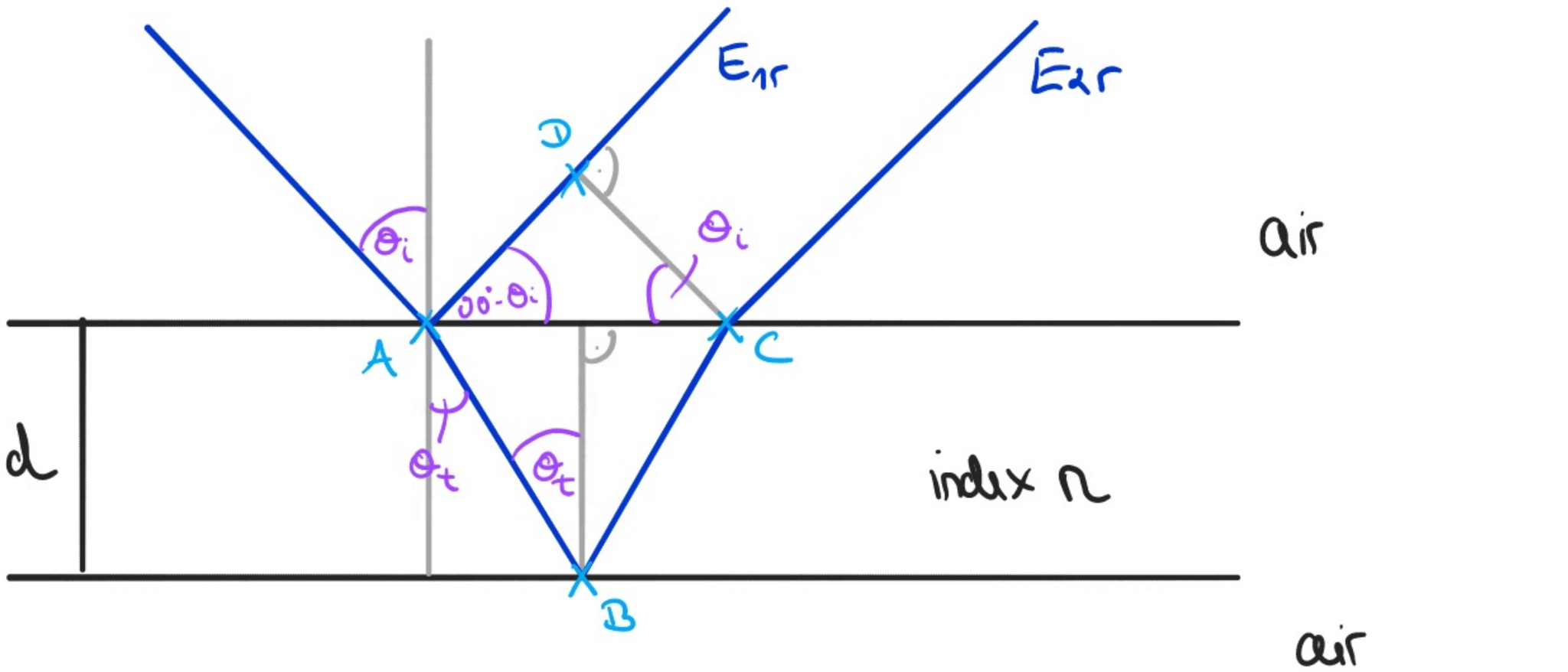
1.) Amplitude - splitting interferometry



An alternative to selecting different sections of a (spatially coherent) wave-front to study interference is to **divide the amplitude** of a source with a partial reflector (e.g. a dielectric interface or a beam splitter) into a **transmitted and reflected component**, which can then be combined to observe interference. For interference to take place, we require the original coherence to be preserved, which implies that the **relative path length** between the two amplitude components **cannot be greater than the coherence length** (otherwise the two components would essentially belong to different wave groups and not have a unique phase-relationship).

Consider the simple case of a **parallel plate of dielectric material** of thickness d. Assume that **absorption is negligible** and only the beams having undergone one reflection, i.e. E_{1r} and E_{2r}, need to be considered. The plate serves as an **amplitude-splitting device** so that E_{1r} and E_{2r} can be con-

sidered as coming from two virtual sources lying behind the film. The reflected rays are leaving the film parallel and can be brought together at a point P when focused from infinity. To analyse the difference in path length between E_{1r} and E_{2r}, we use the following geometry



$$\Delta\,\mathrm{OPL} = n\,\overline{ABC} - \overline{AD} = n\,(\overline{AB}+\overline{BC}) - \overline{AD},$$

where the individual contributions are

$$\overline{AB} = \overline{BC} = d/\cos\vartheta_t,$$

Snell's law:
$\sin\vartheta_i = n \sin\vartheta_t$

$$\overline{AD} = \sin\vartheta_i\,\overline{AC} = n \sin\vartheta_t\, 2\cdot d \tan\vartheta_t,$$

$$\Rightarrow \underline{\Delta\mathrm{OPL}} = 2dn/\cos\vartheta_t - 2dn\sin^2\vartheta_t/\cos\vartheta_t$$

$$= 2dn\,(1-\sin^2\vartheta_t)/\cos\vartheta_t = \underline{2dn\cos\vartheta_t}.$$

The relative phase between the two beams is thus

$$\delta = \vec{k}\cdot(\vec{r}_1 - \vec{r}_2) + (\varepsilon_1 - \varepsilon_2) = k_0 \Delta OPL + (\varepsilon_1 - \varepsilon_2),$$

where k_0 is the free space propagation number. Moreover because of the fact that one beam is internally and the other one externally reflected there will be a relative phase shift of π between the two beams and we have

$$\delta = k_0\, 2dn\cos\vartheta_t \pm \pi \qquad \cos\vartheta_t = \tfrac{1}{n}(n^2 - \sin\vartheta_i)^{1/2}$$

$$= \frac{4\pi}{\lambda_0} d\,(n^2 - \sin^2\vartheta_i)^{1/2} \pm \pi \,.$$

sign doesn't matter, so choose the negative or positive one; changes definition of m

An interference maximum appears when $\delta = 2\pi m$, $m \in \mathbb{Z}$ so

$$2\pi m = \frac{4\pi}{\lambda_0} d(n^2 - \sin^2\vartheta_i)^{1/2} - \pi,$$

$$\rightarrow \quad d\,(n^2 - \sin^2\vartheta_i)^{1/2} = \frac{\lambda_0}{4}(2m+1)\,.$$

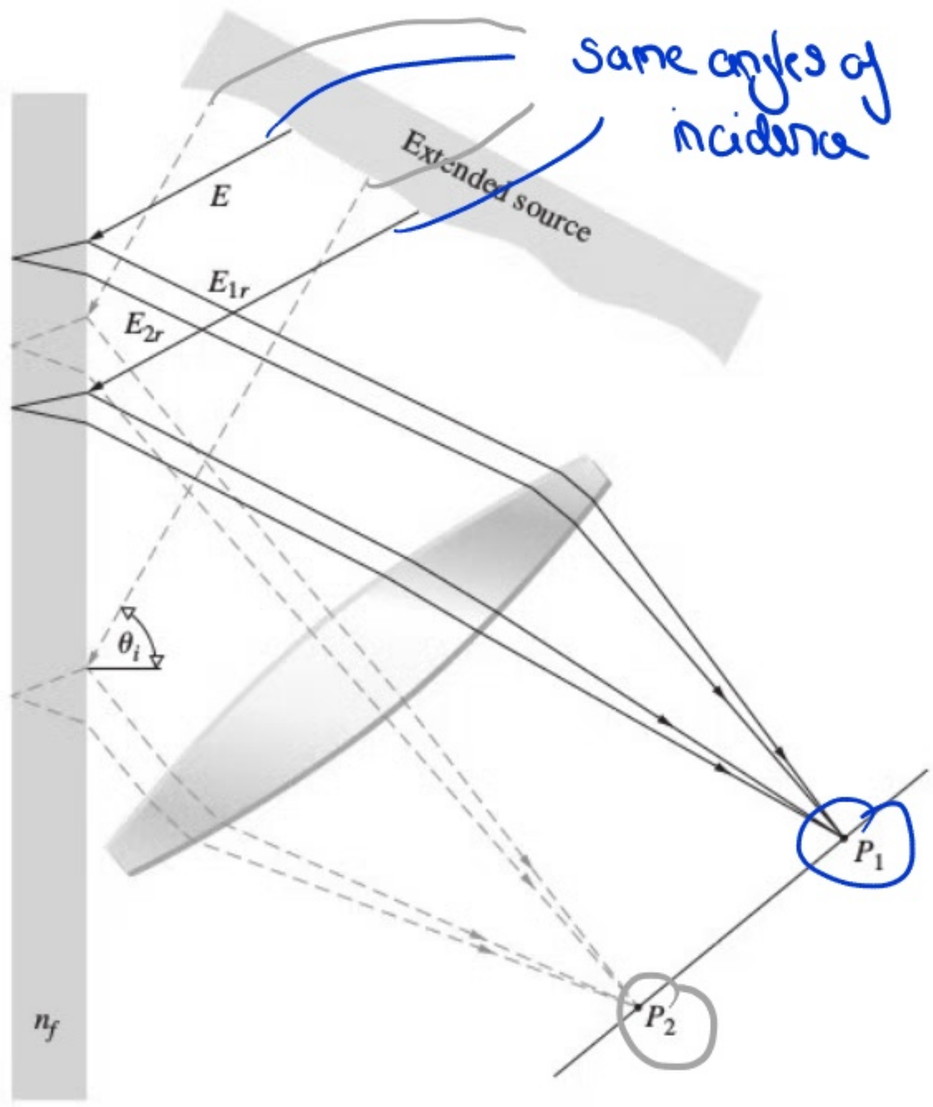
Minima, on the other hand, occur when $\delta = 2\pi(m+1/2)$ so

$$\rightarrow \quad d\,(n^2 - \sin^2\vartheta_i)^{1/2} = \frac{\lambda_0}{4}\,2m\,.$$

Note that we have focused on the interference properties of reflected light. If we were to analyse interference between distinct transmitted light rays (where no relative phase shift is accumulated), we would have to exchange the equations for the occurence of minima and maxima.

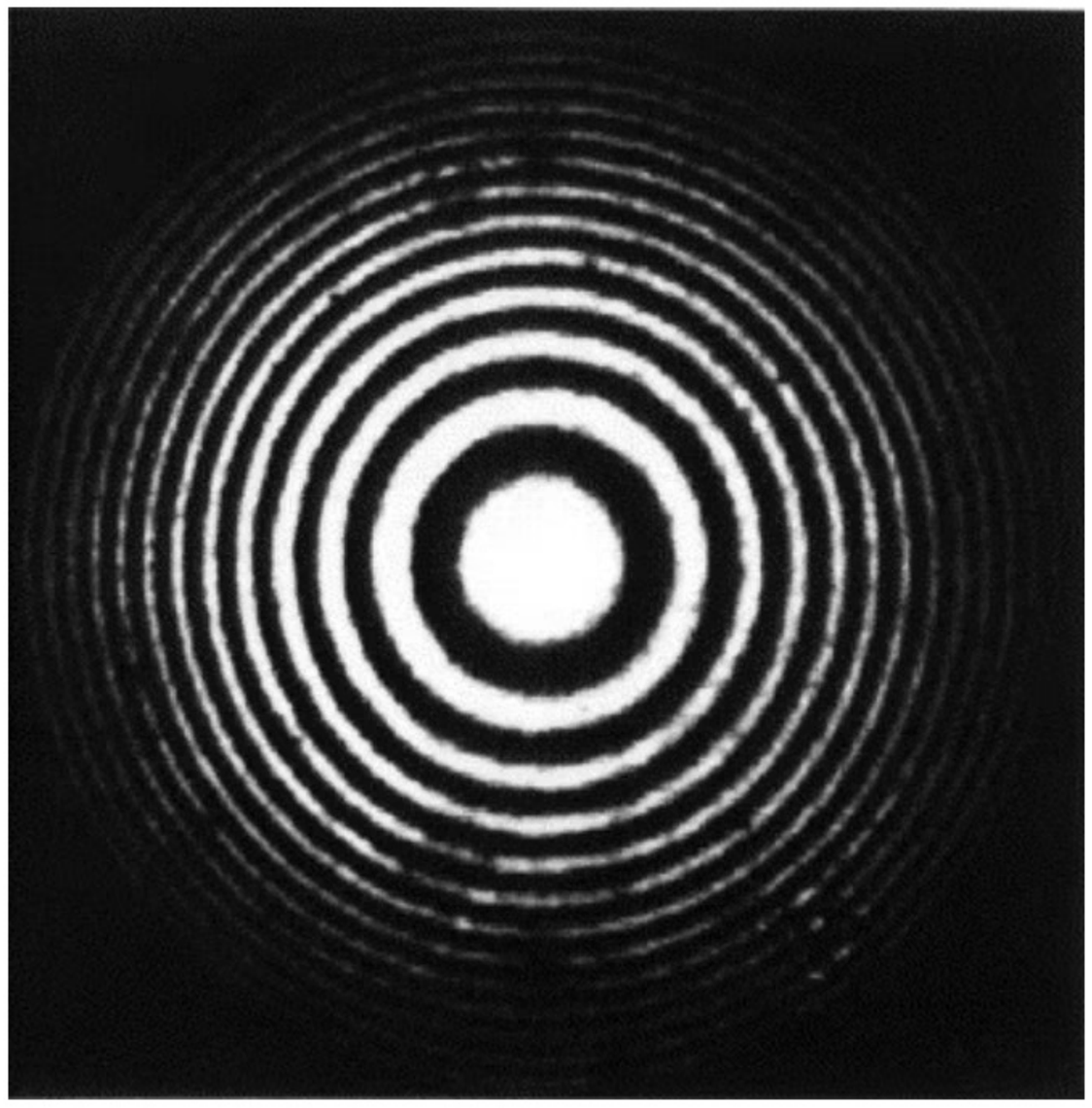
Question: Using a lens to focus the rays from a point source onto a screen, interference can only been studied on a very small section of the screen. How can we use this concept to analyse interference patterns?

As δ depends on Θ_i (the reason why these interference patterns are also referred to as 'fringes of equal inclination') additional intensity maxima and minima can be made visible by using an extended source. Collecting a larger range of incidence angles increases the number of minima/maxima visible on the screen, which are seen as concentric circles, also referred to as Haidinger fringes. Keeping in mind that each source point on the extended source is incoherent with respect to the others, the rays from different source points do not interfere. Their intensities simply add up.



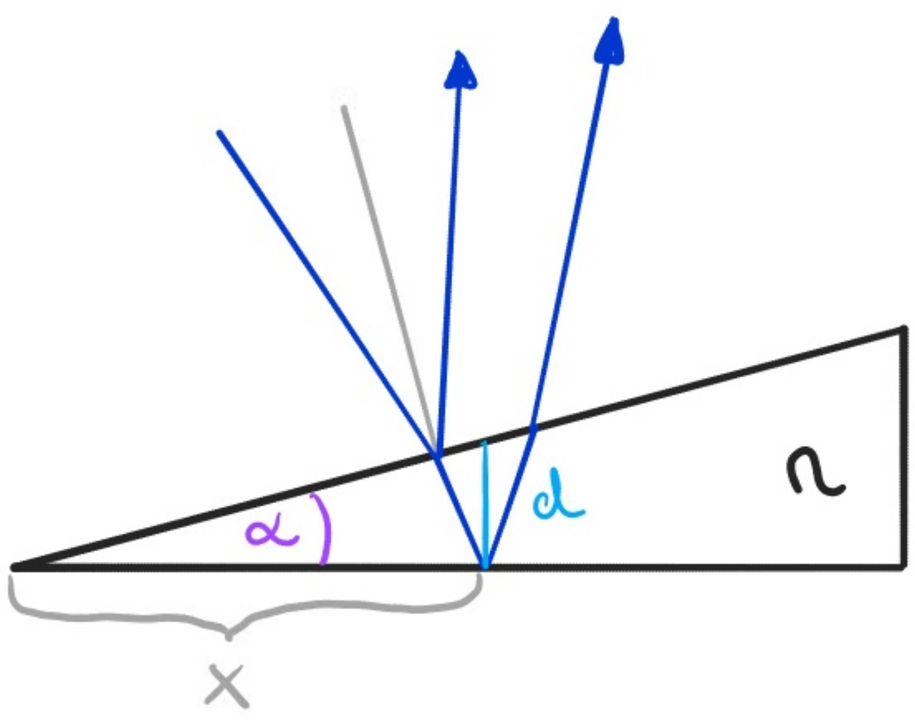
Haidinger fringes / fringes of equal incidence

⇒ concentric circles



Instead of having a thin film of equal width where the incidence angle controls the interference, another class of interference fringes can be created by using surfaces of non uniform separation. In this case, the optical thickness $n \cdot d$ controls the interference properties rather than θ_i and the resulting interference pattern is essentially a topological map of the materials' surface. Consider a wedge-shaped material of index n. When viewed from near normal incidence, the optical path length difference for beams reflected off of the surfaces is given by

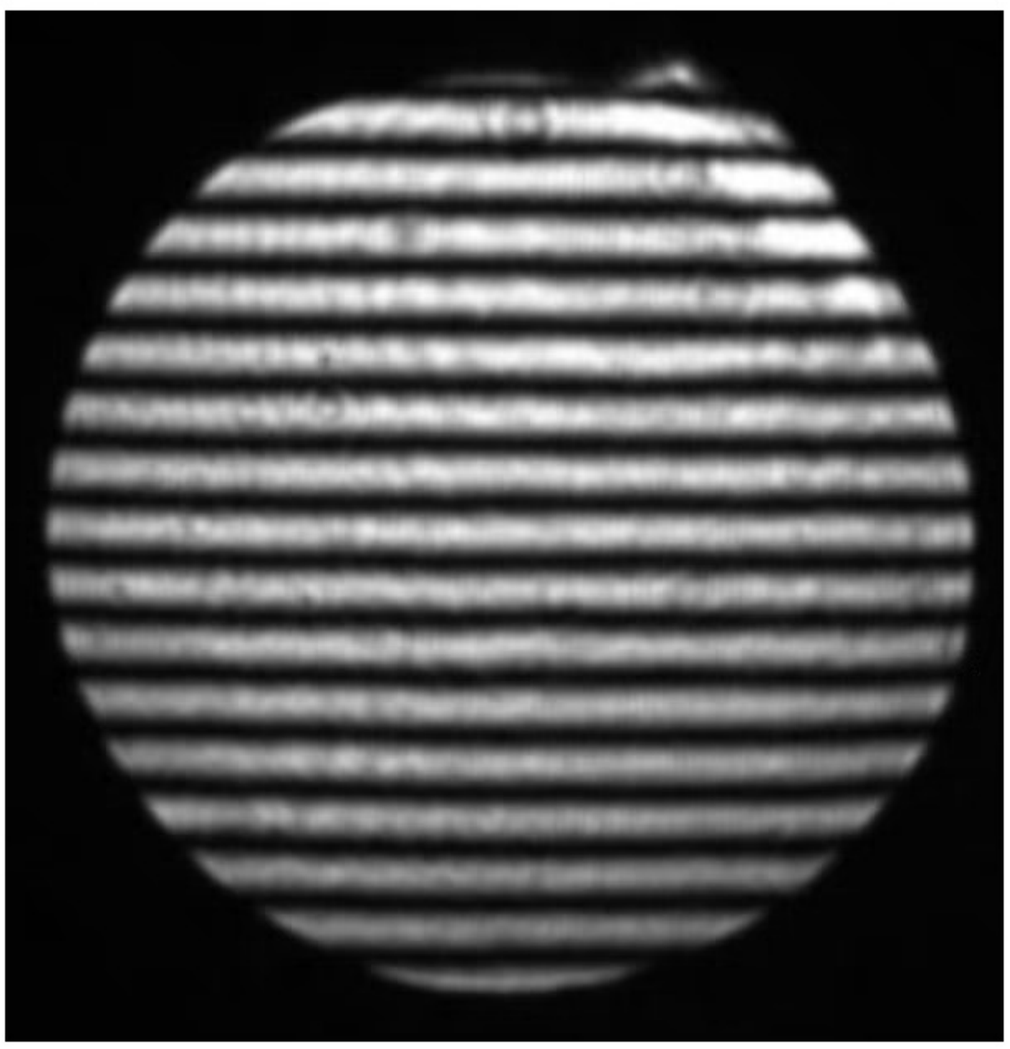
$$\Delta OPL \approx 2nd$$
$$= 2 \cdot n \sin\alpha x$$
$$\approx 2n\alpha x$$



The relative phase is thus given by (accounting for the phase shift)

$$\delta = k_0(r_1 - r_2) - \pi = \frac{4\pi}{\lambda_0} n \alpha x - \pi .$$

When looked at from above the fringe pattern also referred to as 'fringes of equal thickness' (because light from equal d will interfere on the same point on the screen) or 'Fizeau fringes' will look like stripes:



Maxima occur again when $\delta = 2\pi m, m \in \mathbb{Z}$, so that

$$x_m = \frac{\lambda_0}{4\alpha n}(2m+1)$$

The distance between two maxima (or equivalently minima) is:

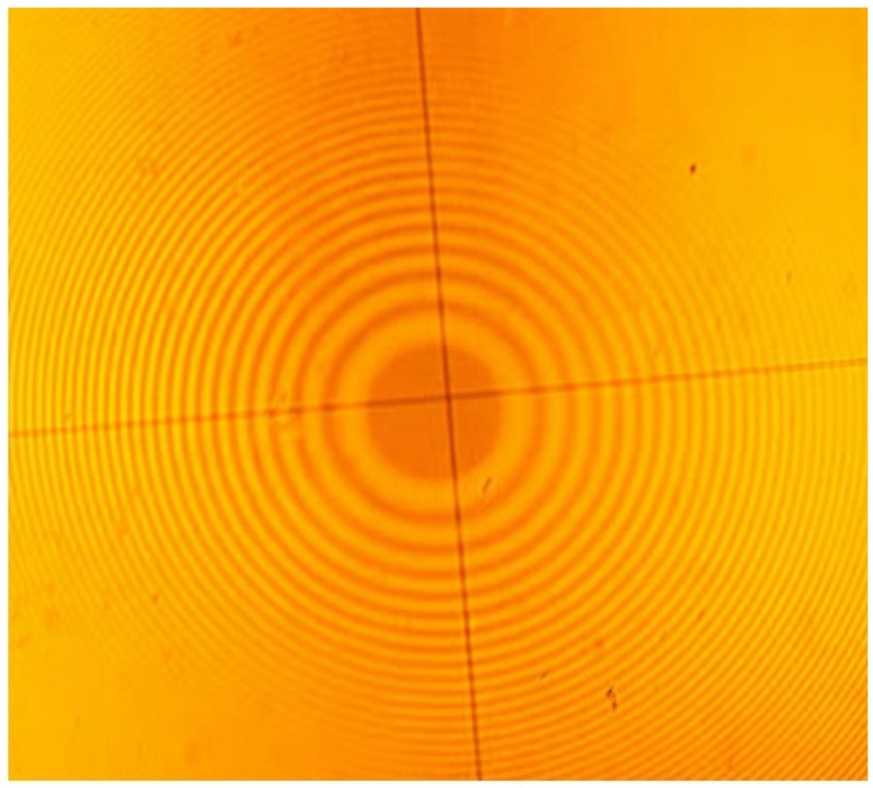
$$\Delta x = x_{m+1} - x_m = \frac{\lambda_0}{4\alpha n} \cdot 2 = \frac{\lambda_0}{2\alpha n} = \frac{\lambda_{medium}}{2\alpha} .$$

This directly implies that the larger the angle α, the smaller the separation between the different fringes. Moreover, the difference in film thickness between two adjacent maxima is $\Delta d = \alpha \Delta x =$

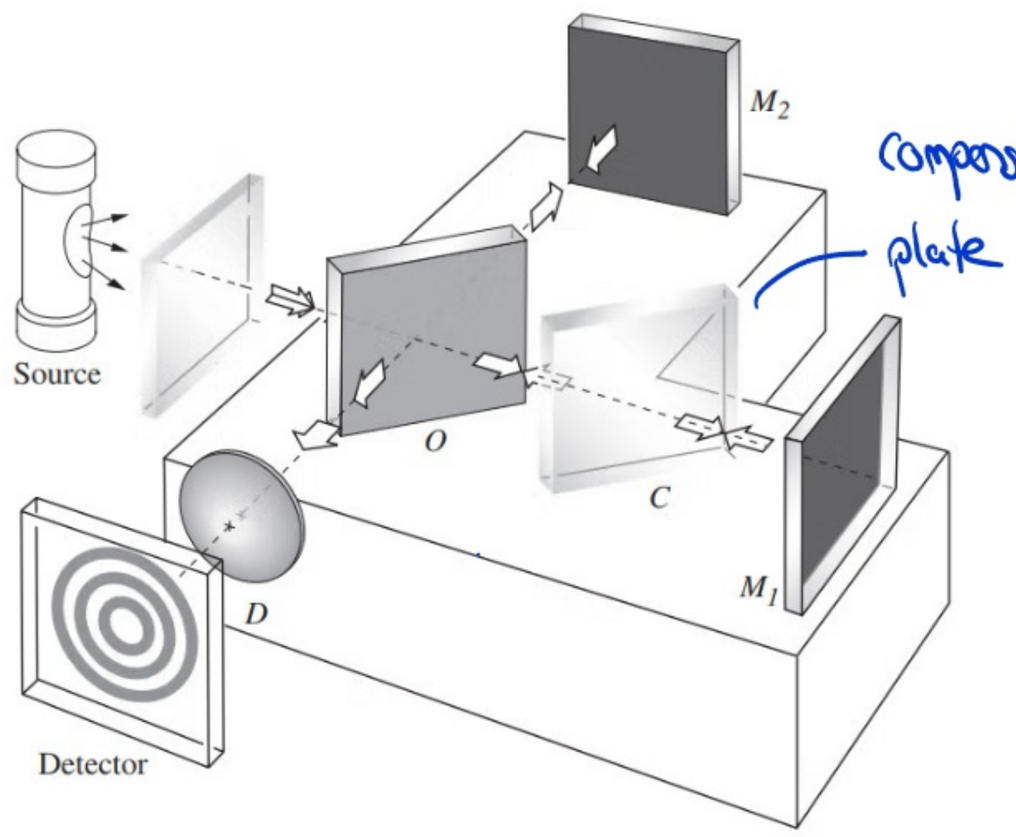
$\lambda_{medium}/2$, so that the difference in path length between two adjacent maxima is exactly λ_{medium} as we would expect. Note that because the interference pattern is so sensitive to the path length traversed in a medium this set-up can essentially be used to measure the width of e.g. a hair by putting it between two glass plates; the width of the hair controls the angle α and hence the distance between two maxima. Knowing the wavelength (in air $\lambda_{medium} \approx \lambda_0$) thus allows one to measure the width of an object separating the two plates at one end.

Question: Is it possible to create spherical patterns using this idea?

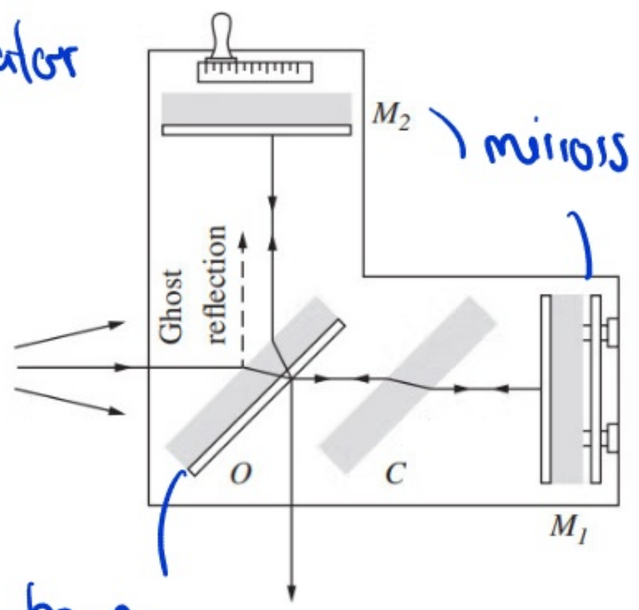
Involving a spherical interface, such as a lens, and combining it with a flat piece of glass ⌓, it is possible to create circular fringes about the point of contact (see PS#5 for more details). These are referred to as Newton rings.



There are an extraordinary number of amplitude-splitting interferometers out there that take advantage of arrangements of beamsplitters and mirrors. The most common one, and historically the most important one, is the Michelson interferometer. The general setup is shown below



compensator plate



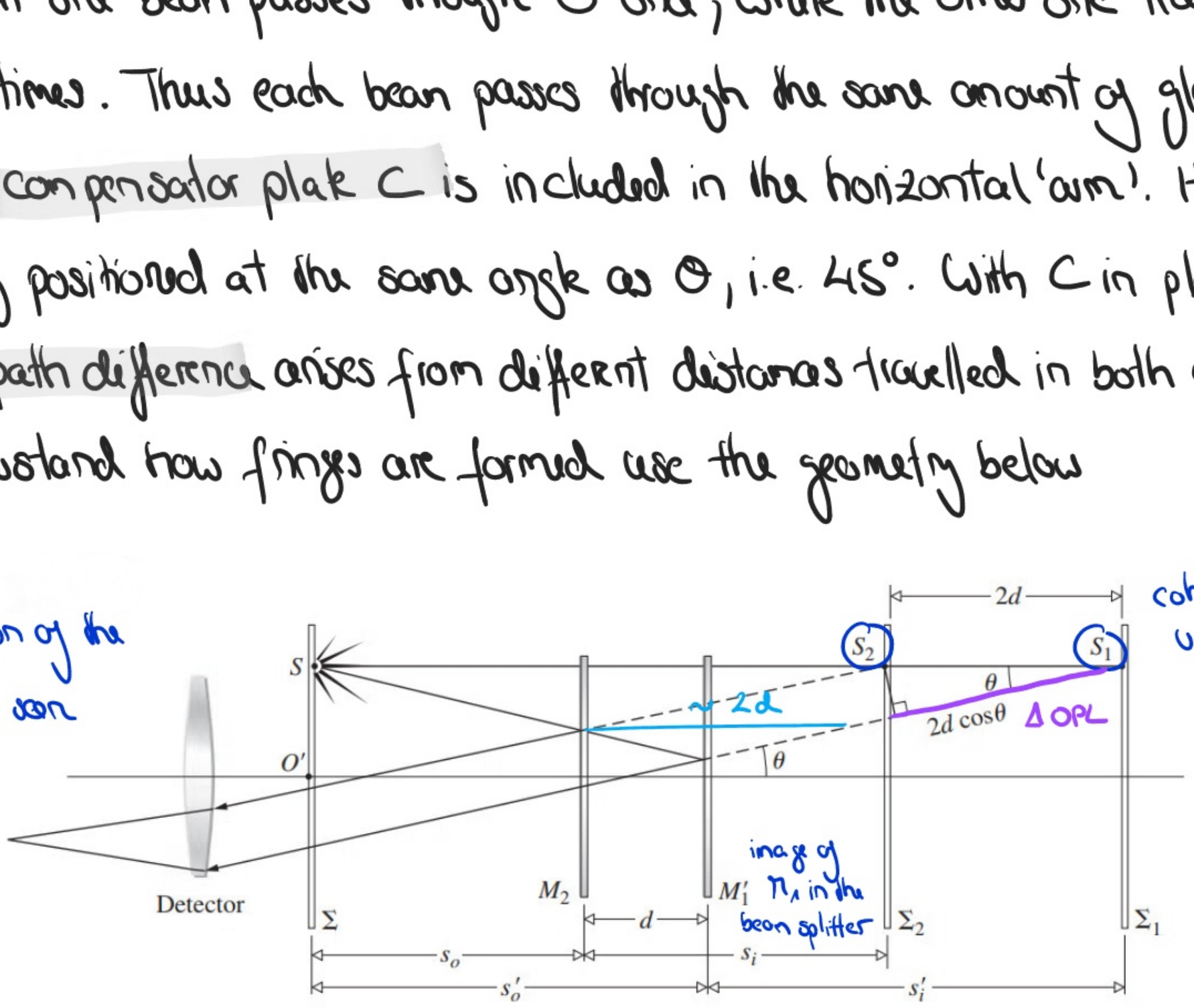
mirrors

beam splitter

Light from a diffuse source (nowadays typically a laser) hits the beam splitter O and is divided into two components, one travelling to the right the other one up. The two beams are reflected by mirrors π_1 and π_2 and return to the beam splitter. Some fractions are transmitted / reflected and deflected toward the detector D, where the waves are united and interfere.

Not that one beam passes through O once, while the other one traverses it three times. Thus each beam passes through the same amount of glass only when a compensator plate C is included in the horizontal 'arm'. It is typically positioned at the same angle as O, i.e. 45°. With C in place any optical path difference arises from different distances travelled in both arms. To understand how fringes are formed use the geometry below

superposition of the sources as seen from detector



coherent virtual sources

Redrawing the different components as if they were in a straight line, we can analyse what happens to light emitted in all directions from a single source – point S, and for an observer at D appears to be coming from points S_1, S_2. The total irradiance should be given by $I = I_1 + I_2 + 2(I_1 I_2)^{1/2} \cos\delta$ (see Lecture 16). From the geometry, we deduce $\Delta OPL \sim 2d\cos\vartheta$.

Moreover, as the OR_2 beam is internally reflected whereas the OR_1 beam is externally reflected, we pick up a 180° phase shift again, so that

$$\delta = k_0 2d\cos\vartheta + \pi = \frac{4\pi}{\lambda_0} d\cos\vartheta + \pi \,.$$

picking + sign for convenience

Destructive interference (rather than constructive interference) happens for

$$\delta = (2m+1)\pi = \frac{4\pi}{\lambda_0} d\cos\vartheta_m + \pi \,,$$

$$\Rightarrow \quad \underline{\underline{\lambda_0 \cdot m = 2d\cos\vartheta_m}} \,, \; m \in \mathbb{Z} \,.$$

If this condition is fulfilled for the point S, any other point on Σ_1 that lies on the circle of radius O'S will satisfy the same condition, so that the observer will see a circular concentric fringe pattern (essentially Haidinger fringes). To see the full pattern a lens is typically required close to the beam splitter. If the source would contain multiple wavelengths each of these would create their own fringe pattern.

This instrument is incredibly useful as it is very sensitive to any changes in length of one of the arms. If we were to decrease d, the rings will move towards the center as ϑ_m decreases because $\cos\vartheta_m$ increases for any given m. The remaining rings broaden until $d=0$ is reached and the central fringe will have spread out over the entire field of view. Due to the reflection phaseshift this corresponds to a minimum (although you might not see this due to imperfections of optical elements). Note that for the geometry above, the concentric circles are fringes of equal inclination since $\vartheta_m = \text{const}$. If the mirrors were slightly tilted, i.e. not parallel any more, we are essentially creating a wedge-shaped air film between them, which causes the appearance of Fizeau fringes. For realistic scenarios both effects are typically combined resulting in all different kinds of shapes.

The interferometer can be used to measure the properties of anything that somehow affects the arm lengths; i.e. you can measure the wavelength of the light source, the index of refraction of a medium placed in one of the arms, you can insert additional polarisers to test the Fresnel-Arago laws, disprove the existence of aether (medium suggested to be the carrier of light waves) or use the it to detect the propagation of gravitational waves through the Earth.

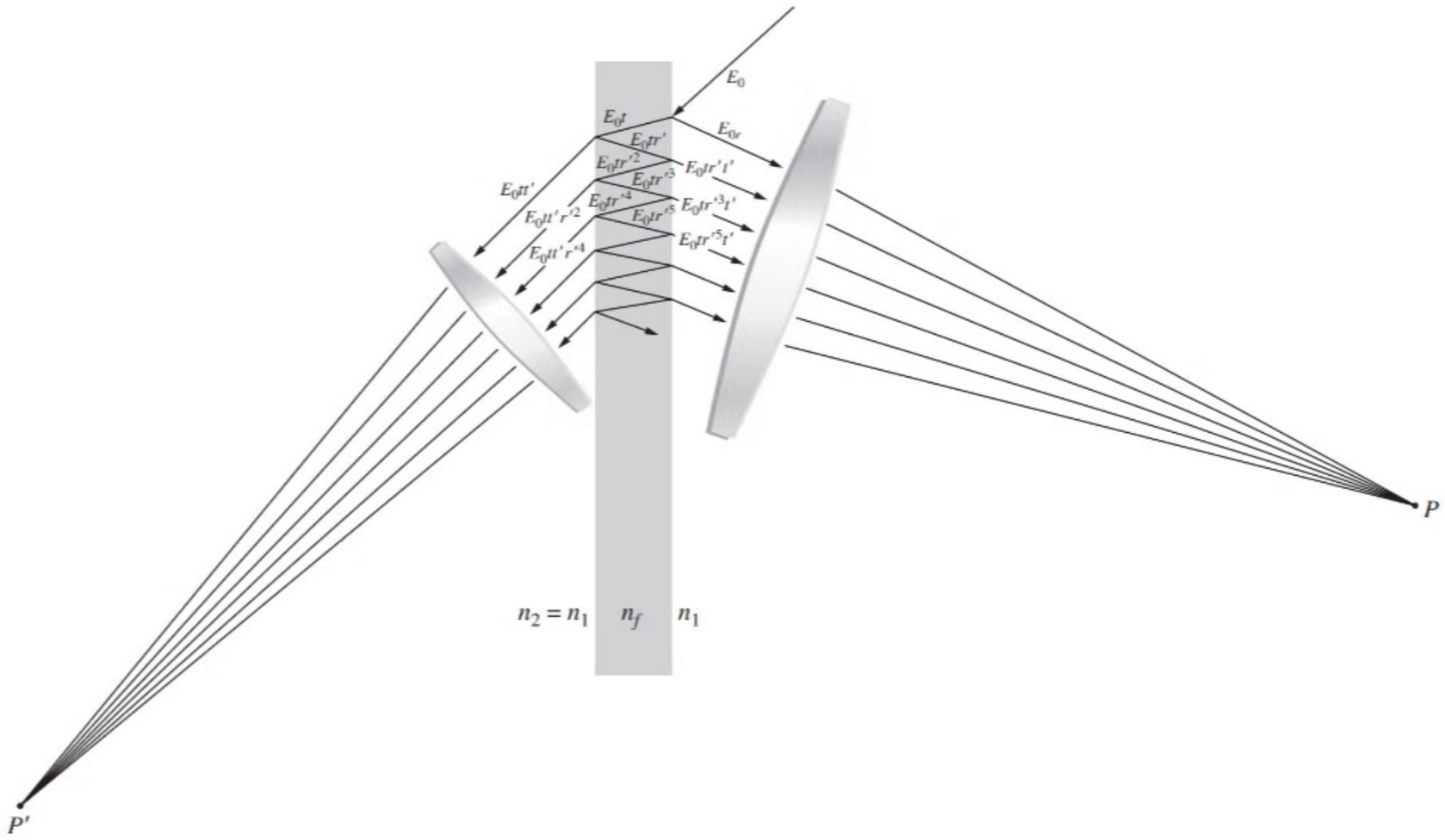
2.) Multi-beam interference

In the examples discussed so far, only two waves were superposed. However, in a lot of scenarios one has to consider a much larger number of mutually coherent waves. Consider a glass plate of width d, that is slightly coated on both sides so that the reflection coefficients r (see Lecture 4) approach unity, generating a large number of multiply internally reflected rays. To study the interference properties of this multi-beam set-up, we have to include all possible numbers of internal reflections, and sum all reflected / transmitted contributions. As the rays remain almost parallel it is sufficient to consider scalar theory and determine

$\Rightarrow$ total reflected amplitude $E_{or} = \sum_i E_{ir}$,

$\Rightarrow$ total transmitted amplitude $E_{ot} = \sum_i E_{it}$.

Using complex notation and remembering that the E field corresponds to the real part, we have $E_{1r} = E_0 r e^{i\omega t}$, $E_{2r} = E_0 t r' t' e^{i\omega t - i\delta}$, ... where E_0 is the amplitude of the incoming wave, r & t the reflection & transmission coefficients for the air to glass interface and r' & t' the coefficients for the glass to air interface (see below). Similarly, $E_{1t} = E_0 t t' e^{i\omega t} e^{-i\delta/2}$, $E_{2t} = E_0 t r'^2 t' e^{i\omega t - i3\delta/2}$, ... Moreover, the difference in path



length between two neighbouring rays is $\Delta OPL = 2n d \cos\theta$ (see page 2), and it is possible to show that $r = -r'$ and $tt' + r^2 = 1$ (see PS#5) when absorption is negligible. Using those relations, the total fields are

$$E_r = E_0 e^{+i\omega t}\left[\frac{r(1-e^{-i\delta})}{1-r^2e^{-i\delta}}\right],$$

$$E_t = E_0 e^{i\omega t}\left[\frac{tt'}{1-r^2e^{-i\delta}}\right].$$ ← ignoring an overall phase factor $e^{-i\delta/2}$, unimportant for I,

which can be translated into the following intensities

$$I_r = \propto \mathrm{Re}[E_r E_r^*] > = < \frac{1}{4}(E_r + E_r^*)^2 >$$

$$= 1/4 < E_r^2 + E_r^{*2} > + 1/2 < E_r E_r^* >$$

$< E_r^2 + E_r^{*2} > \to 0$

$I_0 \equiv \frac{E_0^2}{2}$

$$= I_0 \frac{[2r/(1-r^2)]^2 \sin^2(\delta/2)}{1+[2r/(1-r^2)]^2 \sin^2(\delta/2)},$$

$$I_t = \tfrac{1}{2} E_t \cdot E_t^*$$

$$= I_0 \frac{1}{1 + [2r/(1-r^2)]^2 \sin^2(\delta/2)} \,.$$

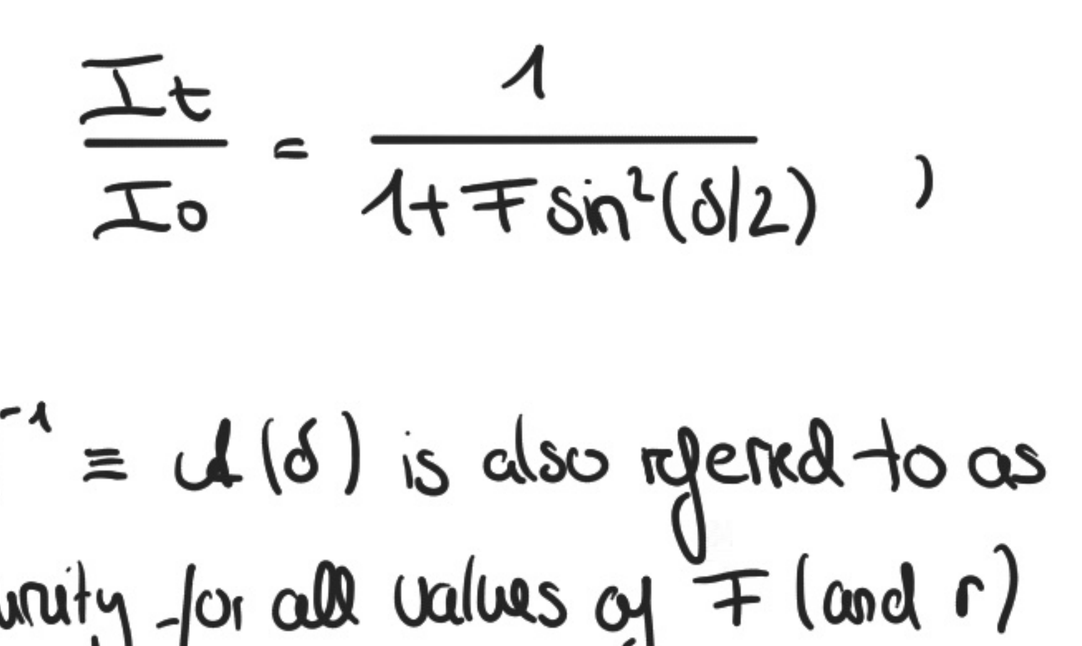
As we would expect if absorption is negligible the intensity of the incident wave is separated into that of the reflected and transmitted beams and we have $I_0 = I_t + I_r$. The form $I_{r,t}$ suggests that we introduce a new quantity, referred to as the coefficient of finesse:

$$F \equiv \left(\frac{2r}{1-r^2}\right)^2 ,$$

so that

$$\frac{I_r}{I_0} = \frac{F \sin^2(\delta/2)}{1 + F \sin^2(\delta/2)} , \qquad \frac{I_t}{I_0} = \frac{1}{1 + F \sin^2(\delta/2)} .$$

where the term $[1 + F \sin^2(\delta/2)]^{-1} \equiv A(\delta)$ is also referred to as the Airy function. $A(\delta)$ is equal to unity for all values of F (and r) if $\delta/2 = m \cdot \pi$, but very small otherwise → for multi beam interference the intensity distribution is no longer sinusoidal but located within small and narrow spikes.



Airy function