Fundamentals of integrated photonic components

Lecture Notes

Prof. Dr. Ileana-Cristina Benea-Chelmus

Hybrid Photonics Laboratory (HYLAB)
Institute for Electro and Microengineering
École Polytechnique Fédérale de Lausanne

December 10, 2025

Contents

Summary

Integrated photonic circuits combine simple blocks that emit, split, combine, couple, confine, disperse, compress, modulate or detect light in a single architecture assembled on a single substrate. A user-defined functionality can then be implemented by combining these components together in a way that they work in parallel or in series.

This course gives an introduction to basic integrated photonic components by considering their fundamental operating principles. The course builds up from basic Maxwell equations in linear, homogeneous, isotropic media. Where not otherwise stated, all materials we consider have these properties. Properties that we derive analytically will then be put at test via finite element calculations performed during exercise class.

We'll start with components that guide light along a user-defined path, such as fibers and waveguides. Then we look at ways to couple light in and out of these components. Once on-chip, we will investigate structures that allow light to pass from one waveguide to another, to split from one waveguide into two waveguides, or to combine several waveguides into a single one. We will then study resonating structures on-chip, such as Fabry-Perot resonators or ring resonators. Finally, we will also add metallic structures in the form of printed antennas to our repertoire.

Fundamental principles that are covered include the theory of total internal reflection at an interface, self consistency condition, multi mode interference effects and light scattering at corrugated gratings.

In addition to a treatment of basic integrated components in the most general form, a variety of material platforms and their applications in both industry and fundamental science will be discussed. We will emphasise each of their unique advantages and disadvantages.

Finally, we will understand how several integrated photonic components on-chip can be assembled together to fulfill an overall functionality that renders them ideal for applications in light manipulation, filtering and routing.

Content and conventions

We summarize below the content of this lecture in more detail.

2.1 Detailed content

1. Introduction
   • Why integrated photonics?
   • Which challenges and opportunities arise from miniaturisation?
   • Which areas benefit currently from integrated photonic circuits?

2. Waveguides
   • How do waveguides work and what are the fundamental modes guided inside a fiber?
   • Which waveguide types exist and why are they useful?
   • Which conditions need to be satisfied for light fields to be guided inside a slab waveguide? What is the spatial profile of guided modes?

3. Fibers
   • How do fibers work and what are the fundamental modes guided inside a fiber?
   • Which fiber types exist and why are they useful?
   • What is the effective mode index, the group index, the group velocity dispersion and the numerical aperture?

4. Couplers
   • How do grating couplers work and what are the parameter that influence their efficiency?
   • How do evanescent couplers work?

5. Multimode interference devices
   • How do multimode interference devices work and what kind of functionalities can one achieve with them?
   • How do x-couplers and y-couplers work?
   • Which parameters matter for the compactness and efficiency of these devices?

6. On-chip resonators
   - What are optical resonators? What is the self-consistency condition?
   - How do the Q-factor, finesse and free spectral range depend on the parameters of a resonator?
   - How are ring resonators different from Fabry-Perot resonators? What are their fundamental modes?
   - What is an overcoupled, undercoupled and critically coupled resonator?

7. Metallic antennas
   - What is a Hertzian dipole?
   - How does the power emitted by a Hertzian dipole depend on angle and frequency?
   - How does the resonance frequency depend on the antenna parameters?
   - What is the far-field pattern of a dipolar antenna?
   - How does the field enhancement depend on the geometry of a printed antenna?

2.2 Conventions

Throughout the script, we will use the conventions summarized in table 2.1.

Please note that different books use different conventions. So do different papers. As a result of this, we will inevitably need to be able to recognize and switch between different conventions.

For example, while Saleh and Teich use the "j" to denote the imaginary part of a complex number and $E_0 e^{j(\omega t - kz)}$ for a plane wave, both Loudon and Boyd's books use i and $E_0 e^{-i(\omega t - kz)}$ (note the sign change!). This leads to a different definition of the Fourier transform (here denoted by $\mathcal{F}\{\}$) for the two cases: $\tilde{a}(\omega) = \mathcal{F}\{a(t)\} = \int_{-\infty}^{\infty} dt a(t) e^{-j\omega t}$ for the first case and $\tilde{a}(\omega) = \mathcal{F}\{a(t)\} = \int_{-\infty}^{\infty} dt a(t) e^{i\omega t}$ for the second case. The inverse Fourier transform is then defined as $a(t) = \mathcal{F}^{-1}\{\tilde{a}(\omega)\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \tilde{a}(\omega) e^{-i\omega t}$.

Note that one can switch from one convention to another by $-j = i$!

A second important convention concerns the coupling and loss rates, g (or κ) and γ, respectively. In some studies they are associated to the amplitude of light (i.e. its electric field), and in other studies to the intensity (i.e. the photon number). In the case where the loss rate γ is defined for the photon number, its associated loss rate for the amplitude is $\gamma/2$.

Convention	Meaning
i	Complex numbers are denoted using i (and not j), as in $a + ib$
$E_0 e^{i(kz-\omega t)}$	Plane waves are denoted in this way (and not $E_0 e^{j(\omega t - kz)}$)
$\mathcal{F}\{a(t)\}$	The Fourier transform is defined as $\tilde{a}(\omega) = \mathcal{F}\{a(t)\} = \int_{-\infty}^{\infty} dt a(t) e^{i\omega t}$
$\mathcal{F}^{-1}\{a(\omega)\}$	The inverse Fourier transform is defined as $a(t) = \mathcal{F}^{-1}\{\tilde{a}(\omega)\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \tilde{a}(\omega) e^{-i\omega t}$
Δ	Frequency detuning from the resonance ω_0, $\Delta = \omega - \omega_0$
κ	Coupling constant for the amplitude (and not the intensity of light), in m^{-1}
α	Attenuation coefficient for the intensity $I(z) = I_0 e^{-\alpha z}$, with $\alpha \sim 2\gamma$, in m^{-1}

Table 2.1: List of most important conventions, their meaning and some formulas used in this lecture.

2.3 Symbols

We summarize in the table 2.2 below the most important physical quantities, their symbols and meaning.

Symbol	Meaning
ω_0	Resonance frequency
$\mathcal{F}$	Finesse
FSR	Free spectral range
GVD	Group velocity dispersion
κ	Coupling constant for the amplitude, in m^{-1}
α	Attenuation coefficient for the intensity $I(z) = I_0 e^{-\alpha z}$, with $\alpha \sim 2\gamma$, in m^{-1}

Table 2.2: List of most important physical quantities, their symbols and some formulas used in this lecture.

General introduction

This chapter introduces the concept of integrated photonic circuits.
Key questions:
• What is an integrated photonic circuit?
• Which role does waveguiding play?
• What is the most generic working principle?
• Which material platforms exist today? What are their particular advantages and disadvantages?
• Which figures of merit are important? Must they be low-loss, broadband, compact?
• Which material properties are especially relevant to passive and active components?
• Where do photonic integrated circuits find applications?
• Overview: importance of photonic integrated circuits, market relevance, use for sensors, detectors, optical computing.
Key concepts: waveguiding, low loss, architecture.

3.1 What is a photonic integrated circuit?

Photonic integrated circuits are microchips that process light to fulfil a desired function. In strong analogy to integrated electronic circuits to which we're much more accustomed, and that process electrons, integrated photonic circuits receive, distribute, combine, synthesise and detect optical signals through combination of basic integrated photonic components into a more complex architecture. At a fundamental level, they are however very different from electronic circuits: unlike electrons, which strongly interact, photons do not interact (virtually) at all. This warrants integrated photonic circuits to be used in scenarios where many different optical signals, for example at different frequencies, or with different polarisations, need to be processed ***in parallel.***

In most cases, integrated photonic circuits aim to emulate and then surpass table-top optical setups, with the particular advantage of compactness. The latter is not only relevant due to the unprecedented possibility to have ultrascaled devices, but also the key to having components that are more energy-efficient, have larger bandwidth, use up less resources, present quantum properties, and are more streamlined than table-top systems can ever be. To achieve this, precise nanoscale patterning plays an important role in the performance and applicability of integrated photonic circuits in various scenarios and under diverse conditions.

In its most simple form, an integrated photonic circuit needs ***to combine active and passive components on a single chip***. Miniaturised light sources, such as ***laser diodes, thermal emitters, or single photon sources*** are currently the subject of intense research. The same applies for miniaturised detectors such as ***avalanche photodiodes, superconducting detectors, or microbolometers***. Other crucial components are for example ***amplifiers***. An in-depth study of these active components - which often rely on fundamental principles of semiconductor physics - goes beyond the scope of the current course. We shall instead study the most basic elements that are typically blended in between an optical source and a detector: ***couplers, waveguides, splitters, resonators, interferometers, and printed contacts and antennas.***

From a standpoint of signal processing, any of these passive components can be described as taking one (or several) input signal(s), and transforming it (them) into one (or several) output signals. As a result of passing through the optical component, the optical signal may change its ***amplitude***, ***phase***, ***propagation properties*** or even ***polarisation***. None of the components we will study in this course can change the ***frequency*** of an optical signal. This requires nonlinear phenomena, for which I refer to other courses.

Owing to their ability to process optical signals, integrated photonic circuits are often employed in scenarios that require:

(a) Combining one or more colors of light in a region of space or time,

(b) Controlled distribution of an optical signal into several channels,

(c) Coherent summation (interference) of several optical signals,

(d) A high light intensity in a particular region of space, e.g. in a resonator,

(e) Filtering and routing of particular wavelengths in a multiplexed fashion.

One of the most appealing properties of integrated photonic circuits is their compatibility with optical fibers. Today's optical fibers are so low-loss in the near-infrared region that optical signals can be sent between

continents with low loss. Light at these wavelengths is therefore an excellent medium to transport information over long distance for long-range communications.

The combination of chip-scale photonic integrated circuits and low-loss optical fibers makes is possible to connect two remote integrated photonic circuits. This is at the core of distributed architectures, largely used for example in data centers.

3.2 Material platforms and their advantages

Over the last 30 years or so, integrated photonic circuits have evolved to the point that they can now address applications in virtually all frequency ranges, as shown in Fig. 3.1.

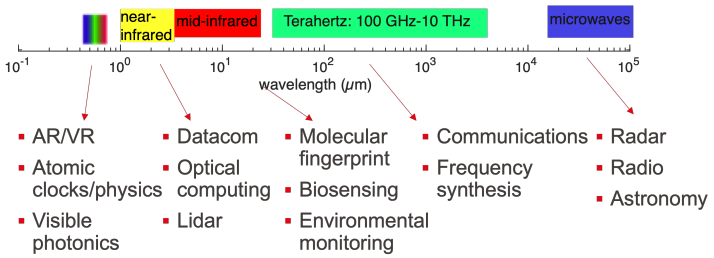

Figure 3.1: Photonic integrated circuits address a variety of applications in all frequency ranges today.

Among the most prominent platforms today are

(a) Silicon-on-insulator,

(b) Silicon nitride,

(c) III-V semiconductors such as indium phosphide,

(d) Lithium niobate,

(e) Germanium.

Silicon-on-insulator is historically one of the most studied platform for integrated photonics. With an indirect bandgap of 1.2 eV, silicon is ideal for applications in the near and low mid-infrared, as it has a transparency window between roughly 1 μm and 7 μm. Its inherent compatibility with CMOS technology makes it very attractive for applications that need electronics and photonics on a single chip. From a photonics perspective, silicon has a high refractive index, which allows low loss waveguides to be

compact and have a sharp bending radius without significant losses. However, its indirect bandgap renders it unsuited for sources and detectors. From a nonlinear optics perspective, silicon has no $\chi^{(2)}$ nonlinearities and a relatively strong two photon absorption.

III-V integrated photonics on InP/InGaAs and InGaAsP is a platform that allows to combine active components in InGaAs (where by tuning the relative concentration of its compounds the bandgap can be tuned between the one of GaAs 1.422 eV and the one of InAs 0.356 eV) with passive components in InP (with a bandgap of 1.34 eV). These are natively well-suited for near-infrared applications.

Silicon nitride is a platform that - interestingly - has a very broad transparency window, and with its bandgap around 5 eV, it allows to address in one single platform near-UV, visible, near-infrared and mid-infrared photonics. Over the last years, this platform has emerged as one of the few leading ones for nonlinear optics, due to its high power handling capabilities, very low propagation loss and low index contrast. Silicon nitride is mostly used for its $\chi^{(3)}$ nonlinearity as it does not possess $\chi^{(2)}$ nonlinearities.

Lithium niobate, leading amongst the few platforms that feature a non-zero $\chi^{(2)}$ nonlinearity, is relatively recent compared to the platforms discussed above. Its high power handling, large refractive index and broad transparency window renders it uniquely suited for applications that combine $\chi^{(2)}$ and $\chi^{(3)}$ nonlinearities on a single chip.

Germanium is typically the go-to material platform for mid-infrared photonics due to its transparency window in this frequency range. It is currently subject of intense research.

Going even beyond the platforms discussed above, photonic integrated circuits have witnessed an extreme diversification. Emerging material platforms are barium titanate, lithium tantalate, silicon carbide, aluminium nitride, or polymers. This access into novel materials has been largely fuelled by developments in nanofabrication and nanotechnologies that nowadays allow to grow high-quality films and pattern these into waveguides which can be only a few 100s of nanometers in width.

On the other hand, diversification also results from possibilities to combine two otherwise distinct material platforms in a single one, using either ***homogeneous*** or ***heterogeneous*** nanofabrication techniques.

3.3 Figures of merit of integrated photonic circuits

Photonic integrated circuits are generally classified according to several major characteristics:

- **Losses** refers to the amount of light lost into channels that do not contribute to the overall purpose of a photonic circuit. In a lossless photonic integrated circuit, energy can be transported along the entire circuit without any decay. Instead, in a lossy photonic circuit, the amount of energy that is being transported becomes less and less, the deeper light penetrates into the circuit.
- **Bandwidth** refers to the ability of a photonic integrated circuit to operate on many different wavelengths. This property is particularly relevant to applications that requires a broad spectral coverage, as for example in spectroscopy or non-linear optics.
- **Passive or active** refers to whether the photonic integrated circuit only performs linear operations on light (i.e. splits, combines and transfers it), or also nonlinear operations (amplification, detection via absorption, generation via a laser diode).
- **Hybrid** refers to whether the photonic circuits is made from one single material system, or if it combines several materials systems, e.g. through heterogeneous integration.
- **Nonlinear** refers to whether the photonic circuit exhibits any optical nonlinearities, such as $\chi^{(2)}$ or $\chi^{(3)}$ nonlinearities. These are essential in frequency conversion, such as second harmonic generation, frequency comb formation, generation of entangled photon pairs, or supercontinuum generation.
- **Footprint** refers to the geometric size of a photonic integrated circuit. This property is particularly important in scenarios where a complex architecture needs to be assembled together to realize a certain functionality.

3.4 Applications and relevance of photonic integrated circuits

Photonic integrated circuits can enable many different applications that we will discuss in detail as we advance through the various components in class:

(a) Wavelength multiplexers and demultiplexers,

(b) Frequency filters and synthesis,

(c) Routing,

(d) Sensing and ranging,

(e) Optical trapping.

Introduction to wave propagation

This chapter recalls the most important properties of light propagation in linear, homogeneous and isotropic media.
Key questions:
• What are Maxwell's equations?
• What is the Helmholtz equation?
• What is dispersion?
• Which are the boundary conditions at a direlectric interface? How can we derive them? How can we derive Snell's law, the law of refraction and the law of reflection from them?
• What are TE and TM modes? How do their reflection and transmission at a dielectric interface depend on the angle of incidence and refractive indices?
• What is total internal reflection and what are its properties?
Key concepts: linear, isotropic, homogeneous, invariant media
Key equations: Maxwell's equations
Literature: Saleh and Teich, chapter 5.1-5.5

4.1 Revision of wave propagation in optical media

Light propagates as electromagnetic waves that oscillate with frequency f and angular frequency $\omega = 2\pi f$ at the speed of light $c_0 = 299792458\ m/s$. Inside a bulk material, its frequency remains unchanged, however its travel speed is $c = \frac{c_0}{n}$ and its wavelength is $\lambda = \frac{\lambda_0}{n}$, with n the real refractive index of the medium.

In the absence of free carriers ($\rho = 0$) and currents ($\mathbf{J} = 0$), Maxwell's

equations take the following form:

$$\nabla \cdot \mathbf{D}(\mathbf{r}, t) = 0 \tag{4.1.1}$$

$$\nabla \times \mathbf{E}(\mathbf{r}, t) = -\frac{\partial \mathbf{B}(\mathbf{r}, t)}{\partial t} \tag{4.1.2}$$

$$\nabla \cdot \mathbf{B}(\mathbf{r}, t) = 0 \tag{4.1.3}$$

$$\nabla \times \mathbf{H}(\mathbf{r}, t) = \frac{\partial \mathbf{D}(\mathbf{r}, t)}{\partial t} \tag{4.1.4}$$

where $\mathbf{r} = (x, y, z)$ a vector in three-dimensional space. $\mathbf{B}$ the magnetic flux density and $\mathbf{H}$ the magnetic field. The two are related through the relationship $\mathbf{B}(\mathbf{r}, t) = \mu_0 \mathbf{H}(\mathbf{r}, t) + \mu_0 \mathbf{M}(\mathbf{r}, t)$. $\mathbf{M}(\mathbf{r}, t)$ is the magnetization. $\mathbf{D}$ is the displacement field and $\mathbf{E}$ is the electric field, and the two are related in homogenous, isotropic media by $\mathbf{D}(\mathbf{r}, t) = \epsilon_0 \mathbf{E}(\mathbf{r}, t) + \mathbf{P}(\mathbf{r}, t)$, with $\mathbf{P}$ the polarization field.

In many cases, it is more convenient to deal with these fields in the frequency domain instead of the temporal domain. Given that any ***monochromatic*** (= single frequency) electromagnetic field $\mathbf{F}(\mathbf{r}, t)$ is described by

$$\mathbf{F}(\mathbf{r}, t) = \frac{1}{2\pi} \Re\{\tilde{\mathbf{F}}(\mathbf{r}, \omega) e^{-i\omega t}\} \tag{4.1.5}$$

with $\tilde{\mathbf{F}}$ being the spectral amplitude of the electromagnetic wave at frequency ω. We can now pass into the frequency domain by taking the Fourier transform $\mathcal{F}\{\mathbf{F}(\mathbf{r}, t)\}$ of all fields. We further use the differentiation identity of the Fourier transform $\mathcal{F}\{\frac{\partial \mathbf{F}(\mathbf{r}, t)}{\partial t}\} = -i\omega \mathcal{F}\{\mathbf{F}(\mathbf{r}, t)\}$ and get to the frequency domain set of Maxwell's equations

$$\nabla \cdot \tilde{\mathbf{D}}(\mathbf{r}, \omega) = 0 \tag{4.1.6}$$

$$\nabla \times \tilde{\mathbf{E}}(\mathbf{r}, \omega) = i\omega \tilde{\mathbf{B}}(\mathbf{r}, \omega) \tag{4.1.7}$$

$$\nabla \cdot \tilde{\mathbf{B}}(\mathbf{r}, \omega) = 0 \tag{4.1.8}$$

$$\nabla \times \tilde{\mathbf{H}}(\mathbf{r}, \omega) = -i\omega \tilde{\mathbf{D}}(\mathbf{r}, \omega). \tag{4.1.9}$$

The polarization field $\mathbf{P}(\mathbf{r}, t)$ describes the ***internal polarisation field generated inside a material in response to an externally incident electric field*** $\mathbf{E}(\mathbf{r}, t)$. In real, linear media, the two are related in that the polarisation field follows the externally applied field with a temporal response that is not instantaneous, but instead described by ***a convolution*** with a function $\chi(\mathbf{r}, \tau)$. What this means is that a polarisation field $\mathbf{P}$ at time t depends on the electric field that is incident onto a material at all times τ prior to t, with a strength given by $\chi(\mathbf{r}, \tau)$:

$$\mathbf{P}(\mathbf{r}, t) = \epsilon_0 \int_{-\infty}^{t} \chi(\mathbf{r}, t - \tau) \mathbf{E}(\mathbf{r}, \tau) d\tau \tag{4.1.10}$$

$$\tilde{\mathbf{P}}(\mathbf{r}, \omega) = \epsilon_0 \tilde{\chi}(\mathbf{r}, \omega) \tilde{\mathbf{E}}(\mathbf{r}, \omega). \tag{4.1.11}$$

$\tilde{\chi}(\mathbf{r},\omega)$ is called the electric susceptibility. We find that the electric susceptibility is generally dependent on the frequency ω of the incident electric field. This phenomenon is referred to as ***dispersion***. Such frequency dependency implies that the response of the material is ***not instantaneous***.

The relation between the displacement field and the electric field can then be rewritten:

$$\tilde{\mathbf{D}}(\mathbf{r},\omega) = \epsilon_0(1+\tilde{\chi}(\mathbf{r},\omega))\tilde{\mathbf{E}}(\mathbf{r},\omega) \tag{4.1.12}$$

$$= \epsilon_0\tilde{\epsilon}_r(\mathbf{r},\omega)\tilde{\mathbf{E}}(\mathbf{r},\omega) \tag{4.1.13}$$

Similarly, the magnetization $\mathbf{M}(\mathbf{r},t)$ depends on the externally applied magnetic field as $\mathbf{M}(\mathbf{r},t) = \chi_m\mathbf{H}(\mathbf{r},t)$. χ_m is the relative magnetic susceptibility that indicates the degree of magnetization of a material in response to an applied magnetic field. In non-magnetic materials, $\chi_m = 0$. The underscore $_m$ is used to distinguish between the electric and magnetic susceptibility.

Finally, the complex refractive index of the medium is related to the complex dielectric function:

$$\tilde{\epsilon}_r(\mathbf{r},\omega) = 1+\tilde{\chi}(\mathbf{r},\omega) = \Re\{\tilde{\epsilon}_r(\mathbf{r},\omega)\} - i\Im\{\tilde{\epsilon}_r(\mathbf{r},\omega)\} \tag{4.1.14}$$

$$= (\tilde{n}_r(\mathbf{r},\omega) - i\tilde{n}_i(\mathbf{r},\omega))^2 \tag{4.1.15}$$

$$= \underbrace{\tilde{n}_r(\mathbf{r},\omega)^2 - \tilde{n}_i(\mathbf{r},\omega)^2}_{\Re\{\tilde{\epsilon}_r(\mathbf{r},\omega)\}} - i\underbrace{2\tilde{n}_r(\mathbf{r},\omega)\tilde{n}_i(\mathbf{r},\omega)}_{\Im\{\tilde{\epsilon}_r(\mathbf{r},\omega)\}} \tag{4.1.16}$$

Note that the refractive index generally contains information about the loss and gain in a medium through the imaginary part.

When the material response is ***instantaneous and linear***, $\tilde{\chi}(\omega) = \chi^{(1)} = \chi = const.$ and we can write that:

$$\mathbf{P}(\mathbf{r},t) = \epsilon_0\chi^{(1)}\mathbf{E}(\mathbf{r},t) \tag{4.1.17}$$

$$\tilde{\mathbf{P}}(\mathbf{r},\omega) = \epsilon_0\chi^{(1)}\tilde{\mathbf{E}}(\mathbf{r},\omega). \tag{4.1.18}$$

4.2 Plane waves in linear media

Recasting Maxwell's equations above we can retrieve the vectorial form of wave equations for the electric and magnetic fields by using $\nabla\times(\nabla\times\mathbf{F}) = \nabla(\nabla\cdot\mathbf{F}) - \nabla^2\mathbf{F}$:

$$\nabla^2\tilde{\mathbf{E}}(\mathbf{r},\omega) + \nabla(\frac{\nabla\tilde{\epsilon}_r(\mathbf{r},\omega)}{\tilde{\epsilon}_r(\mathbf{r},\omega)}\cdot\tilde{\mathbf{E}}(\mathbf{r},\omega)) + k_0^2\tilde{\epsilon}_r(\mathbf{r},\omega)\tilde{\mathbf{E}}(\mathbf{r},\omega) = 0 \tag{4.2.1}$$

$$\nabla^2\tilde{\mathbf{H}}(\mathbf{r},\omega) + \frac{\nabla\tilde{\epsilon}_r(\mathbf{r},\omega)}{\tilde{\epsilon}_r(\mathbf{r},\omega)}\times(\nabla\times\tilde{\mathbf{H}}(\mathbf{r},\omega)) + k_0^2\tilde{\epsilon}_r(\mathbf{r},\omega)\tilde{\mathbf{H}}(\mathbf{r},\omega) = 0 \tag{4.2.2}$$

Here, $k_0 = \frac{2\pi}{\lambda_0}$ is the wavenumber in vacuum, with $\omega = c_0 k_0$. In weakly inhomogenous media, the dielectric function $\epsilon_r(\mathbf{r}, \omega)$ does not change significantly over lengthscales similar to an optical wavelength, the wave equations can be simplified to

$$\nabla^2 \tilde{\mathbf{E}}(\mathbf{r}, \omega) + k_0^2 \tilde{\epsilon}_r(\mathbf{r}, \omega) \tilde{\mathbf{E}}(\mathbf{r}, \omega) = 0 \tag{4.2.3}$$

$$\nabla^2 \tilde{\mathbf{H}}(\mathbf{r}, \omega) + k_0^2 \tilde{\epsilon}_r(\mathbf{r}, \omega) \tilde{\mathbf{H}}(\mathbf{r}, \omega) = 0 \tag{4.2.4}$$

Generalising, we can simply write

$$\nabla^2 \tilde{U}(\mathbf{r}, \omega) + k_0^2 \tilde{\epsilon}_r(\mathbf{r}, \omega) \tilde{U}(\mathbf{r}, \omega) = 0 \tag{4.2.5}$$

which is the famous ***Helmholtz equation*** applied to the electric and magnetic fields, where $\tilde{U}$ can be any of the x, y, z-components of either the electric or the magnetic fields $(E_x, E_y, E_z, H_x, H_y, H_z)$. Another way to interpret the Helmholtz equation is that the x, y, z components of the electric field and magnetic field are decoupled and that similar equations as the two above can be found for each component individually.

In the case of isotropic homogenous media, the dielectric function does not depend on space and hence $\epsilon_r(\mathbf{r}, \omega) = \epsilon_r(\omega)$. The solution of the wave equation above are plane waves:

$$\mathbf{E}(\mathbf{r}, t) = \frac{1}{2\pi} \Re\{\tilde{\mathbf{E}}(\mathbf{r}, \omega) e^{-i\omega t}\} = \frac{1}{2\pi} \Re\{\mathbf{E}_0 e^{i(\mathbf{kr} - \omega t)}\} \tag{4.2.6}$$

$$\mathbf{H}(\mathbf{r}, t) = \frac{1}{2\pi} \Re\{\tilde{\mathbf{H}}(\mathbf{r}, \omega) e^{-i\omega t}\} = \frac{1}{2\pi} \Re\{\mathbf{H}_0 e^{i(\mathbf{kr} - \omega t)}\} \tag{4.2.7}$$

where $\mathbf{k}$ is the wavevector pointing into the direction that is perpendicular to the planes of equal phase, also called phase fronts, and has a magnitude $\mathbf{k}|^2 = \epsilon_r |k_0|^2$. $\mathbf{E}_0$ and $\mathbf{B}_0$ are the complex amplitudes of the fields and contain also the overall phase.

For a plane wave, we find that

$$\mathbf{H}_0 = \frac{1}{\omega \mu_0} \mathbf{k} \times \mathbf{E}_0 \tag{4.2.8}$$

$$\mathbf{E}_0 = -\frac{1}{\omega \epsilon_0 \epsilon_r} \mathbf{k} \times \mathbf{H}_0 \tag{4.2.9}$$

$$\mathbf{k} \cdot \mathbf{E}_0 = 0 \tag{4.2.10}$$

$$\mathbf{k} \cdot \mathbf{H}_0 = 0. \tag{4.2.11}$$

The equations above describe that $\mathbf{k}$, $\mathbf{E}_0$ and $\mathbf{B}_0$ are mutually orthogonal and that there is no phase difference between $\mathbf{E}_0$ and $\mathbf{H}_0$.

4.3 Laws of refraction and reflection at an interface

In the following, we will derive what happens to electromagnetic fields that are incident onto an interface as a function of the incident angle and their polarisation as shown in Fig. 4.1. We shall assume without loss of generality that the plane of incidence is the yz plane and denote $\mathbf{n}$ the vector normal to the interface.

A wave with corresponding electric and magnetic field $\mathbf{E}_1$ and $\mathbf{H}_1$ that is incident onto an interface of refractive index n_1 and n_2 will be partially reflected into a wave with corresponding electric and magnetic field $\mathbf{E}_3$ and $\mathbf{H}_3$ and partially transmitted into a wave with corresponding electric and magnetic field $\mathbf{E}_2$ and $\mathbf{H}_2$. The amount of reflected and transmitted wave depends on the polarisation of the incident wave (described through $\mathbf{E}_1$), on the angle of incidence θ_1 and on the refractive indices of the two media.

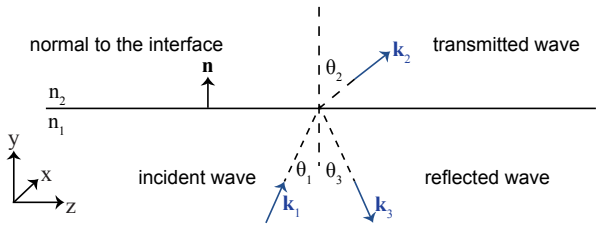

Figure 4.1: Reflection and transmission at an interface between two media.

4.3.1 Boundary conditions

Irrespective of the incident polarisation, the following boundary conditions must be fulfilled in the case of homogenous, isotropic media without any free charges:

Tangential components of $\mathbf{E}$ and $\mathbf{H}$ are continuous:

$$(\mathbf{E}_1 + \mathbf{E}_3 - \mathbf{E}_2) \times \mathbf{n} = 0 \tag{4.3.1}$$

$$(\mathbf{H}_1 + \mathbf{H}_3 - \mathbf{H}_2) \times \mathbf{n} = 0. \tag{4.3.2}$$

Normal components of $\mathbf{D}$ and $\mathbf{B}$ are continuous:

$$(n_1^2\mathbf{E}_1 + n_1^2\mathbf{E}_3 - n_2^2\mathbf{E}_2) \cdot \mathbf{n} = 0 \tag{4.3.3}$$

$$(\mathbf{B}_1 + \mathbf{B}_3 - \mathbf{B}_2) \cdot \mathbf{n} = 0. \tag{4.3.4}$$

A result of these boundary conditions is that at an interface, the longitudinal components of the wave vector must be conserved:

$$k_{1,z} = k_{2,z} = k_{3,z}, \tag{4.3.5}$$

that Snells law is applied

$$n_1 \sin(\theta_1) = n_2 \sin(\theta_2), \tag{4.3.6}$$

and that the angle of reflected wave is the same as the angle of the incident wave:

$$\theta_1 = \theta_3 \tag{4.3.7}$$

The wave vectors have the following magnitudes:

$$|k_1|^2 = k_{1,y}^2 + k_{1,z}^2 = n_1^2 k_0^2 = |k_3|^2 \tag{4.3.8}$$

$$|k_2|^2 = k_{2,y}^2 + k_{2,z}^2 = n_2^2 k_0^2. \tag{4.3.9}$$

Let's demonstrate the above for any arbitrary wave that is incident in the yz plane. We use eq. 4.3.3 and 4.2.9 and the rule for the scalar triple product $(\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}))$ we find that

$$(\frac{n_1^2}{n_1^2}\mathbf{k}_1 \times \mathbf{H}_1 + \frac{n_1^2}{n_1^2}\mathbf{k}_3 \times \mathbf{H}_3 - \frac{n_2^2}{n_2^2}\mathbf{k}_2 \times \mathbf{H}_2) \cdot \mathbf{n} = 0 \tag{4.3.10}$$

$$\mathbf{H}_1 \cdot (\mathbf{n} \times \mathbf{k}_1) + \mathbf{H}_3 \cdot (\mathbf{n} \times \mathbf{k}_3) - \mathbf{H}_2 \cdot (\mathbf{n} \times \mathbf{k}_2) = 0 \tag{4.3.11}$$

$$\mathbf{H}_1 \mathbf{e}_x k_1 \sin(\theta_1) + \mathbf{H}_3 \mathbf{e}_x k_3 \sin(\theta_3) - \mathbf{H}_2 \mathbf{e}_x k_2 \sin(\theta_2) = 0 \tag{4.3.12}$$

where $k_1 = |\mathbf{k}_1| = n_1 k_0$, $k_3 = |\mathbf{k}_3| = n_1 k_0$ and $k_2 = |\mathbf{k}_2| = n_2 k_0$. $\mathbf{e}_x$ is the unit vector along the x-direction. Using the continuity of the tangential component of the **H**-field, we find

$$\mathbf{H}_1 \mathbf{e}_x + \mathbf{H}_3 \mathbf{e}_x - \mathbf{H}_2 \mathbf{e}_x = 0 \tag{4.3.13}$$

and find overall that

$$\mathbf{H}_1 \mathbf{e}_x (k_1 \sin(\theta_1) - k_2 \sin(\theta_2)) + \mathbf{H}_3 \mathbf{e}_x (k_3 \sin(\theta_3) - k_2 \sin(\theta_2)) = 0 \tag{4.3.14}$$

In analogy to this, we can now use eq. 4.3.4 and 4.2.9 and find:

$$(\mathbf{k}_1 \times \mathbf{E}_1 + \mathbf{k}_3 \times \mathbf{E}_3 - \mathbf{k}_2 \times \mathbf{E}_2) \cdot \mathbf{n} = 0 \tag{4.3.15}$$

$$\mathbf{E}_1 \cdot (\mathbf{n} \times \mathbf{k}_1) + \mathbf{E}_3 \cdot (\mathbf{n} \times \mathbf{k}_3) - \mathbf{E}_2 \cdot (\mathbf{n} \times \mathbf{k}_2) = 0 \tag{4.3.16}$$

$$\mathbf{E}_1 \mathbf{e}_x k_1 \sin(\theta_1) + \mathbf{E}_3 \mathbf{e}_x k_3 \sin(\theta_3) - \mathbf{E}_2 \mathbf{e}_x k_2 \sin(\theta_2) = 0 \tag{4.3.17}$$

Using the continuity of the tangential component of the **E**-field, we find

$$\mathbf{E}_1 \mathbf{e}_x + \mathbf{E}_3 \mathbf{e}_x - \mathbf{E}_2 \mathbf{e}_x = 0 \tag{4.3.18}$$

and find overall that

$$\mathbf{E}_1 \mathbf{e}_x (k_1 \sin(\theta_1) - k_2 \sin(\theta_2)) + \mathbf{E}_3 \mathbf{e}_x (k_3 \sin(\theta_3) - k_2 \sin(\theta_2)) = 0 \tag{4.3.19}$$

Both eq. 4.3.14 and 4.3.19 need to be satisfied simultaneously for any incident angle and any orientation of $\mathbf{H}_1$ and $\mathbf{E}_1$ that satisfy the plane wave rules discussed in the previous section. This is the case if the following equations are fulfilled simultaneously

$$k_1 \sin(\theta_1) - k_2 \sin(\theta_2) = 0, \; and \tag{4.3.20}$$

$$k_3 \sin(\theta_3) - k_2 \sin(\theta_2) = 0 \tag{4.3.21}$$

and hence $\theta_1 = \theta_3$ and $n_1 \sin(\theta_1) = n_2 \sin(\theta_2)$. With $k_{1,z} = k_1 \sin\theta_1$, $k_{3,z} = k_1 \sin\theta_3$ and $k_{2,z} = k_2 \sin\theta_2$, it follows directly that the transverse component of the wave vector is conserved at an interface $k_{1,z} = k_{3,z} = k_{2,z}$.

These laws of reflection and refraction will be extremely important throughout our lecture, for example in the case of waveguides and fibers, resonators and gratings.

4.3.2 TE and TM modes

In our treatment so far, the exact orientation of the electric and magnetic fields of a plane wave were arbitrary, as long as they fulfill the plane wave rules of eq. 4.2.9 and 4.2.8. This means that all properties we derived so far apply to waves where the corresponding electric field can have any orientation. To differentiate between the various orientations, one speaks about ***the polarisation of a wave***. This denotes the axis along which the ***electric field*** of a wave oscillates. For example, an x-polarised beam has its electric field oscillating along the x-axis.

However, when a plane wave is incident onto a surface, it is more useful to define the type of wave with respect to the orientation of the electric and

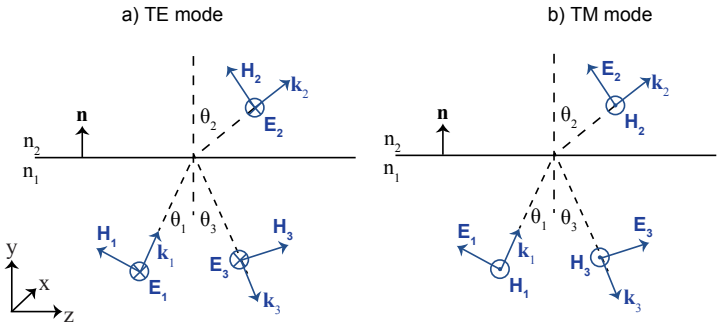

Figure 4.2: Transverse-electric (TE) and transverse magnetic (TM) modes incident onto an inferface that delimits two media of refractive index n_1 (incident medium) and n_2.

magnetic fields to the plane of incidence. We therefore distinguish two types of waves: transverse electric (TE) and transverse magnetic (TM) waves, shown in Fig. 4.2.

For a so-called TE-polarised wave, the electric field $\mathbf{E}_1$ is perpendicular to the plane of incidence and the magnetic field $\mathbf{H}_1$ is parallel to the plane of incidence. For a TM-polarized light, the magnetic field $\mathbf{H}_1$ is perpendicular to the plane of incidence and the electric field $\mathbf{E}_1$ is parallel to it.

The reflection and transmission coefficients of the electric fields where $\mathbf{E}_2 = t\mathbf{E}_1$ and $\mathbf{E}_3 = r\mathbf{E}_1$ are then obtained by considering the boundary conditions for the electric and magnetic fields in these two cases.

This will be treated in the exercise class, but we find:

$$r_{TE}^E = \frac{\mathbf{E}_3}{\mathbf{E}_1} = \frac{k_{1,y} - k_{2,y}}{k_{1,y} + k_{2,y}} = \frac{n_1 \cos\theta_1 - n_2 \cos\theta_2}{n_1 \cos\theta_1 + n_2 \cos\theta_2} \tag{4.3.22}$$

$$t_{TE}^E = \frac{\mathbf{E}_2}{\mathbf{E}_1} = 1 + r_{TE} = \frac{2n_1 \cos\theta_1}{n_1 \cos\theta_1 + n_2 \cos\theta_2} \tag{4.3.23}$$

$$r_{TM}^H = \frac{\mathbf{H}_3}{\mathbf{H}_1} = \frac{n_2^2 k_{1,y} - n_1^2 k_{2,y}}{n_2^2 k_{1,y} + n_1^2 k_{2,y}} = \frac{n_2 \cos\theta_1 - n_1 \cos\theta_2}{n_2 \cos\theta_1 + n_1 \cos\theta_2} \tag{4.3.24}$$

$$t_{TM}^H = \frac{\mathbf{H}_2}{\mathbf{H}_1} = 1 + r_{TM} = \frac{2n_2 \cos\theta_1}{n_2 \cos\theta_1 + n_1 \cos\theta_2} \tag{4.3.25}$$

To demonstrate the above, it may be useful to remind that $\mathbf{E} = \frac{c_0 \mathbf{B}}{n}$ (NB: when considering a wave propagating into a certain direction, mind the sign between $\mathbf{E}$ and $\mathbf{B}$!). Note that $T \neq |t|^2$ but rather that $T = 1 - r^2$.

It is worthwhile mentioning here that often times, the reflection and transmission coefficients of the TM mode are also rather defined in terms of the electric fields $\mathbf{E}_1$, $\mathbf{E}_2$ and $\mathbf{E}_3$ rather than the magnetic fields $\mathbf{H}_1$, $\mathbf{H}_2$ and $\mathbf{H}_3$. In that case we have the following:

$$r_{TM}^E = \frac{\mathbf{E}_3}{\mathbf{E}_1} = \frac{n_1 \mathbf{H}_3}{n_1 \mathbf{H}_1} \frac{n_2^2 k_{1,y} - n_1^2 k_{2,y}}{n_2^2 k_{1,y} + n_1^2 k_{2,y}} = \frac{n_2 \cos\theta_1 - n_1 \cos\theta_2}{n_2 \cos\theta_1 + n_1 \cos\theta_2} \tag{4.3.26}$$

$$t_{TM}^E = \frac{\mathbf{E}_2}{\mathbf{E}_1} = \frac{n_1 \mathbf{H}_2}{n_2 \mathbf{H}_1} = \frac{n_1}{n_2}(1 + r_{TM}) = \frac{2n_1 \cos\theta_1}{n_2 \cos\theta_1 + n_1 \cos\theta_2} \tag{4.3.27}$$

4.3.3 Derivation of the above

To compute the transmission and reflection coefficients we must use the two other boundary conditions of eq. 4.3.2 and 4.3.1, together with eq. 4.2.8 and 4.2.9.

We start first with eq. 4.3.2:

$$(\mathbf{k}_1 \times \mathbf{E}_1 + \mathbf{k}_3 \times \mathbf{E}_3 - \mathbf{k}_2 \times \mathbf{E}_2) \times \mathbf{n} = 0 \tag{4.3.28}$$

$$(\mathbf{n} \cdot \mathbf{E}_1) \cdot \mathbf{k}_1 + (\mathbf{n} \cdot \mathbf{E}_3) \cdot \mathbf{k}_3 - (\mathbf{n} \cdot \mathbf{E}_2) \cdot \mathbf{k}_2 - (\mathbf{n} \cdot \mathbf{k}_1) \cdot \mathbf{E}_1 - (\mathbf{n} \cdot \mathbf{k}_3) \cdot \mathbf{E}_3 + (\mathbf{n} \cdot \mathbf{k}_2) \cdot \mathbf{E}_2 = 0 \tag{4.3.29}$$

For a TE-mode, all products $\mathbf{n} \cdot \mathbf{E}_1 = \mathbf{n} \cdot \mathbf{E}_3 = \mathbf{n} \cdot \mathbf{E}_2 = 0$ (since perpendicular). In addition, $\mathbf{n} \cdot \mathbf{k}_1 = k_1 \cos(\theta_1)$, $\mathbf{n} \cdot \mathbf{k}_3 = -k_3 \cos(\theta_3)$ (mind the orientation of the vectors!) and $\mathbf{n} \cdot \mathbf{k}_2 = k_2 \cos(\theta_2)$. With $\mathbf{E}_1 + \mathbf{E}_3 = \mathbf{E}_2$, Snells law and the law of reflection, we then find

$$r_{TE} = \frac{\mathbf{E}_3}{\mathbf{E}_1} = \frac{k_{1,y} - k_{2,y}}{k_{1,y} + k_{2,y}} = \frac{n_1 \cos\theta_1 - n_2 \cos\theta_2}{n_1 \cos\theta_1 + n_2 \cos\theta_2} \tag{4.3.30}$$

$$t_{TE} = \frac{\mathbf{E}_2}{\mathbf{E}_1} = 1 + r_{TE} = \frac{2n_1 \cos\theta_1}{n_1 \cos\theta_1 + n_2 \cos\theta_2} \tag{4.3.31}$$

Similarly, we now also treat eq. 4.3.1 and eq. 4.2.9

$$(\frac{1}{n_1^2}\mathbf{k}_1 \times \mathbf{H}_1 + \frac{1}{n_1^2}\mathbf{k}_3 \times \mathbf{H}_3 - \frac{1}{n_2^2}\mathbf{k}_2 \times \mathbf{H}_2) \times \mathbf{n} = 0 \tag{4.3.32}$$

$$(\frac{1}{n_1^2}\mathbf{n} \cdot \mathbf{H}_1) \cdot \mathbf{k}_1 + \frac{1}{n_1^2}(\mathbf{n} \cdot \mathbf{H}_3) \cdot \mathbf{k}_3 - \frac{1}{n_2^2}(\mathbf{n} \cdot \mathbf{H}_2) \cdot \mathbf{k}_2 - \frac{1}{n_1^2}(\mathbf{n} \cdot \mathbf{k}_1) \cdot \mathbf{H}_1 - \frac{1}{n_1^2}(\mathbf{n} \cdot \mathbf{k}_3) \cdot \mathbf{H}_3 + \frac{1}{n_2^2}(\mathbf{n} \cdot \mathbf{k}_2) \cdot \mathbf{H}_2 = 0 \tag{4.3.33}$$

For a TE-mode, all products $\mathbf{n} \cdot \mathbf{H}_1 = \mathbf{n} \cdot \mathbf{H}_3 = \mathbf{n} \cdot \mathbf{H}_2 = 0$ (since perpendicular). In addition, $\mathbf{n} \cdot \mathbf{k}_1 = k_1 \cos(\theta_1)$, $\mathbf{n} \cdot \mathbf{k}_3 = -k_3 \cos(\theta_3)$ (mind the orientation of the vectors!) and $\mathbf{n} \cdot \mathbf{k}_2 = k_2 \cos(\theta_2)$. With $\mathbf{H}_1 + \mathbf{H}_3 = \mathbf{H}_2$, Snells law and the law of reflection, we then find

$$r_{TM} = \frac{\mathbf{H}_3}{\mathbf{H}_1} = \frac{n_2^2 k_{1,y} - n_1^2 k_{2,y}}{n_2^2 k_{1,y} + n_1^2 k_{2,y}} = \frac{n_2 \cos\theta_1 - n_1 \cos\theta_2}{n_2 \cos\theta_1 + n_1 \cos\theta_2} \tag{4.3.34}$$

$$t_{TM} = \frac{\mathbf{H}_2}{\mathbf{H}_1} = 1 + r_{TM} = \frac{2n_2 \cos\theta_1}{n_2 \cos\theta_1 + n_1 \cos\theta_2} \tag{4.3.35}$$

We note one important aspect of the equations above. While the expressions of the reflection and transmission coefficients as a function of the wavevector components are generally valid, the expressions as a function of angle and refractive index are valid in that form only for cases where total internal reflection does not occur. We will see in the next section how these must be modified in the particular case of total internal reflection.

4.3.4 Total internal reflection at a dielectric interface

We have studied in the previous section how a wave with arbitrary polarisation behaves when incident onto an interface. We have found that the

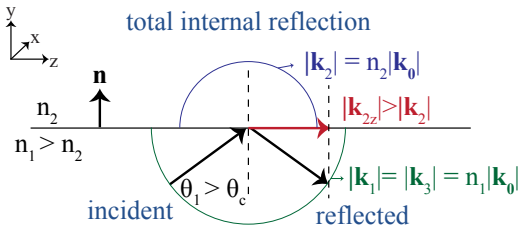

Figure 4.3: At total internal reflection, the refracted wave propagates along the waveguide interface. Its y component is imaginary (has an exponential decay) to accommodate the fact that $|k_{2z}| > |k_2|$.

angle of incidence influences both the magnitudes and the phases of the reflected and transmitted waves. In the case of total internal reflection, t describes the transmitted electric field just behind the interface. It is an evanescently decaying field which does not propagate as T=0. From the formulas above, one can see that total internal reflection occurs when the reflection coefficient is one, hence when $\theta_2 = \frac{\pi}{2}$. In that situation, $\theta_1 = \theta_c$ (for total internal reflection):

$$\sin\theta_c = \frac{n_2}{n_1}. \tag{4.3.36}$$

For $\theta_1 > \theta_c$, the transverse wave vector component of the transmitted wave must be purely imaginary since both of the following conditions need to be satisfied $k_{2z} = k_{3z}$ and $k_2 = n_2 k_0$ and $k_3 = n_1 k_0$ with $n_1 > n_2$. This is shown in Fig. 4.4 can be proven in the following way $n_2^2 k_0^2 = |k_2|^2 = k_{2y}^2 + k_{2z}^2 = k_{2y}^2 + k_{1z}^2 \overset{k_{1z}=n_1 k_0 \sin(\theta_1)}{=} k_{2y}^2 + k_0^2 n_1^2 \sin(\theta_1)^2 \overset{\theta_1 > \theta_c = \frac{n_2}{n_1}}{>} k_{2y}^2 + n_2^2 k_0^2$. Comparing the first and last term of this equation, we immediately find that $k_{2y} = i|k_{2y}|$ must be imaginary.

With the reflection and transmission coefficients we have calculated above, we can show that total internal reflection is accompanied by a phase shift:

$$r_{TE} = \frac{k_{1,y} - k_{2,y}}{k_{1,y} + k_{2,y}} = \frac{n_1\cos(\theta_1) - \sqrt{n_2^2 - n_1^2\sin^2(\theta_1)}}{n_1\cos(\theta_1) + \sqrt{n_2^2 - n_1^2\sin^2(\theta_1)}} = |r_{TE}|e^{i\phi_{r,TE}} \tag{4.3.37}$$

and

$$r_{TM} = \frac{n_2^2 k_{1,y} - n_1^2 k_{2,y}}{n_2^2 k_{1,y} + n_1^2 k_{2,y}} = \frac{n_1 n_2^2\cos(\theta_1) - n_1^2\sqrt{n_2^2 - n_1^2\sin^2(\theta_1)}}{n_1 n_2^2\cos(\theta_1) + n_1^2\sqrt{n_2^2 - n_1^2\sin^2(\theta_1)}} = |r_{TM}|e^{i\phi_{r,TM}} \tag{4.3.38}$$

For a TE mode $\phi_{r,TE} = arg(r_{TE})$ given by $\tan\frac{\phi_{r,TE}}{2} = \sqrt{\frac{\cos^2\theta_c}{\cos^2\theta_1} - 1}$. At $\theta_1 = \theta_c$, $\phi_{r,TE} = 0$. For a TM mode, $\phi_{r,TM} = arg(r_{TM})$ given by $\tan\frac{\phi_{r,TM}}{2} = \frac{-1}{\sin^2\theta_c}\sqrt{\frac{\cos^2\theta_c}{\cos^2\theta_1} - 1}$. At $\theta_1 = \theta_c$, $\phi_{r,TM} = 0$.

This means that at an angle above the one of total internal reflection, the transversal component of the transmitted field is decaying exponentially and no energy flows into medium 2. This principle will be used for wave guiding.

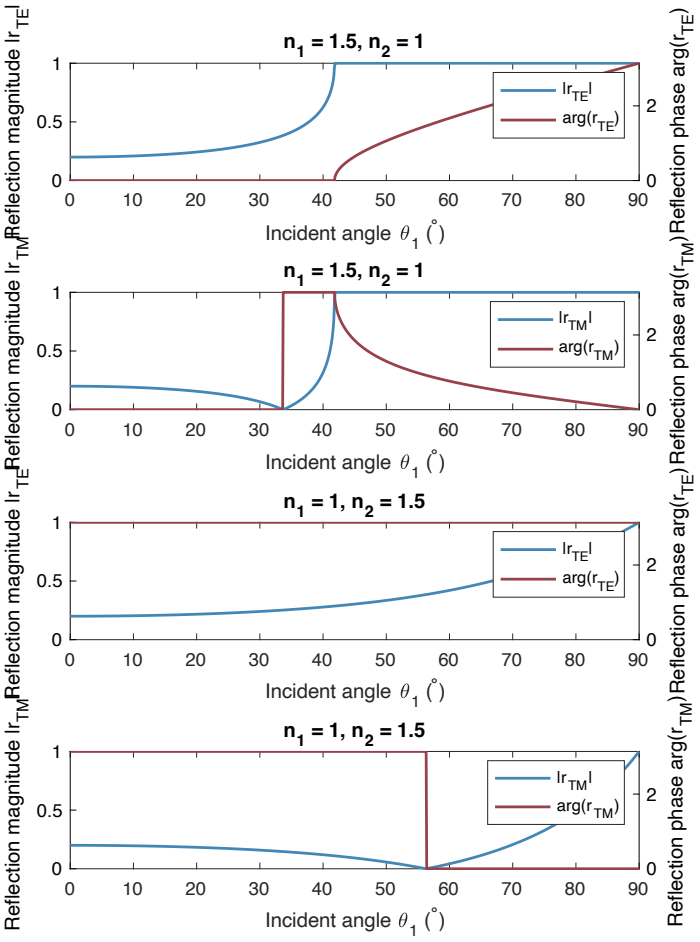

Figure 4.4: Electric field reflection coefficients and their phases for TE and TM modes for two scenarios ($n_1 = 1.5$ and $n_2 = 1$) and vice-versa ($n_1 = 1$ and $n_2 = 1.5$).

Optical waveguides

This chapter introduces waveguides as a mean to confine propagating electromagnetic waves to a specific region of space.
Key questions:
• What is a waveguide and what are waveguide modes?
• What is the effective and the group refractive index and what is the group velocity dispersion?
• How do the various mode properties depend on the parameters of a waveguide?
• What are the fundamental modes of slab waveguides? How do their properties depend on the mode order $_m$?
• What is waveguide dispersion? Explain the various dispersion orders.
• What is phase velocity? What is group velocity?
Key concepts: total internal reflection, self-consistency condition, evanescent wave, effective refractive index, group refractive index
Key equations: effective refractive index, cut-off frequency, multimode and single mode, group velocity dispersion, propagation constant
Literature: Saleh and Teich, chapter 9

In many applications of integrated photonics, light is not propagating through free-space but instead either on-chip, along waveguides, or along optical fibers. This enables control over the various properties of light by nanoscale patterning of dielectric media, made possible through nanofabrication techniques. For example, the energy of guided light can be confined to the core of a waveguide instead of its energy being spread over a large area in free-space. This is very useful e.g. to limit any interference and scattering that can otherwise easily distort optical signals in free space. Additionally, propagation speed, phase of light and other properties about which we will learn in this course, can be significantly different inside waveguides from free space, all depending - within certain bounds - largely on the nanoscale patterning of materials. This is an important resource to any engineer or scientist that aims at creating application specific photonic integrated circuits.

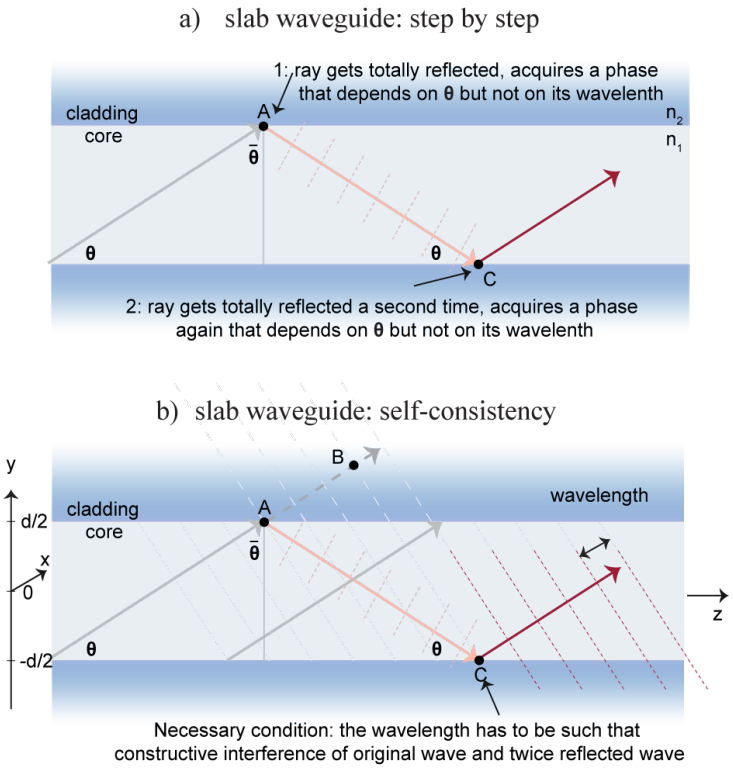

Figure 5.1: Waveguide modes of a slab waveguide. (a) Step-by-step evolution of a wave in a slab. b) self-consistency condition needs to be satisfied where the wave that reflects twice needs to be in-phase with the original wave.

In this chapter, we will understand the most basic principles behind guided modes inside waveguides and their properties on the nanoscale parameters.

5.1 Dielectric slab waveguides

A waveguide is a photonic structure that can guide light along a well-defined path. Unlike plane waves which fill up the entire space and propagate, waveguide modes ***are confined*** to a well-defined part of space. Unlike resonator modes (which we will discuss in section 9) that are also confined to a region but have zero group velocity, ***waveguide modes propagate***.

The most simple waveguide is a ***slab waveguide***. This is a photonic structure composed of three dielectric layers of infinite dimension in xz-

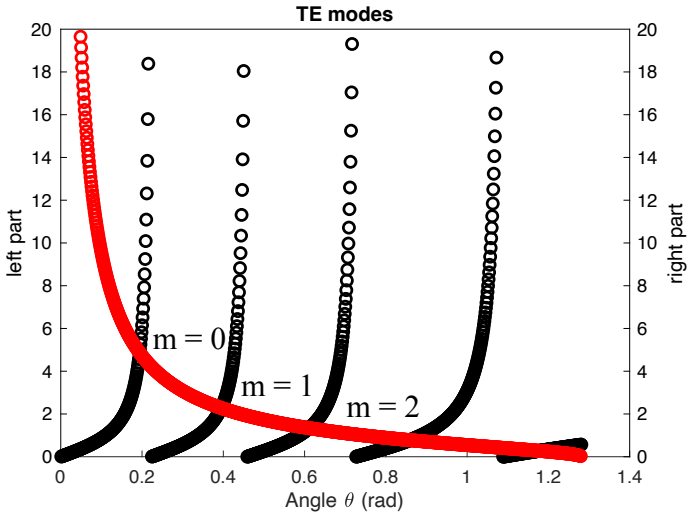

Figure 5.2: Graphical solution of the transcendental equation for a slab waveguide with core refractive index $n_1 = 3.5$ and cladding $n_2 = 1$, at a wavelength of $\lambda_0 = 1.55$ μm. We find guiding angles $\theta_0 = 0.19$, $\theta_1 = 0.38$ and $\theta_2 = 0.62$ rad, corresponding to effective indices of $n_{eff,0} = 3.47$, $n_{eff,1} = 3.25$, $n_{eff,2} = 2.84$.

plane, as shown in Fig. 5.1. The two outer media have a refractive index n_2 and the inner layer has a refractive index n_1. For the rest of this section, we assume all media to be lossless and hence n_1 and n_2 to be real. Guided modes are electromagnetic waves that propagate inside the inner layer and do not leak energy into the two outer layers. Consequently, ***total internal reflection*** of the guided wave must occur at the interface between the two materials.

In the following we will extend the discussion to a slab waveguide, which features two such interfaces as shown in Fig. 5.1. Note here the change in the definition of θ compared to the interface, but I wanted to stay compatible with Saleh and Teich (aside from their definition of θ_c which is messed up.)

The guided modes inside the waveguide must satisfy ***the self-consistency condition***. This is an extremely important principle in optics and we'll find it back also in the case of resonators. This implies that the interference of the original wave and the twice reflected wave must be constructive. In this way, the intensity inside the waveguide is not suppressed by interference, leading to disappearance of waveguiding. Given that the magnitude of the wave vector inside the core material is $|\mathbf{k}_1| = n_1 k_0$ and that the wave experiences a phase shift at the dielectric boundary $\phi_r = arg(r)$ with r the reflection coefficients introduced in the previous chapter, the self-

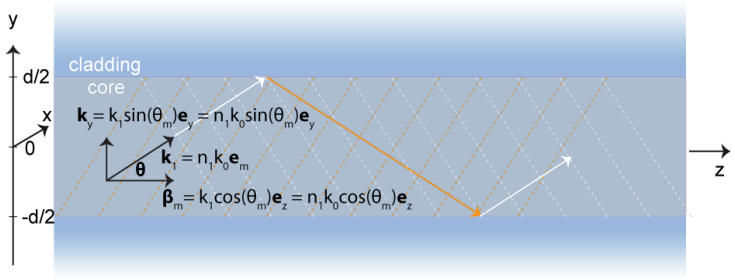

Figure 5.3: Graphical representation of the decomposition of the wave vector inside the core material along the two principal axes z and y for any m as $\mathbf{k}_1 = n_1 k_0 \cos(\theta_m)\mathbf{e}_z + n_1 k_0 \sin(\theta_m)\mathbf{e}_y$. Here $\mathbf{e}_m$ is the unit vector that describes the propagation direction of $\mathbf{k}_1$.

consistency condition can be formulated mathematically in the following way:

$$2n_1 k_0 AC - 2\phi_r = 2\pi m + n_1 k_0 (AC + AB) \tag{5.1.1}$$

With $AC = d/\sin\theta$ and $AB = AC\cos\left(\pi - 2\bar{\theta}\right)$, we find that

$$2n_1 k_0 d \sin\theta - 2\phi_r = 2\pi m. \tag{5.1.2}$$

We find a ***discrete set of modes*** that satisfy the self-consistency condition. This discrete set must fulfill a ***transcendental equation*** that depends on θ_m of the form

$$\tan\left(\frac{n_1 k_0 d \sin\theta_m}{2} - \frac{\pi}{2}m\right) = \tan\left(\frac{\phi_r}{2}\right) = \sqrt{\frac{\cos^2\bar{\theta}_c}{\cos^2\bar{\theta}_m} - 1}. \tag{5.1.3}$$

Solutions to this transcendental equations can be found graphically, by plotting the left and the right side of equation 5.1.3 as a function of θ_m and finding where they intersect. An example is shown in Fig. 5.2 for a slab waveguide with a core refractive index of $n_1 = 3.5$ and cladding $n_2 = 1$, for $m = 0, 1, 2, 3$. The intersections of the two curves appears at guiding angles $\theta_0 = 0.19$, $\theta_1 = 0.38$ and $\theta_2 = 0.62$ rad, corresponding to effective indices of $n_1 \cos\theta_0 = 3.47$, $n_1 \cos\theta_1 = 3.25$, $n_1 \cos\theta_2 = 2.84$.

Knowing the angle allows us to write the wave vector inside the core material for any m as $\mathbf{k}_1 = k_z\mathbf{e}_z + k_y\mathbf{e}_y = n_1 k_0 \cos(\theta_m)\mathbf{e}_z + n_1 k_0 \sin(\theta_m)\mathbf{e}_y$, as schematically shown in Fig. 5.3. We define the ***propagation constant*** as the component in forward $z-$direction $\beta_m = k_z = n_1 k_0 \cos\theta_m$, where the angles θ_m lie between 0 and $\theta_c = \frac{\pi}{2} - \bar{\theta}_c$.

To find the field distribution inside the slab waveguide, we must solve the ***Helmholtz equation***. The self-consistency condition we derived gives us

a good intuition on how the solutions must look like. Given the translational symmetry in x-direction, the solutions do not depend on x, but only on y and z. In addition, we found that the two sets of TE-modes and TM-modes behave differently when incident onto an interface. Without loss of generality, we will only discuss TE-modes, but the same principles can be applied to TM-modes. Given that the Helmholtz equation is homogeneous, we make an Ansatz of ***separation of variables***:

$$E_x(y,z) = u_m(y)h_m(z) \tag{5.1.4}$$

Plugging this into the Helmholtz equation, we find that we must satisfy simultaneously that

$$\frac{1}{u_m(y)}\frac{\partial^2}{\partial y^2}u_m(y) + \frac{1}{h_m(z)}\frac{\partial^2}{\partial z^2}h_m(z) + k_0^2 n_1^2 = 0, |y| \leq \frac{d}{2} \tag{5.1.5}$$

$$\frac{1}{u_m(y)}\frac{\partial^2}{\partial y^2}u_m(y) + \frac{1}{h_m(z)}\frac{\partial^2}{\partial z^2}h_m(z) + k_0^2 n_2^2 = 0, |y| > \frac{d}{2} \tag{5.1.6}$$

You may have seen this kind of equations in other lectures! Then you know that a common approach is to assume

$$\frac{1}{h_m(z)}\frac{\partial^2}{\partial z^2}h_m(z) = -A^2 \tag{5.1.7}$$

$$\frac{1}{u_m(y)}\frac{\partial^2}{\partial y^2}u_m(y) = A^2 - k_0^2 n_1^2, |y| \leq \frac{d}{2} \tag{5.1.8}$$

$$\frac{1}{u_m(y)}\frac{\partial^2}{\partial y^2}u_m(y) = A^2 - k_0^2 n_2^2, |y| > \frac{d}{2} \tag{5.1.9}$$

Here, let's first let our intuition guide us. We have seen that due to self-consistency the propagation constant in z-direction is β_m. Indeed, we can now assume $A = \beta_m$ and find $h_m(z) = e^{i\beta_m z}$. We then find:

$$\frac{1}{u_m(y)}\frac{\partial^2}{\partial y^2}u_m(y) = k_0^2 n_1^2 \cos^2\theta_m - k_0^2 n_1^2 = -k_0^2 n_1^2 \sin^2\theta_m, |y| \leq \frac{d}{2} \tag{5.1.10}$$

$$\frac{1}{u_m(y)}\frac{\partial^2}{\partial y^2}u_m(y) = \underbrace{k_0^2 n_1^2 \cos^2\theta_m - k_0^2 n_2^2}_{=\beta_m^2 - k_0^2 n_2^2 = \gamma_m^2 > 0}, |y| > \frac{d}{2} \tag{5.1.11}$$

The solutions for $u_m(y)$ are then inside the waveguide ($|y| < \frac{d}{2}$)

$$u_m(y) \propto \begin{cases} \cos(n_1 k_0 \sin\theta_m y), & |y| \leq \frac{d}{2},\ m = 0,2,4,... \\ \sin(n_1 k_0 \sin\theta_m y), & |y| \leq \frac{d}{2},\ m = 1,3,5,... \end{cases} \tag{5.1.12}$$

and outside the waveguide ($|y| \geq \frac{d}{2}$)

$$u_m(y) \propto \begin{cases} e^{-\gamma_m y}, & y > \frac{d}{2} \\ e^{\gamma_m y}, & y < -\frac{d}{2} \end{cases} \tag{5.1.13}$$

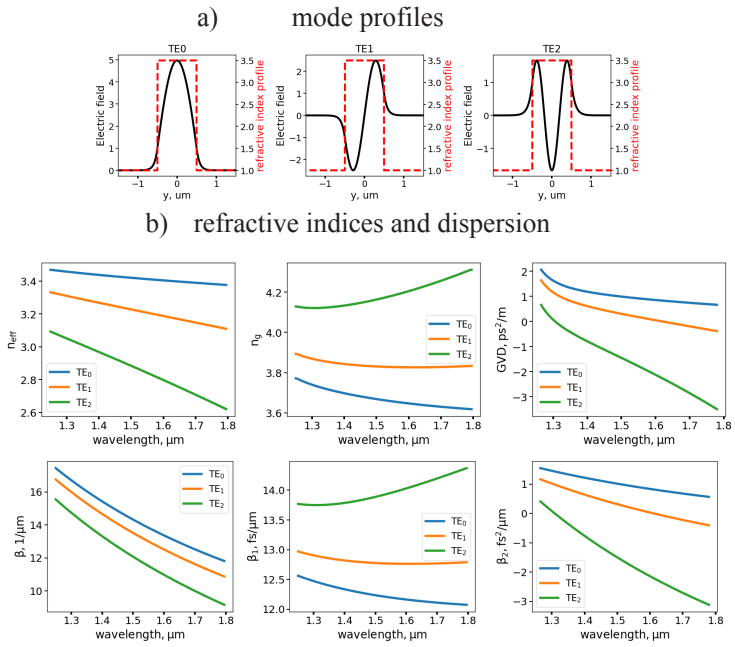

Figure 5.4: Guided modes inside a slab waveguide with core refractive index $n_1 = 3.5$ and cladding $n_2 = 1$, their refractive indices and dispersion.

where $\gamma_m^2 = \beta_m^2 - n_2^2 k_0^2$.

We notice that the phase accumulation $e^{i\beta_m z}$ of propagating guided modes along z is linked to the so-called ***effective refractive index*** $n_{eff}(\omega)$, which is defined as

$$\beta_m(\omega) = n_{eff,m}(\omega) k_0. \tag{5.1.14}$$

The waveguide modes are orthogonal and normalized, meaning that

$$\int_{-\infty}^{\infty} dy |u_m(y)|^2 = 1, \ \forall m \tag{5.1.15}$$

$$\int_{-\infty}^{\infty} dy u_m(y) u_n^*(y) = \delta_{mn}. \tag{5.1.16}$$

We plot these solutions for the first three modes in Fig. 5.4. Mathematically, we also recognise that we can introduce an ***effective width*** of the slab waveguide d_m, that is different for each mode, but that satisfies the condition that

$$n_1 k_0 \sin(\theta_m) d_m = (m+1)\pi. \tag{5.1.17}$$

In essence, this means that the transverse wave vector component of every mode is in such way that the wave exhibits a multiple integer of half cycles

along the total thickness of the slab. Looking at their mode profiles, we find that indeed, the TE_0 mode has one half-cycle, the TE_1 mode has one full cycle, and that the TE_2 mode has one and a half full cycles. The effective width takes into account the penetration depth into the cladding, the so-called ***Goos-Haenchen shift.*** In the case of high-confinement waveguides, where only a negligible part of the light is in the cladding material, we can safely approximate $d_m \approx d$. This will be very useful in chapter 8.

Finally, a slab waveguide that can support several transverse modes can be populated with any arbitrary superposition of all these transverse modes. Such arbitrary superposition can be written as a sum, satisfies Maxwell's equation through the superposition principle, and can have arbitrary weights a_m

$$\mathbf{E}(\mathbf{r},t) = E_x(y,z)e^{-i\omega t}\mathbf{e}_x = \sum_m a_m u_m(y) e^{-i(\omega t - \beta_m z)}\mathbf{e}_x, \ \forall y,z \quad (5.1.18)$$

Note that while the above equation is defined for a fixed frequency ω (the sum is over the transverse mode index), it can be expanded to accommodate also various frequencies. Such waveguides are then highly multimode.

5.2 Dispersion in waveguides

In our brief revision of Maxwell's equations, we have introduced the notion of ***dispersion*** as being the phenomenon where the response of a dielectric medium to an external field depends on the frequency. In the case of waveguides discussed here, we find that their propagation constant $\beta_m = k_z = n_1 k_0 \cos(\theta_m)$ depends strongly on the frequency at which the wave oscillates, both through k_0 and through θ_m which is a function of k_0. Naturally, this leads to a ***frequency-dependent propagation constant*** $\beta_m(\omega)$, where m stands for the transverse mode number.

In addition to the dispersion of the cladding and the core materials, the dispersion also depends on the introduced material patterning (the waveguide dimensions). The latter is typically refered to as ***waveguide dispersion***. In the following, we omit the mode order index m, but note that all equations are valid for any m.

We start by defining the so-called ***phase velocity***, which represents the speed at which fronts of equal phase travel in space:

$$v_{ph} = \frac{\omega}{\beta(\omega)} = \frac{c_0}{n_{eff}(\omega)}. \quad (5.2.1)$$

In the case where the effective index does not depend on frequency ω, all phase fronts travel at the same speed, irrespective of their frequency. This

means that if two phase fronts interfered constructively at $z = 0$, they will continue to interfere constructively as they propagate in space, meaning for any $z \neq 0$. To exemplify, let's consider a broadband electromagnetic wave

$$E(z,t) = \frac{1}{2\pi}\int d\omega \tilde{E}(\omega) e^{i(\beta z - \omega t)} \tag{5.2.2}$$

and with $\beta(\omega) = \beta$ independent of frequency we find

$$E(z,t) = \frac{1}{2\pi} e^{i\beta z}\int d\omega \tilde{E}(\omega) e^{-i\omega t} = e^{i\beta z} E(z=0,t) \tag{5.2.3}$$

The wave at $z \neq 0$ is therefore the same as at $z = 0$, up to a propagation phase determined by βz.

If the effective index does depend on frequency, the situation will be more complicated, since different colors will propagate at different speeds, and they may interfere in a non-trivial way. This is only relevant for light that is not monochromatic, as is the case of pulses of light that contain a broadband spectrum. To exemplify, let's consider a broadband electromagnetic wave

$$E(z,t) = \frac{1}{2\pi}\int d\omega \tilde{E}(\omega) e^{i(\beta(\omega) z - \omega t)} \tag{5.2.4}$$

Now the propagation constant $\beta(\omega)$ depends on ω. Typically, the spectrum of light is centered around a central frequency ω_0 and has a width $\Delta\omega$. We can therefore replace $\omega = \omega_0 + \Delta\omega$ and rewrite

$$E(z,t) = \frac{1}{2\pi} e^{-i\omega_0 t}\int d\Delta\omega \tilde{E}(\omega_0 + \Delta\omega) e^{i\beta(\omega_0+\Delta\omega) z} e^{-i\Delta\omega t} \tag{5.2.5}$$

To grasp the effects of dispersion on broadband light sources, it is instructive to perform a Taylor expansion of the propagation constant $\beta(\omega)$ around a frequency of interest ω_0:

$$\beta(\omega = \omega_0 + \Delta\omega) = \beta(\omega_0) + \Delta\omega \underbrace{\frac{\partial}{\partial\omega}\beta(\omega)\Big|_{\omega=\omega_0}}_{\beta_1} + \frac{1}{2}\Delta\omega^2 \underbrace{\frac{\partial^2}{\partial\omega^2}\beta(\omega)\Big|_{\omega=\omega_0}}_{\beta_2} + ... \tag{5.2.6}$$

$$\beta(\omega = \omega_0 + \Delta\omega) = \beta(\omega_0) + \Delta\omega\beta_1 + \frac{1}{2}\Delta\omega^2\beta_2 + ... \tag{5.2.7}$$

where β_1 is the ***first order dispersion***, β_2 is the ***second order dispersion*** and so on. Please note that $\beta(\omega_0)$, β_1 and β_2 have different units!

Going back to our equation 5.2.5, we find

$$E(z,t) = \frac{1}{2\pi} e^{i(\beta(\omega_0) z - \omega_0 t)}\int d\Delta\omega \underbrace{\tilde{E}(\omega_0 + \Delta\omega) e^{i(\Delta\omega\beta_1 + \frac{1}{2}\Delta\omega^2\beta_2 + ...) z}}_{\tilde{A}(\Delta\omega, z)} e^{-i\Delta\omega t} \tag{5.2.8}$$

We define a complex temporal envelope function $A(z,t)$ that is the inverse Fourier transform of the complex spectrum $\tilde{A}(\Delta\omega, z)$ with $A(z,t) = \mathcal{F}^{-1}\{\tilde{A}(\Delta\omega, z)\}$ that changes as the pulse propagates along z in a way that depends on the properties of $\beta(\omega_0 + \Delta\omega)$.

$$\frac{\partial}{\partial z}A(z,t) = i(\Delta\omega\beta_1 + \frac{1}{2}\Delta\omega^2\beta_2 + ...)A(z,t) \tag{5.2.9}$$

We eventually find that the speed at which $A(z,t)$ moves is the so-called ***group velocity***:

$$\frac{1}{v_g} = \frac{\Delta\omega\beta_1 + \frac{1}{2}\Delta\omega^2\beta_2 + ...}{\Delta\omega} \overset{\beta_2 \sim 0}{=} \beta_1 \tag{5.2.10}$$

From here, we can define the ***group index*** which is $n_g = \frac{c_0}{v_g} = c_0\beta_1 = c_0\frac{\partial\beta}{\partial\omega}|_{\omega_0} = c_0\frac{\partial(n_{eff}(\omega)\frac{\omega}{c_0})}{\partial\omega}|_{\omega_0}$. By doing some simple math , we find the formula for the group index:

$$n_g(\omega) = n_{eff}(\omega) + \omega\frac{\partial}{\partial\omega}n_{eff}(\omega) \tag{5.2.11}$$

Finally, we also introduce the ***group velocity dispersion (GVD)*** which is simply

$$GVD = \beta_2. \tag{5.2.12}$$

In Fig. 5.4 we show exemplarily the effective, group indices of silion waveguides of thickness 1 μm for TE_0, TE_1 and TE_2 mode as a function of wavelength. We find, as expected that the effective index of modes increases with the order m and that the group index is strongly different depending on the mode.

5.2.1 Other waveguide types

There are many different types of waveguides, among which the most prominent are shown in Fig. 5.5. As a general rule of thumb, all of them are made from a high-index material in combination with a low-index material. Most often, the light is guided inside the high-index material, but can be different! E.g. slot and plasmonic waveguides guide the light in between two rails of high-index material (or metal for the case of the plasmonic waveguide). The ***confinement*** provided by the waveguides can be carefully engineered by proper choice of dimensions and materials. Also, their evanescent decay inside the cladding can be engineered, and with it the coupling to other waveguides. We show in the figure panels g)-h) exemplary mode profiles of fields guided in waveguides d)-f).

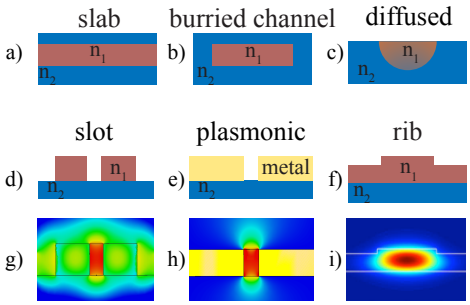

Figure 5.5: Waveguides come in many shapes and flavors and often are composed of a core that is surrounded by a cladding. Depending on the exact waveguide dimensions, the guided modes may be located inside the high index material (e.g. for slab, buried, rib waveguides) or outside it (e.g. for both the slot and the plasmonic waveguides, the guided modes are located in between the two rails). a)-f) show different waveguide geometries and g)-i) show the guided mode profiles of the waveguides shown in d)-f).

Another useful property of waveguides is that one can engineer their dispersion. This is extensively used in all photonic platforms that exist today.

In the simulation class you will study another type of waveguides, namely ridge waveguides. These support mixed TE and TM waveguides, so-called TEM waveguides.

Optical fibers

This chapter introduces fibers as a mean to confine propagating electromagnetic waves to a specific region of space.
Key questions:
• What is an optical fiber and which kind of fibers exist?
• How is a fiber different from a waveguide?
• What is the numerical aperture of a fiber?
• What are the fundamental modes of fibers? How do they depend on $\mathbf{r}$, ϕ and z?
• What are Bessel functions of first and second kind? Which one describes the spatial profile of a fiber mode and in which area of space?
• How do the various mode properties depend on the parameters of a fiber (core diameter, refractive index contrast)?// **Key concepts: step index fibers, graded index fibers, single mode fibers, multi-mode fibers, acceptance angle, numerical aperture**
Key equations: numerical aperture, propagation constant
Literature: Saleh and Teich, chapter 10

Optical fibers are made from a core, typically made from silica glass of high chemical purity, and a cladding, also from silica, that has a slightly different refractive index. This slight change in refractive index is introduced by adding dopants to the silica glass, such as titanium. The most simple optical fibers are ***step-index fibers***, where the core and the cladding typically have a shallow refractive index contrast. This constrains optical fibers to have core dimensions that are typically much larger than on-chip waveguides, with ***single mode fibers*** having field diameters on the order of 8 μm. Instead, ***multi-mode fibers*** have large core diameters up to 200 μm and can guide many higher-order transverse modes. Step-index, multi-mode fibers however have large modal dispersion - meaning that different transverse modes propagate at different speeds, because of their different effective refractive indices. To overcome this problem, graded index fibers can be used which, instead, have a nonhomogenous core refractive index that gets increasingly lower when going from the center of the core towards the cladding, reaching a minimal value at the core-cladding

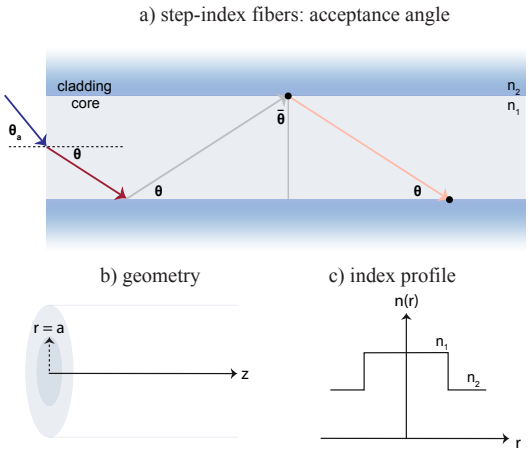

Figure 6.1: Optical fibers: acceptance angle, geometry and refractive index profile.

interface that equals the constant refractive index of the cladding. This allows the various refractive indices to be equalized.

6.1 Step-index fibers

A step-index fiber is a cylindrical type of waveguide with a core refractive index n_1 and a cladding refractive index n_2, as shown in Fig. 6.1. A useful quantity is the ***relative change in refractive index***

$$\Delta = \frac{n_1^2 - n_2^2}{n_1^2 + n_2^2} \simeq \frac{n_1 - n_2}{n_1}.$$

Just like in a waveguide, only rays with an angle $\theta < \frac{\pi}{2} - \bar{\theta}_c$ undergo total internal reflection. Because $n_2 \simeq n_1$, these angles θ are very small, and guided rays are approximately parallel to the propagation direction.

The critical angle for total internal reflection determines the ***acceptance angle*** θ_a of a fiber. This angle determines the cone angle of external rays that will be efficiently coupled into the fiber. Incident rays with angles greater than θ_a will be transmitted at the fiber facet but will also be transmitted at the core-cladding interface, instead of being totally internally reflected. This leads to leakage of energy outside of the fiber core and eventually to light's energy not remaining confined to the core.

If the incident light comes from air onto the fiber, then the acceptance angle and the critical angle are related through $\sin\theta_a = n_1 \sin\left(\frac{\pi}{2} - \bar{\theta}_c\right) =$

$n_1\sqrt{1-\cos^2(\frac{\pi}{2}-\bar{\theta}_c)} = n_1\sqrt{1-(\frac{n_2}{n_1})^2} = \sqrt{n_1^2-n_2^2}$. From here we introduce the definition of ***the numerical aperture*** NA as $\theta_a = \sin^{-1} NA$ and find

$$NA = \sqrt{n_1^2 - n_2^2} \simeq n_1\sqrt{2\Delta}. \tag{6.1.1}$$

A large numerical aperture therefore requires a large refractive index contrast between the core and the cladding.

6.2 Guided modes in step-index fibers

In this section, we will derive the properties of guided modes inside fibers, and exploit the cylindrical symmetry these exhibit. We will find that their propagation constants depend on the degrees of freedom of these modes.

We start by describing a fiber with core radius a in cylindrical coordinates. We therefore have a radially dependent refractive index that obeys

$$n(\mathbf{r}) = \begin{cases} n_1, \mathbf{r} \leq a \\ n_2, \mathbf{r} > a. \end{cases} \tag{6.2.1}$$

The electric and magnetic fields of each of the guided modes must satisfy the ***Helmholtz equation*** we derived in section 4.2, that we simplify for a ***monochromatic*** wave (one single frequency ω):

$$\nabla^2\tilde{\mathbf{E}}(\mathbf{r}) + k_0^2\tilde{\epsilon}_r(\mathbf{r})\tilde{\mathbf{E}}(\mathbf{r}) = 0 \tag{6.2.2}$$

$$\nabla^2\tilde{\mathbf{H}}(\mathbf{r}) + k_0^2\tilde{\epsilon}_r(\mathbf{r})\tilde{\mathbf{H}}(\mathbf{r}) = 0 \tag{6.2.3}$$

where $\tilde{\epsilon}_r(\mathbf{r}) = n^2(\mathbf{r})$. Each of the components of these vectorial fields $\tilde{\mathbf{E}} = \begin{pmatrix} E_r \\ E_\phi \\ E_z \end{pmatrix}$ and $\tilde{\mathbf{H}} = \begin{pmatrix} H_r \\ H_\phi \\ H_z \end{pmatrix}$ must fulfil the Helmholtz equation. To simplify the notation, we represent each component of the fields above as a scalar field in a compacted form

$$\nabla^2 U(\mathbf{r}) + k_0^2 n^2(\mathbf{r}) U(\mathbf{r}) = 0. \tag{6.2.4}$$

with $U(\mathbf{r})$ being any of the components E_r, E_ϕ, E_z and H_r, H_ϕ, H_z. We can now evaluate this Helmholtz equation in cylindrical coordinates $\mathbf{r} = (r, \phi, z)$ by introducing $\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2}$. We further make the Ansatz of separation of variables and assume that our solution must be of the form

$$U(r, \phi, z) = u(r)h(\phi)e^{i\beta z} \tag{6.2.5}$$

Plugging this into the Helmholtz equation, we find that we must fulfill

$$\underbrace{\frac{r^2}{u(r)}\left(\frac{\partial^2 u(r)}{\partial r^2} + \frac{1}{r}\frac{\partial u(r)}{\partial r}\right) - \beta^2 r^2 + k_0^2 n^2 r^2}_{l^2} + \underbrace{\frac{1}{h(\phi)}\frac{\partial^2 h(\phi)}{\partial \phi^2}}_{-l^2} = 0 \tag{6.2.6}$$

Since the two parts of the equation above each depend on only one single variable (r and ϕ, respectively), they must be each equal to a constant and all must sum up to zero. Hence, we set them to be equal to l^2 and $-l^2$, respectively. We can now solve the ϕ-dependent equation and find that

$$h(\phi) = A_+ e^{i(l\phi+\phi_0)} + A_- e^{-i(l\phi+\phi_0)} \tag{6.2.7}$$

with A_+ and A_- so that they fulfill the boundary conditions. Now since continuity of any field after a rotation of 2π must hold and hence $h(0) = h(2\pi)$, we find that l can take any integer value $l = 0, 1, 2, 3....$ Why don't we also consider negative indices? Well, negative indices would lead to an overdetermined space (alternatively, we can say that $h(\phi) = e^{i(\phi l+\phi_0)}$ and any supersposition thereof for any $l = 0, \pm 1, \pm 2$ and so on). Note: in the remainder of this subsection, we will always write just the $e^{il\phi}$ term for easy notation and math.

After some math, we find that the function $u(r)$ that describes the radial distribution of the fields must fulfill

$$\begin{cases} \frac{\partial^2 u(r)}{\partial r^2} + \frac{1}{r}\frac{\partial u(r)}{\partial r} + (n_1^2 k_0^2 - \frac{l^2}{r^2} - \beta^2)u(r) = 0, r \leq a \\ \frac{\partial^2 u(r)}{\partial r^2} + \frac{1}{r}\frac{\partial u(r)}{\partial r} + (n_2^2 k_0^2 - \frac{l^2}{r^2} - \beta^2)u(r) = 0, r > a \end{cases} \tag{6.2.8}$$

We have learnt in the previous section, that $\beta = n_{eff} k_0$, with n_{eff} taking values in between n_1 and n_2. Much like in the case of slab waveguides, we can define a parameter $\gamma^2 = \beta^2 - n_2^2 k_0^2$ and a transverse component of the wave vector $k_\perp^2 = n_1^2 k_0^2 - \beta^2$. Note that both $\gamma, k_\perp > 0$ are real numbers. With these definitions, we rewrite the equations for the core and the cladding as

$$\begin{cases} r^2 \frac{\partial^2 u(r)}{\partial r^2} + r\frac{\partial u(r)}{\partial r} + (k_\perp^2 r^2 - l^2)u(r) = 0, r \leq a \\ r^2 \frac{\partial^2 u(r)}{\partial r^2} + r\frac{\partial u(r)}{\partial r} - (\gamma^2 r^2 + l^2)u(r) = 0, r > a. \end{cases} \tag{6.2.9}$$

We substitute in the first equation $\tilde{r} = k_\perp r$ (and in the second equation $\tilde{r} = \gamma r$), and use that $\frac{\partial u}{\partial r} = \frac{\partial \tilde{r}}{\partial r}\frac{\partial u}{\partial \tilde{r}} = k_\perp \frac{\partial u}{\partial \tilde{r}}$ and $\frac{\partial^2 u}{\partial r^2} = \frac{\partial}{\partial r}(\frac{\partial u}{\partial r}) = \frac{\partial \tilde{r}}{\partial r}\frac{\partial}{\partial \tilde{r}}(\frac{\partial u}{\partial r}) = k_\perp \frac{\partial}{\partial \tilde{r}}(k_\perp \frac{\partial u}{\partial \tilde{r}}) = k_\perp^2 \frac{\partial^2 u}{\partial \tilde{r}^2})$ (and identically $\frac{\partial u}{\partial r} = \gamma \frac{\partial u}{\partial \tilde{r}}$ and $\frac{\partial^2 u}{\partial r^2} = \gamma^2 \frac{\partial^2 u}{\partial \tilde{r}^2}$) in the second equation) and find

$$\begin{cases} \tilde{r}^2 \frac{\partial^2 u(\tilde{r})}{\partial \tilde{r}^2} + \tilde{r}\frac{\partial u(\tilde{r})}{\partial \tilde{r}} + (\tilde{r}^2 - l^2)u(\tilde{r}) = 0, \tilde{r} \leq k_\perp a \\ \tilde{r}^2 \frac{\partial^2 u(\tilde{r})}{\partial \tilde{r}^2} + \tilde{r}\frac{\partial u(\tilde{r})}{\partial \tilde{r}} + (-\tilde{r}^2 - l^2)u(\tilde{r}) = 0, \tilde{r} > a\gamma. \end{cases} \tag{6.2.10}$$

This type of equations are standard and have solutions that take the form of Bessel functions. Being a differential equation of second order, there are two independent solutions to each differential equation:

$$u_l(r) = AJ_l(k_\perp r) + BY_l(k_\perp r), r \leq a \tag{6.2.11}$$

$$u_l(r) = CK_l(\gamma r) + DI_l(\gamma r), r > a \tag{6.2.12}$$

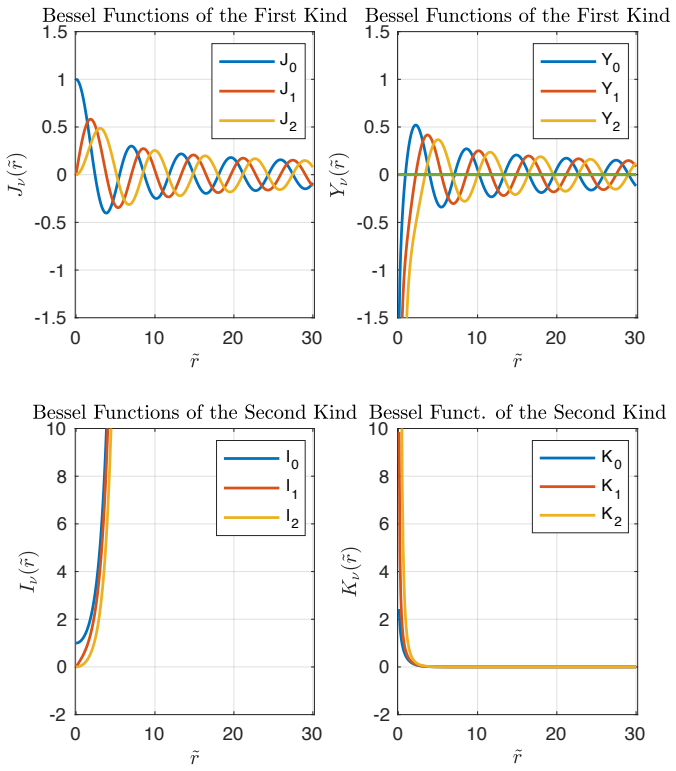

Figure 6.2: Bessel functions of first and second kind for orders $\nu = 0, 1, 2$.

These are Bessel functions of first kind and order l for the core (since l is integer value as we will show in an instant) and the modified Bessel function of second kind and order l for the cladding. To give you a flavor for the shape of these functions, we plot all four of them in Fig. 6.2. We see clearly that $J_l(\tilde{r})$ and $Y_l(\tilde{r})$ are harmonic functions with an additional decay as a function of $\tilde{r}$. Indeed, they can be approximated for large $\tilde{r}$ by $J_l(\tilde{r}) = \sqrt{\frac{2}{\pi z}}\cos\big(\tilde{r} - \frac{l\pi}{2} - \frac{\pi}{4}\big) + ...$ and $Y_l(\tilde{r}) = \sqrt{\frac{2}{\pi z}}\sin\big(\tilde{r} - \frac{l\pi}{2} - \frac{\pi}{4}\big) + ...$ $I_l(\tilde{r})$ and $K_l(\tilde{r})$ instead resemble much more an exponential function. If we compare, we notice that they decay as a function of $\tilde{r}$ even faster than an exponential. Indeed, they can be approximated for large $\tilde{r}$ as $I_l(\tilde{r}) = \frac{e^{\tilde{r}}}{\sqrt{2\pi\tilde{r}}} - f(l,\tilde{r})$ and $K_l(\tilde{r}) = \sqrt{\frac{\pi}{2\tilde{r}}}e^{-\tilde{r}} - f(l,\tilde{r})$.

Now we add to our analysis a second important ingredient. We have learnt that in the case of no free charges, both $\nabla \cdot H = 0$ and $\nabla \cdot E = 0$. This means that none of these fields can diverge (in other words, all fields must be numerically bounded). This automatically discards the Bessel functions which diverge, namely Y_l (which diverges at $\tilde{r} = 0$) and I_l (which diverges at $\tilde{r} = \infty$). Consequently, we have $B = D = 0$. This simplifies our equations to

$$U(r,\phi,z) = AJ_l(k_\perp r)e^{i(l\phi+\phi_0)}e^{i\beta z}, r \le a \tag{6.2.13}$$
$$U(r,\phi,z) = CK_l(\gamma r)e^{i(l\phi+\phi_0)}e^{i\beta z}, r > a \tag{6.2.14}$$

or any combination of $\pm l$.

Since any of the r, ϕ, z components of the E-field and H-field must take the above shape, it must hold that

$$\frac{\partial E_{\phi,r}}{\partial z} = i\beta E_{\phi,r} \tag{6.2.15}$$
$$\frac{\partial E_{\phi,r}}{\partial \phi} = il E_{\phi,r} \tag{6.2.16}$$
$$\frac{\partial H_{\phi,r}}{\partial z} = i\beta H_{\phi,r} \tag{6.2.17}$$
$$\frac{\partial H_{\phi,r}}{\partial \phi} = il H_{\phi,r} \tag{6.2.18}$$

Note that the β and l must be the same for all components and both E- and H-field!

To find the total solutions, we can now choose any two independent field components and calculate the other components from them using Maxwell equations. Knowing that we must satisfy $\nabla \times \mathbf{H} = -i\omega\epsilon_0\epsilon_r\mathbf{E}$ and $\nabla \times \mathbf{E} = i\omega\mu_0\mathbf{H}$, we simply choose, without loss of generality, the two independent fields to be E_z and H_z and impose:

$$E_z(r,\phi,z) = A_1 J_l(k_\perp r)e^{il\phi}e^{i\beta z}, r \le a \tag{6.2.19}$$
$$E_z(r,\phi,z) = C_1 K_l(\gamma r)e^{il\phi}e^{i\beta z}, r > a \tag{6.2.20}$$

and

$$H_z(r,\phi,z) = A_2 J_l(k_\perp r) e^{il\phi} e^{i\beta z}, r \leq a \tag{6.2.21}$$
$$H_z(r,\phi,z) = C_2 K_l(\gamma r) e^{il\phi} e^{i\beta z}, r > a \tag{6.2.22}$$

or any sum of $\pm l$. What would be ideal to have is that all other six components are expressed only as a function of these two. We then calculate the other six components using:

$$E_r = \frac{1}{-i\omega\epsilon_0\epsilon_r}(\frac{1}{r}\frac{\partial H_z}{\partial\phi} - \frac{\partial H_\phi}{\partial z}) \tag{6.2.23}$$
$$E_\phi = \frac{1}{-i\omega\epsilon_0\epsilon_r}(\frac{\partial H_r}{\partial z} - \frac{\partial H_z}{\partial r}) \tag{6.2.24}$$
$$H_r = \frac{1}{i\omega\mu_0}(\frac{1}{r}\frac{\partial E_z}{\partial\phi} - \frac{\partial E_\phi}{\partial z}) \tag{6.2.25}$$
$$H_\phi = \frac{1}{i\omega\mu_0}(\frac{\partial E_r}{\partial z} - \frac{\partial E_z}{\partial r}) \tag{6.2.26}$$

Using the set of derivatives of eq. 6.2.18, we can simplify

$$E_r = \frac{1}{-i\omega\epsilon_0\epsilon_r}(\frac{1}{r}\frac{\partial H_z}{\partial\phi} - i\beta H_\phi) \tag{6.2.27}$$
$$E_\phi = \frac{1}{-i\omega\epsilon_0\epsilon_r}(i\beta H_r - \frac{\partial H_z}{\partial r}) \tag{6.2.28}$$
$$H_r = \frac{1}{i\omega\mu_0}(\frac{1}{r}\frac{\partial E_z}{\partial\phi} - i\beta E_\phi) \tag{6.2.29}$$
$$H_\phi = \frac{1}{i\omega\mu_0}(i\beta E_r - \frac{\partial E_z}{\partial r}) \tag{6.2.30}$$

Reshuffling, we get a nice set of equations where all components E_r, E_ϕ, H_r and H_ϕ are expressed only in terms of derivatives of z-components E_z and H_z:

$$E_r = \frac{i}{k_0^2\epsilon_r - \beta^2}(\frac{\omega\mu_0}{r}\frac{\partial H_z}{\partial\phi} + \beta\frac{\partial E_z}{\partial r}) \tag{6.2.31}$$
$$E_\phi = \frac{i}{k_0^2\epsilon_r - \beta^2}(\frac{\beta}{r}\frac{\partial E_z}{\partial\phi} - \omega\mu_0\frac{\partial H_z}{\partial r}) \tag{6.2.32}$$
$$H_r = \frac{i}{k_0^2\epsilon_r - \beta^2}(\beta\frac{\partial H_z}{\partial r} - \frac{\omega\epsilon_0\epsilon_r}{r}\frac{\partial E_z}{\partial\phi}) \tag{6.2.33}$$
$$H_\phi = \frac{i}{k_0^2\epsilon_r - \beta^2}(\frac{\beta}{r}\frac{\partial H_z}{\partial\phi} + \omega\epsilon_0\epsilon_r\frac{\partial E_z}{\partial r}) \tag{6.2.34}$$

in addition to

$$\frac{\partial E_z}{\partial r} = A_1 k_\perp J_l'(k_\perp r) e^{il\phi} e^{i\beta z}, r \leq a \tag{6.2.35}$$

$$\frac{\partial E_z}{\partial r} = C_1 \gamma K_l'(\gamma r) e^{il\phi} e^{i\beta z}, r > a \tag{6.2.36}$$

$$\frac{\partial E_z}{\partial \phi} = ilE_z, \forall r \tag{6.2.37}$$

$$\frac{\partial E_z}{\partial z} = i\beta E_z, \forall r \tag{6.2.38}$$

and similarly

$$\frac{\partial H_z}{\partial r} = A_2 k_\perp J_l'(k_\perp r) e^{il\phi} e^{i\beta z}, r \leq a \tag{6.2.39}$$

$$\frac{\partial H_z}{\partial r} = C_2 \gamma K_l'(\gamma r) e^{il\phi} e^{i\beta z}, r > a \tag{6.2.40}$$

$$\frac{\partial H_z}{\partial \phi} = ilH_z, \forall r \tag{6.2.41}$$

$$\frac{\partial H_z}{\partial z} = i\beta H_z, \forall r \tag{6.2.42}$$

6.2.1 Boundary conditions

Now that we have all equations we need, we can apply the boundary conditions with regards to the continuity of E_z, H_z, E_ϕ and H_ϕ since they are all tangential to the interface. We have therefore at $r = a$

$$A_1 J_l(k_\perp a) = C_1 K_l(\gamma a), E_z - \text{cont.} \tag{6.2.43}$$

$$A_2 J_l(k_\perp a) = C_2 K_l(\gamma a), H_z - \text{cont.} \tag{6.2.44}$$

$$\frac{1}{k_0^2 n_1^2 - \beta^2}(\frac{\beta}{a} ilA_1 J_l(k_\perp a) - \omega\mu_0 A_2 k_\perp J_l'(k_\perp a))$$
$$= \frac{1}{k_0^2 n_2^2 - \beta^2}(\frac{\beta}{a} ilC_1 K_l(\gamma a) - \omega\mu_0 C_2 \gamma K_l'(\gamma a)), E_\phi - \text{cont.} \tag{6.2.45}$$

$$\frac{1}{k_0^2 n_1^2 - \beta^2}(\frac{\beta}{a} ilA_2 J_l(k_\perp a) + \omega\epsilon_0 n_1^2 A_1 k_\perp J_l'(k_\perp a))$$
$$= \frac{1}{k_0^2 n_2^2 - \beta^2}(\frac{\beta}{a} ilC_2 K_l(k_\perp a) + \omega\epsilon_0 n_2^2 C_1 \gamma K_l'(\gamma a)), H_\phi - \text{cont.} \tag{6.2.46}$$

By further using that $k_0^2 n_1^2 - \beta^2 = k_\perp^2$ and $k_0^2 n_2^2 - \beta^2 = -\gamma^2$, introducing generic variables $X = k_\perp a$ and $Y = \gamma a$ that trivially satisfy $X^2 + Y^2 = V^2 = k_0^2 a^2 (n_1^2 - n_2^2)$, and substituting $C_1 = \frac{A_1 J_l(k_\perp a)}{K_l(\gamma a)}$ and $C_2 = \frac{A_2 J_l(k_\perp a)}{K_l(\gamma a)}$ we simplify:

$$\frac{1}{X^2}(\beta il A_1 J_l(X) - \omega\mu_0 A_2 X J_l'(X))$$
$$= \frac{-1}{Y^2}(\beta il A_1 J_l(X) - \omega\mu_0 Y \frac{A_2 J_l(X)}{K_l(Y)} K_l'(Y)) \quad (6.2.47)$$
$$\frac{1}{X^2}(\beta il A_2 J_l(X) + \omega\epsilon_0 n_1^2 A_1 X J_l'(X))$$
$$= \frac{-1}{Y^2}(\beta il A_2 J_l(X) + \omega\epsilon_0 n_2^2 \frac{A_1 J_l(X)}{K_l(Y)} Y K_l'(Y)) \quad (6.2.48)$$

After some cumbersome reshuffling we get

$$(\frac{1}{X^2} + \frac{1}{Y^2})\beta il A_1 = \omega\mu_0 A_2 (\frac{J_l'(X)}{X J_l(X)} + \frac{K_l'(Y)}{Y K_l(Y)}) \quad (6.2.49)$$
$$(\frac{1}{X^2} + \frac{1}{Y^2})\beta il A_2 = \omega\epsilon_0 A_1 (-n_1^2 \frac{J_l'(X)}{X J_l(X)} - n_2^2 \frac{K_l'(Y)}{Y K_l(Y)}) \quad (6.2.50)$$

and FINALLY the transcendental equation, substituting $\omega^2\mu_0\epsilon_0 = k_0^2$

$$(\frac{1}{X^2} + \frac{1}{Y^2})^2 \frac{\beta^2 l^2}{k_0^2}$$
$$= (\frac{J_l'(X)}{X J_l(X)} + \frac{K_l'(Y)}{Y K_l(Y)})(n_1^2 \frac{J_l'(X)}{X J_l(X)} + n_2^2 \frac{K_l'(Y)}{Y K_l(Y)}) \quad (6.2.51)$$

This is the most generic equation that describes the possible solutions for X and Y for any a, n_1 and n_2. We see that the left and right part of the equation both depend on X and Y which are ***not*** independent variables. Instead, remember that $Y = \sqrt{V^2 - X^2}$. This means that we can simply plot again, like in the case of the slab waveguide the left and right part as a function of X (remember $\beta^2 = k_0^2 n_1^2 - \frac{X^2}{a^2}$) and find the crossings of the left and right hand side geometrically.

This allows us to find X first, and from there $k_\perp$, β and γ. We find right away that β depends on l, and that for each l several solutions, indexed by m are possible. This is part of the exercise class.

We can now solve this for the special case of $l = 0$ (in this course we will restrict our discussion to this, but note that an entire zoology of modes can exist in a fiber).

We can easily see that one of the following two equations need to be fulfilled:

$$\frac{J_0'(X)}{X J_0(X)} + \frac{K_0'(Y)}{Y K_0(Y)} = 0 \quad (6.2.52)$$

or

$$n_1^2 \frac{J_0'(X)}{X J_0(X)} + n_2^2 \frac{K_0'(Y)}{Y K_0(Y)} = 0 \quad (6.2.53)$$

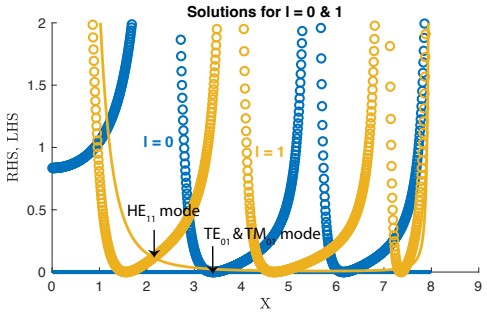

Figure 6.3: Hybrid modes as found by solving the left and right hand side (LHS, shown as full line, and RHS shown as circles) of eq. 6.2.51 of an optical fiber for $n_1 = 1.4495$, $n_2 = 1.4360$ and $a = 10$ μm, and corresponding $V = 8$ at a wavelength of $\lambda_0 = 1.55$ μm. We plot the solutions for $l = 0$ (which gives the TE and TM modes as we show below) and $l = 1$ (whose fundamental is the linearly polarised mode LP_{01}, as we shall see). Note that HE_{11} with solution at $X = 2.13$ is the fundamental mode (has the smallest X, and hence the highest effective refractive index).

6.2.2 Solutions to the Helmholtz equations

Let's now consider a bit more carefully the solutions to the Helmholtz equations, starting from the general definition of $h(\phi)$ from eq. 6.2.7. For the time being, it was very handy to work with complex solutions, since the exponentials are easy to take care of mathematically. However, since our fields need to be real, it makes sense for us to now rather look at the ***real*** solutions of eq. 6.2.7 which are evident right away:

$$h(\phi) = \begin{cases} \sin(l\phi + \phi_0), or \\ \cos(l\phi + \phi_0) \end{cases} \tag{6.2.54}$$

for $l = 0, 1, 2, 3, ...$ Once again, we restrict to positive integers in order to not overdefine our space. Trivially, these solutions can be found by imposing $A_+ = A_-$ and $A_+ = -A_-$.

In principle, based on all we discussed so far, both E_z and H_z can be any of the two, with no particular relation between the two. However, this turns out to not be the case! If we consider the dependence of any of E_r, E_ϕ, H_r and H_ϕ on H_z and E_z of equations 6.2.34, it turns out that typically the derivative $\frac{\partial}{\partial r}$ of one should yield the same as $\frac{\partial}{\partial \phi}$ of the other. This allows us to conclude that if the electric field E_z depends like $\cos(l\phi + \phi_0)$, then the magnetic field must depend like $\sin(l\phi + \phi_0)$. Hence, the solutions that actually fulfill Helmholtz equations must be of

the form:

$$E_z(r,\phi,z) = A_1 J_l(k_\perp r)\cos(l\phi+\phi_0)e^{i\beta z}, r \le a \tag{6.2.55}$$

$$E_z(r,\phi,z) = \frac{A_1 J_l(k_\perp a)}{K_l(\gamma a)} K_l(\gamma r)\cos(l\phi+\phi_0)e^{i\beta z}, r > a \tag{6.2.56}$$

and

$$H_z(r,\phi,z) = A_2 J_l(k_\perp r)\sin(l\phi+\phi_0)e^{i\beta z}, r \le a \tag{6.2.57}$$

$$H_z(r,\phi,z) = \frac{A_2 J_l(k_\perp a)}{K_l(\gamma a)} K_l(\gamma r)\sin(l\phi+\phi_0)e^{i\beta z}, r > a \tag{6.2.58}$$

These two sets of equations, together with the boundary conditions and the dependencies of all other components give all possible solutions of guided modes inside a waveguide. In the following, we identify a few mode families.

6.2.3 TE and TM modes

From this starting point, we can now consider the special case where $l = 0$.

TE modes

Let's assume for a moment that $\phi_0 = \frac{\pi}{2}$. In this case, we find that

$$E_z(r,\phi,z) = 0, r \le a \tag{6.2.59}$$

$$E_z(r,\phi,z) = 0, r > a \tag{6.2.60}$$

and

$$H_z(r,\phi,z) = A_2 J_0(k_\perp r)e^{i\beta z}, r \le a \tag{6.2.61}$$

$$H_z(r,\phi,z) = \frac{A_2 J_0(k_\perp a)}{K_0(\gamma a)} K_0(\gamma r)e^{i\beta z}, r > a \tag{6.2.62}$$

Using eq. 6.2.34 we can easily identify that in this case it follows that

$$\begin{cases} E_z = 0, E_\phi \neq 0, E_r = 0 \\ H_z \neq 0, H_\phi = 0, H_r \neq 0 \end{cases} \tag{6.2.63}$$

Hence, the electric field has only a ϕ component given by

$$E_\phi = -\frac{i\omega\mu_0}{k_0^2\epsilon_r - \beta^2}\frac{\partial H_z}{\partial r} \tag{6.2.64}$$

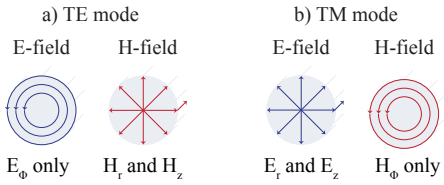

Figure 6.4: TE and TM modes in an optical fiber.

It is circulating around the z-axis and it is always tangential to the interface. Hence the name ***TE mode***. It must be conserved at the interface, i.e. at $r = a$, which leads to the following boundary condition:

$$\frac{1}{k_0^2 n_1^2 - \beta^2} k_\perp J_0'(k_\perp a) = \frac{1}{k_0^2 n_2^2 - \beta^2} \frac{J_0(k_\perp a)}{K_0(\gamma a)} \gamma K_0'(\gamma a) \tag{6.2.65}$$

$$\frac{J_0'(k_\perp a)}{k_\perp J_0(k_\perp a)} = -\frac{K_0'(\gamma a)}{\gamma K_0(\gamma a)} \tag{6.2.66}$$

Isn't this cool? We compare this with eq. 6.2.52 and find it is the same. We can now attribute this solution to the TE mode. We schematically represent it in Fig. 6.4 a.

TM modes

Let's now assume instead that $\phi_0 = 0$. In this case, we find that

$$E_z(r, \phi, z) = A_1 J_0(k_\perp r) e^{i\beta z}, r \leq a \tag{6.2.67}$$

$$E_z(r, \phi, z) = \frac{A_1 J_0(k_\perp a)}{K_0(\gamma a)} K_0(\gamma r) e^{i\beta z}, r > a \tag{6.2.68}$$

and

$$H_z(r, \phi, z) = 0, r \leq a \tag{6.2.69}$$

$$H_z(r, \phi, z) = 0, r > a \tag{6.2.70}$$

Using eq. 6.2.34 we can easily identify that in this case it follows that

$$\begin{cases} E_z \neq 0, E_\phi = 0, E_r \neq 0 \\ H_z = 0, H_\phi \neq 0, H_r = 0 \end{cases} \tag{6.2.71}$$

Hence, the magnetic field has only a ϕ component:

$$H_\phi = \frac{i\omega\epsilon_0\epsilon_r}{k_0^2\epsilon_r - \beta^2} \frac{\partial E_z}{\partial r} \tag{6.2.72}$$

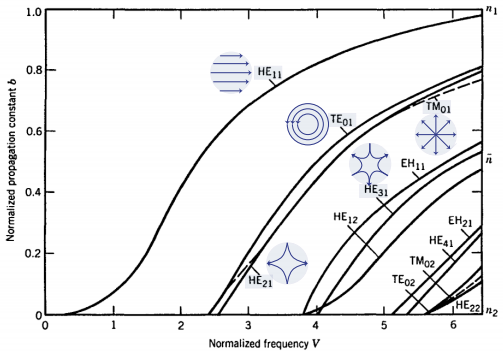

Figure 6.5: Propagation constant of various fiber modes as a function of V. The blue lines represent the electric field.

It is circulating around the z-axis and it is always tangential to the interface. Hence the name ***TM mode***. It must be conserved at the interface:

$$\frac{n_1^2}{k_0^2 n_1^2 - \beta^2} k_\perp J_0'(k_\perp a) = \frac{n_2^2}{k_0^2 n_2^2 - \beta^2} \frac{J_0(k_\perp a)}{K_0(\gamma a)} \gamma K_0'(\gamma a) \tag{6.2.73}$$

$$\frac{J_0'(k_\perp a)}{k_\perp J_0(k_\perp a)} = -\frac{n_2^2}{n_1^2} \frac{K_0'(\gamma a)}{\gamma K_0(\gamma a)} \tag{6.2.74}$$

Isn't this cool? We compare this with eq. 6.2.53 and find it is the same. We can now attribute this solution to the TM mode. The electric field instead has a radial component, as well as a z-component. We schematically represent it in Fig. 6.4 b.

Hybrid modes

All other modes that solve the transcendental equation 6.2.51 are so-called ***hybrid modes***. They exist for any $l \leq 1$ and are two-fold degenerate with respect to a rotation of $\phi = \pi/2$ around the z-axis. The exact naming is a bit complicated but results from whether it is the H_z or the E_z component that is dominant. We won't treat them in detail in this class, but please consult Fig. 6.5 for a summary of these modes.

6.2.4 Weakly guiding fibers

So far, we have kept our discussion about the exact properties of the step-index fiber general. All prior conclusions hold therefore for any type of step-index fiber, regardless of the relative refractive index Δ.

A large proportion of fibers that you will encounter in the labs however have one useful property: they typically have a low index contrast where $n_1 \simeq n_2 \simeq n_{eff} = n$, and hence $\beta \simeq nk_0$ (which is the case for most glass fibers, but may not be applicable for any fibers, e.g. semiconductor fibers!). In this case, drastic simplifications can be made to the math we derived so far.

Weakly guiding TE and TM modes for $l = 0$

For $l = 0$ we find therefore

$$\frac{J_0'(X)}{XJ_0(X)} \simeq -\frac{K_0'(Y)}{YK_0(Y)} \tag{6.2.75}$$

We can further simplify by using the identity for the derivatives of the Bessel functions

$$J_l'(X) = \frac{l}{X}J_l(X) - J_{l+1}(X) \tag{6.2.76}$$

$$K_l'(Y) = \frac{l}{Y}K_l(Y) - K_{l+1}(Y) \tag{6.2.77}$$

and find

$$\frac{J_1(X)}{XJ_0(X)} \simeq -\frac{K_1(Y)}{YK_0(Y)} \tag{6.2.78}$$

TE and TM modes have therefore almost the same effective refractive index (which makes sense, since all refractive indices are $\simeq n$, and actually even $X \simeq Y$).

Weakly guiding modes for any $l \neq 0$

We can however generalise the above discussion for any l. If the refractive index contrast between core and cladding is very small, that means that both the electric and magnetic field vectors will be ***approximately*** conserved across the interface. (The tangential component is always conserved, while the normal component of the D-field is conserved. If the refractive index is almost the same across the interface, then also the normal component of the electric field is ***approximately*** conserved and with it the electric field vector). In turn, this means that the electric field and magnetic field have ***no discontinuities***, or no jumps. This means that also their derivative is conserved. This reasoning allows us to cast all solutions to the Helmholtz equation into one single transcendental equation in which both the fields themselves and their derivatives are conserved across interfaces. Overall, this can be expressed by having their ratio being conserved at any position along the interface:

$$\frac{k_\perp J_l'(k_\perp a)}{J_l(k_\perp a)} = \frac{\gamma K_l'(k_\gamma a)}{K_l(\gamma a)} \tag{6.2.79}$$

which is equivalent to

$$\frac{XJ_l'(X)}{J_l(X)} = \frac{YK_l'(Y)}{K_l(Y)} \tag{6.2.80}$$

We simplify using the derivative properties of the Bessel functions in eq. 6.2.77 and find for any l must hold:

$$\frac{XJ_{l+1}(X)}{J_l(X)} = \frac{YK_{l+1}(Y)}{K_l(Y)} \tag{6.2.81}$$

For sake of completeness, we mention that this equation can also appear in various books in the following shape:

$$\frac{XJ_{l-1}(X)}{J_l(X)} = -\frac{YK_{l-1}(Y)}{K_l(Y)} \tag{6.2.82}$$

which one retrieves after simple math using the identities $J_{l+1} = \frac{2l}{X}J_l(X) - J_{l-1}$ and $K_{l+1} = K_{l-1} + \frac{2l}{Y}K_l$. This equation is well suited to find the linearly polarised modes of a fiber. These are modes that, much in contrast to the TE and TM modes discussed before, have predominantly one single polarisation, oriented either along the $x-$ or along the $y - axis$. These are so-called $\mathrm{LP_{lm}}$ ***modes***. The reasoning behind the nomenclature will become clear in a moment. We plot in the figure below the left and right had side of the transcendental equation above to find the solutions for $l = 0$. In this particular choise of n_1, n_2 and a, we find 3 distinct modes for one single l that we label with the index m.

The conclusion of the long story short is that altogether, this introduces yet a dependency of $u(r)$ on both m and l:

$$U(r, \phi, z) = u_{lm}(r)e^{il\phi}e^{i\beta_{lm}z} \tag{6.2.83}$$

When comparing our results to the ones of a slab waveguide, this makes sense. Also in that case, β took only discrete values, and this discreteness is translated by trivial math also to both $k_\perp$ and γ.

Now, it is important to understand the dependency of the mode profile on ϕ.

For the case that $l = 0$, the mode profile is fully symmetric around the z-axis. It turns out that the LP_{01} mode is generally the fundamental mode of the fiber since it has no cut-off. For no matter which V, this mode will always exist.

For the case that $l = 1$, the phase of the electric field sweeps from 0 to 2π during one revolution around the z-axis. We hence have the first excited mode of the fiber.

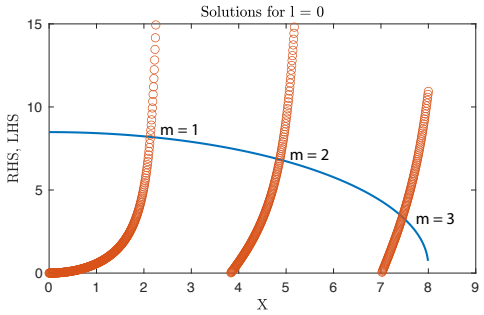

Figure 6.6: LP_{0m} modes as found by solving the left and right hand side (LHS, shown as circles, and RHS shown as full line) of eq. 6.2.81 of an optical fiber for $n_1 = 1.4495$, $n_2 = 1.4360$ and $a = 10$ μm, and corresponding $V = 8$ at a wavelength of $\lambda_0 = 1.55$ μm. We find three solutions, corresponding to $X = 2.13, 4.86$ and 7.45, respectively, that we label with yet another index $m = 1, 2, 3$. Note that LP_{01} is the fundamental mode.

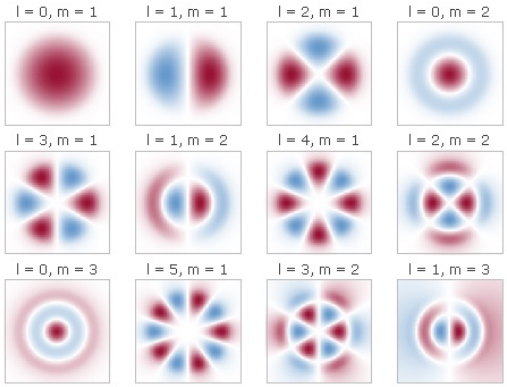

Figure 6.7: LP_{lm} modes of an optical fiber for various m and l.

For the case of $l > 1$, the phase of the electric field sweeps from 0 to $2\pi\ l$ times during one revolution around the z-axis. These are the higher excited modes of the fiber.

We plot in Fig. 6.7 the various solutions of the modes in an optical fiber.

Waveguide couplers

This chapter introduces the notion of waveguide couplers by means of grating or evanescent couplers.
Key questions: Overview:
• Which conditions need to be satisfyied to efficiently couple from fiber to chip?
• Why are edge couplers more broadband than grating couplers? What about evanescent couplers?
• How do grating couplers work? What are the advantages of forward and backward couplers?
• How do directional couplers work? What needs to be fulfilled for maximal energy transfer and why?
• What is the beating length? Which geometrical parameters influence it and in which way?
• What are eigenmodes of the coupled system?
Key concepts: mode matching, wave vector diagrams, momentum matching, beam splitter
Key equations: Forward and backward grating couplers, directional couplers, symmetric and asymmetric mode, coupling coefficient κ
Literature: Saleh and Teich, chapter 9.4

Light being efficiently guided along waveguides, with virtually no light escaping the waveguide at all opens the question about how to possibly send in light into these waveguides in first place?

There are a few approaches to efficiently couple light either from free space or from an optical fiber into a waveguide, most notably the use of ***edge couplers***, ***grating couplers*** and ***evanescent couplers***. A simplified overview is given in Fig. 7.1.

Edge couplers are typically made from a tapered waveguide which allows to shape the two-dimensional cross-section of a guided mode and terminates exactly at the edge of the chip. An optical fiber (for example a lensed fiber) can then be brought into the immediate vicinity of the edge of the chip and light scattered out of the fiber can be efficiently coupled into the waveguide. While we do not discuss this type of couplers in this lec-

ture, it should be noted that an efficient coupling is achieved if the mode of the fiber matches in all degrees of freedom (transverse mode profile, polarisation and frequency) the ones supported by the tapered waveguide. In general, edge couplers are very broad bandwidth and hence suitable for broadband applications. However, they require several complicated post-fabrication steps, such as high quality facet polishing, and high-resolution optical alignment. This renders them unsuited for wafer-level testing and high-volume manufacturing.

Evanescent couplers instead use evanescent coupling between a tapered fiber (which is typically pulled with great precision) and a on-chip waveguide. The principle behind is very similar to that of the evanescent coupling between two waveguide modes that we discuss in the next section 7.2. By careful control of the diameter of the pulled tapered fiber and its distance from the waveguide into which light needs to be coupled, near-unity coupling can be achieved. However, this technique is very tricky, as typically highly precise and stable positioners need to be employed not only to bring the fiber into the correct vicinity of the waveguide but also to keep it there for prolonged durations.

Grating couplers use corrugated waveguides to couple light from the free space into a waveguide. The performance of these couplers depends entirely on the nanoscale dimensions of the corrugated waveguides by nanofabrication techniques. We will treat their working principle in detail in the next section.

7.1 Grating couplers

The principle of grating couplers is summarised in Fig. 7.2. They are a corrugated waveguide where the height of the waveguide is altered by h with a ***periodicity*** Λ. Since we now know that the effective refractive index of optical modes depends on the transverse dimensions of the waveguide, in our case, this means that the uncorrugated waveguide has an effective index $n_{eff,1}$ and the corrugated one an effective index $n_{eff,2}$. At this point, it is also useful to introduce the concept of ***duty cycle***, which relates the

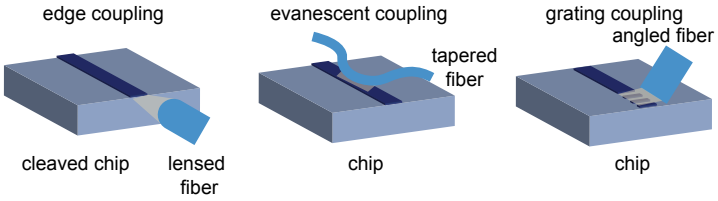

Figure 7.1: An overview of the most popular fiber-to-waveguide couplers.

length of the waveguide that is not corrugated to the total period of the grating. For sake of simplicity, we shall consider only gratings with 50 % duty cycle, meaning that the corrugated and not corrugated sections have equal length.

Owing to reciprocity, these waveguide gratings have dual purpose: the same grating can couple light in from free space ***at a specific wavelength and angle***, or, by operating in reverse, out into free space ***at the same angle and wavelength***.

For simplicity, without loss of generality, we discuss the working principle behind coupling of light from the waveguide into free space. Let's go through the grating step by step. Light that is propagating along the waveguide will be partially scattered out when it encounters the first period of the grating. The scattered wave can be approximated in first order by an elementary spherical wave. The portion of light which does not get scattered out, will continue being guided along the waveguide for a distance Λ, acquire a phase delay of $e^{i(k_0 n_{eff,1}+k_0 n_{eff,2})\frac{\Lambda}{2}}$ (note the 50% duty cycle!) and part of it will again be scattered later at the second period of the grating, leading to the emission of a second spherical wave. This process will occur as many times as the number of periods contained in the grating. ***Efficient out-coupling*** occurs if these various elementary waves interfere constructively in the far-field. In this case, the scattered field can carry energy away.

We distinguish now two different types of gratings, which need to fulfil slightly different interference conditions: ***forward couplers*** (Fig. 7.2 a) and ***backward couplers*** (Fig. 7.2 b). In all scenarios we assume the waveguide to be buried in a cladding of refractive index n_2.

The individual scattered circular waves will interfere constructively in the forward direction if the following condition is satisfied

$$k_0 n_{eff,1}\frac{\Lambda_f}{2} + k_0 n_{eff,2}\frac{\Lambda_f}{2} - k_0 n_2 \Delta d = m2\pi. \tag{7.1.1}$$

with $\Delta d = \Lambda_f \sin\theta_m$ and $m = 0, 1, ...$ Note the definition of θ in Fig. 7.2 a with respect to the normal, being a positive number. If we now search for the angle θ we find it to be dependent on the wavelength $\lambda_0 = \frac{2\pi}{k_0}$:

$$\sin\theta_m = \frac{k_0 n_{eff,1}\frac{\Lambda_f}{2} + k_0 n_{eff,2}\frac{\Lambda_f}{2} - m2\pi}{k_0 n_2 \Lambda_f} \tag{7.1.2}$$

We note that, in general, light can be scattered into several angles θ_m, called grating orders. Their number depends primarily on the periodicity Λ_f. However, we can show that, actually, $m = 0$ cannot satisfy the above condition since, in general, $n_{eff,1}, n_{eff,2} > n_2$. In addition, the equation above has only one solution for $m = 1$, if for $m = 2$ the right part of the equation becomes negative. In general, it is highly desired that light

is scattered in one single grating order, which makes its collection with a single optical fiber more straightforward.

Let's now consider the special case of $m = 1$ for the case of very shallow corrugation and hence $n_{eff,1} \simeq n_{eff,2}$:

$$k_0 n_{eff,1} \frac{\Lambda_f}{2} + k_0 n_{eff,1} \frac{\Lambda_f}{2} - k_0 n_2 \Lambda_f \sin\theta_1 = 2\pi. \tag{7.1.3}$$

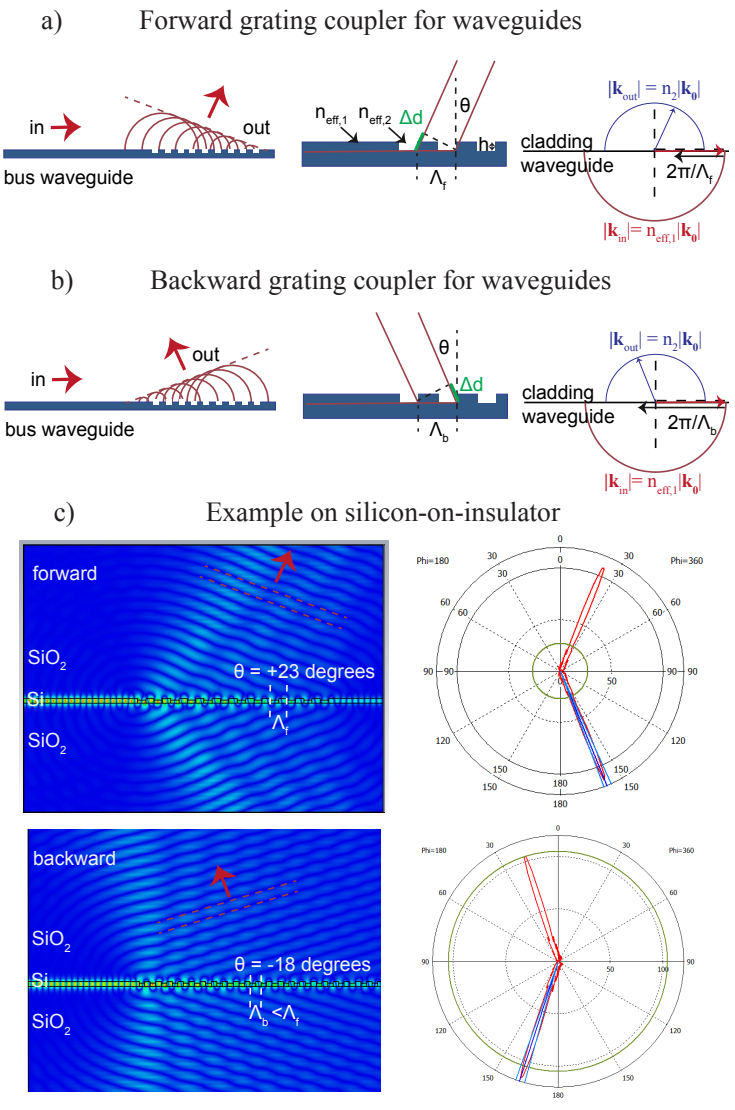

Figure 7.2: On-chip grating couplers.

and find

$$k_0 n_{eff,1} - \frac{2\pi}{\Lambda_f} = k_0 n_2 \sin\theta_1. \tag{7.1.4}$$

The meaning of this equation can be understood graphically from the principle of conservation of momentum, sketched in the right panel of Fig. 7.2 a. The grating provides the necessary momentum to light to be scattered out of the waveguide.

In a second step, let's also analyse the case of backward couplers. The individual scattered circular waves will interfere constructively in the backward direction if the following condition is satisfied

$$k_0 n_{eff,1} \frac{\Lambda_b}{2} + k_0 n_{eff,2} \frac{\Lambda_b}{2} + k_0 n_2 \Delta d = m 2\pi. \tag{7.1.5}$$

with $\Delta d = \Lambda_b \sin\theta_m$ and $m = 0, 1, ...$ Note the definition of θ in Fig. 7.2 b with respect to the normal, being a positive number. If we now search for the angle θ we find it to be dependent on the wavelength $\lambda_0 = \frac{2\pi}{k_0}$:

$$\sin\theta_m = \frac{m2\pi - k_0 n_{eff,1} \frac{\Lambda_b}{2} - k_0 n_{eff,2} \frac{\Lambda_b}{2}}{k_0 n_2 \Lambda_b} \tag{7.1.6}$$

Yet again, $m = 0$ provides no solution to this equation. Searching for the solution for $m = 1$, we find

$$\frac{2\pi}{\Lambda_b} - k_0 n_{eff,1} = k_0 n_2 \sin\theta_1. \tag{7.1.7}$$

Also this equation can be understood from the principle of conservation of momentum: the corrugated waveguide provides a momentum to the guided light that is sufficient to scatter it out into an angle θ at wavelength λ_0. The corresponding graphical representation of the momentum conservation is shown in the right panel of Fig. 7.2 b.

If we now compare the conditions on the periodicity for forward and backward grating couplers, we find that the periodicity for backward gratings is typically smaller than the one for forward gratings at the same wavelength, duty cycle and angle θ: $\Lambda_b < \Lambda_f$.

We show finite element simulations for the two types of grating couplers in Fig. 7.2 c for a silicon waveguide corrugated by $h = 80$ nm and a duty cycle of 50 % for two grating periods of $\Lambda_f = 770$ nm and $\Lambda_b = 520$ nm. When the guidd wave reaches the grating, part of it is scattered out. The interference if the various scattered waves build up an overall wavefront (shown with the red dashed line) that evidently propagates at a well-defined angle of $\theta = 23°$ for the forward gratings and $\theta = 18°$ for the backward gratings. This is confirmed by the farfield plots shown in the two right panels of Fig. 7.2 c.

In general, the width of the grating must be on the order of the size of the optical fiber mode. For this reason, an adiabatic taper is required to ensure a smooth transition from a typical waveguide towards the grating coupler. This adiabatic taper must have such a small angle that it does

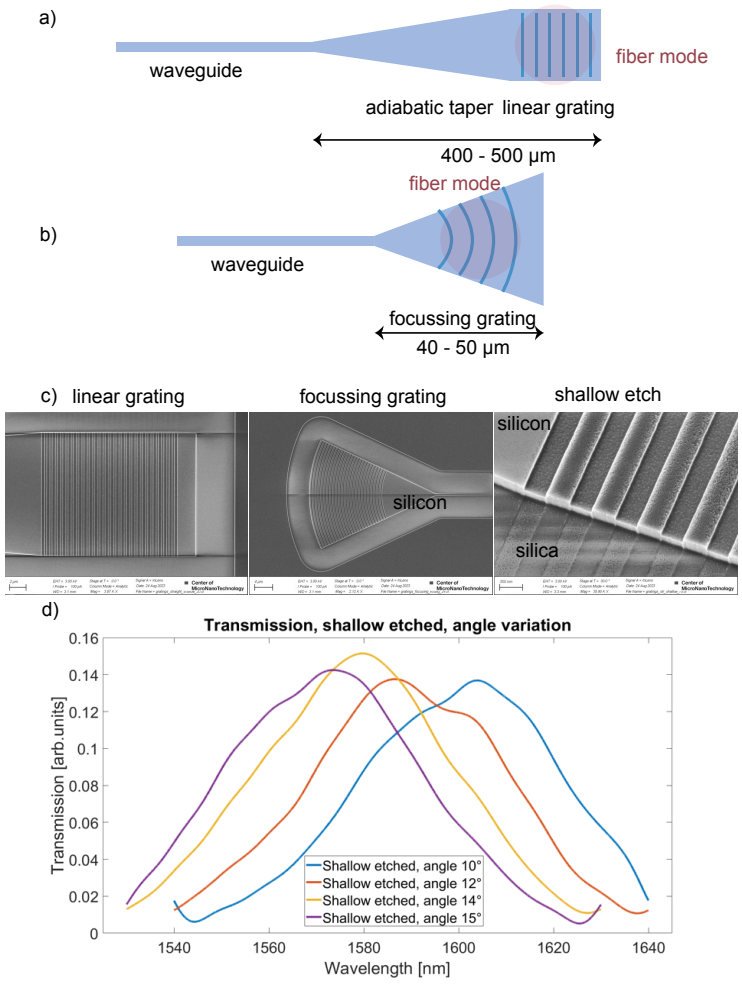

Figure 7.3: The two most prevalent types of grating couplers are straight linear gratings (a) and focusing gratings (b), here shown for a burried silicon waveguide with silica cladding. The advantage of focusing gratings is their compact size. In c) we show scanning electron microscope images of these gratings. Here we used a shallow etch of 80 nm into silicon. d) Typical tranmission through a waveguide with in- and out-coupling focusing tapers shows a characteristic bandwidth of 40-50 nm and its shifting with angle.

not significantly distort the optical beam. The phasefronts that arrive at the straight grating must be planar. On the other hand, focusing gratings can be much more compact. In that case, a short taper with a large angle is used, which leads to rounded phase fronts. The gratings need then to follow these phase fronts and are also circular, as shown in Fig. 7.3 b. In Fig. 7.3 c you can see scanning electron microscope images of such devices fabricated in our lab.

7.2 Directional couplers

In this section we will treat the case of ***directional couplers*** that make use of the coherent transfer of energy from one waveguide into another waveguide that is brought into its immediate vicinity. The evanescent tail of the optical field guided in one waveguide is overlapping with the core region of the other waveguide. As a result, this field starts building up a guided mode in the second waveguide, and the latter becomes stronger and stronger as a function of propagation length. A necessary requirement is however, that the coupled fields ad up constructively as the wave builds up in the second waveguide, see Fig. 7.4 a and b. We speak in this case about a form of phase matching.

Let's consider the special case depicted in Fig. 7.4 b where two waveguides are in the immediate vicinity of each other, such that the spatial overlap of a guided optical field inside one waveguide with the second waveguide is non-negligible. In this case, as the guided waveguide modes depend strongly on the space variable $\mathbf{r}$, it makes sense to consider the problem in the three-dimensional space. We can model the two-waveguide system as being one where $\mathbf{E}_a(\mathbf{r},t)$ is the guided mode in waveguide a and $\mathbf{E}_b(\mathbf{r},t)$ is the guided mode in waveguide b, see Fig. 7.4 a.

In the case of ***weak coupling***, one can assume that a part of the energy of the optical field initially guided in waveguide a will leak into waveguide b and vice-versa. The coupling constants κ_{ba} and κ_{ab} quantify the electric field that leaks from waveguide a into waveguide b and vice-versa per unit length. In the following we assume, without loss of generality, that $\kappa_{ba} = \kappa_{ab} = \kappa$. Weak coupling means that this rate is much smaller than the propagation constants of the relevant optical fields: $\kappa_{ab} << \beta_a$ and $\kappa_{ba} << \beta_b$. In this scenario, we can write set of coupled differential equations that describe the entire system as a function of space as:

$$\frac{d\mathbf{E}_a(\mathbf{r},t)}{dz} = i\beta_a\mathbf{E}_a(\mathbf{r},t) - i\kappa\mathbf{E}_b(\mathbf{r},t) \tag{7.2.1}$$

$$\frac{d\mathbf{E}_b(\mathbf{r},t)}{dz} = i\beta_b\mathbf{E}_b(\mathbf{r},t) - i\kappa\mathbf{E}_a(\mathbf{r},t). \tag{7.2.2}$$

When the two propagation constants are the same $\beta_a = \beta_b = \beta$ in the two waveguides - the system of coupled waveguides can be rewritten as

$$\frac{d\mathbf{E}_a(\mathbf{r},t)}{dz} = i\beta\mathbf{E}_a(\mathbf{r},t) - i\kappa\mathbf{E}_b(\mathbf{r},t) \tag{7.2.3}$$

$$\frac{d\mathbf{E}_b(\mathbf{r},t)}{dz} = i\beta\mathbf{E}_b(\mathbf{r},t) - i\kappa\mathbf{E}_a(\mathbf{r},t). \tag{7.2.4}$$

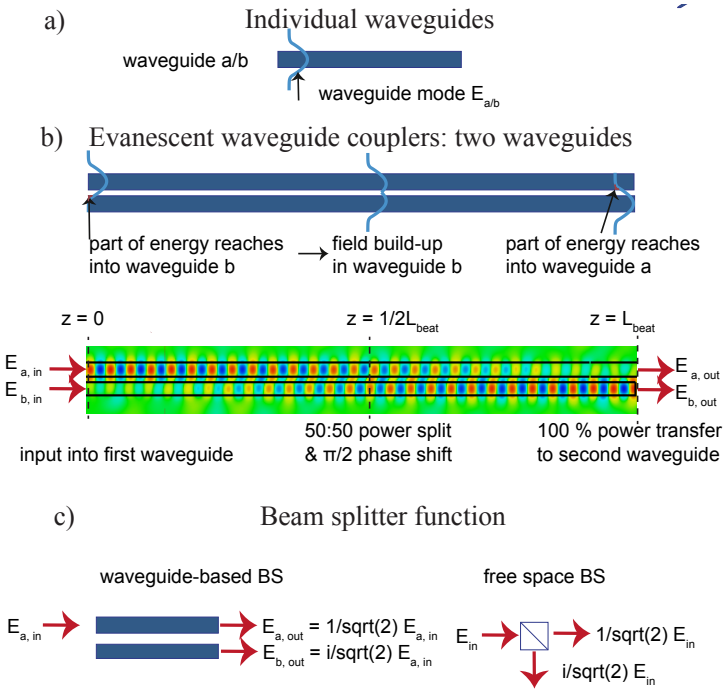

Figure 7.4: Evanescent waveguide couplers (directional couplers). These couplers are realised by bringing two otherwise separated dielectric waveguides (a)) into each other's vicinity (b). In this way, the optical field guided inside one of the waveguides excites a guided mode inside the second waveguide. After a length L the energy is transferred from waveguide a to waveguide b. The amplitude and phase of the light guided in each waveguide $\begin{pmatrix} E_{a,out} \\ E_{b_out} \end{pmatrix}$ can be described at any position z through the transfer matrix 7.2.32 and the input conditions $\begin{pmatrix} E_{a,in} \\ E_{b,in} \end{pmatrix}$. d) It turns out that the transfer matrix is equivalent to the one of a beam splitter. At $z = \frac{1}{2}L_{beat}$, the transfer matrix is equivalent to a 50:50 beam splitter.

and we can either sum them up or subtract the them to get:

$$\frac{d(\mathbf{E}_a(\mathbf{r},t)+\mathbf{E}_b(\mathbf{r},t))}{dz} = i(\beta-\kappa)(\mathbf{E}_a(\mathbf{r},t)+\mathbf{E}_b(\mathbf{r},t)) \tag{7.2.5}$$

$$\frac{d(\mathbf{E}_a(\mathbf{r},t)-\mathbf{E}_b(\mathbf{r},t))}{dz} = i(\beta+\kappa)(\mathbf{E}_a(\mathbf{r},t)-\mathbf{E}_b(\mathbf{r},t)). \tag{7.2.6}$$

This means that the sum or the difference of the two modes

$$\mathbf{E}_{s,as}(\mathbf{r},t) = \frac{1}{2}(\mathbf{E}_a(\mathbf{r},t) \pm \mathbf{E}_b(\mathbf{r},t)). \tag{7.2.7}$$

are now modes of the new coupled system, with corresponding propagation constants $\beta_{as,s} = \beta \pm \kappa$.

Let's analyse the particular case of two identical waveguides located at positions $y = \pm\frac{h}{2} := \pm(a+\frac{d}{2})$ that guide the same optical mode with field profile $\mathbf{E}_a(\mathbf{r},t) = \mathbf{E}_b(\mathbf{r},t)$. The subscripts will hence only be used to stress out whether the mode is located in waveguide a or in waveguide b.

We note that $\mathbf{E}_s(\mathbf{r},t)$ is a symmetric mode and $\mathbf{E}_{as}(\mathbf{r},t)$ is an asymmetric mode. Both of them will be guided with no loss if launched into the waveguide system.

The next interesting question is what will happen if we launch light into only one of the two waveguides, say into waveguide a? This is equivalent to

$$\mathbf{E}_{in} = \mathbf{E}_a(y, z=0). \tag{7.2.8}$$

As we will rigorously derive below, the optical field will be oscillating between the two waveguides if e.g. we feed an optical lightfield only into one of the two waveguides, as shown in Fig. 7.4 b. The length over which the lightfield oscillates once back and forth between the two waveguides depends strongly on the difference in effective refractive index for the symmetric and antisymmetric modes. Let's call it L_{period}.

To demonstrate the coherent transfer from one waveguide to another, let's consider the following. An input mode localized in one of the waveguides at $z = 0$ can be written as a superposition of the symmetric and antisymmetric modes with equal weights:

$$\mathbf{E}_{in}(y,z=0) = \mathbf{E}_s(y,z=0) + \mathbf{E}_{as}(y,z=0) = \mathbf{E}_a(y,z=0). \tag{7.2.9}$$

Each of these modes will propagate with their own propagation constants $\beta_{s,as}$. Hence, at a distance z the total electric field will be

$$\mathbf{E}_{out}(y,z) = \mathbf{E}_s(y,z=0)e^{i\beta_s z} + \mathbf{E}_{as}(y,z=0)e^{i\beta_{as}z} \tag{7.2.10}$$

$$= e^{i\beta_s z}(\mathbf{E}_s(y,z=0) + \mathbf{E}_{as}(y,z=0)e^{i(\beta_{as}-\beta_s)z}). \tag{7.2.11}$$

To find L_{period}, we must satisfy that the total field at $z = L_{period}$ is the same as at $z = 0$ up to an overall phase: $\mathbf{E}_{out}(y, z = L_{period}) =$

$e^{i\phi}\mathbf{E}_{in}(y, z = 0)$. From here, we find that

$$(\beta_{as} - \beta_s)L_{period} = 2\pi \tag{7.2.12}$$

$$L_{period} = \frac{2\pi}{\Delta\beta} = \frac{\lambda_0}{\Delta n} = \frac{\pi}{\kappa} \tag{7.2.13}$$

where $\Delta n = n_{eff,1} - n_{eff,2}$.

7.2.1 Coupling from one waveguide into a second waveguide

Let's now consider a second special case, where the two coupled waveguides stop after a length of $z = L_{beat} = \frac{L_{period}}{2}$. We find

$$\mathbf{E}_{out}(y, z = L_{beat}) = \mathbf{E}_s(y, z = 0)e^{i\beta_s L_{beat}} + \mathbf{E}_{as}(x, y, z = 0, t)e^{i\beta_{as} L_{beat}} \tag{7.2.14}$$

$$= e^{i\beta_s L_{beat}}(\mathbf{E}_s(y, z = 0) + \mathbf{E}_{as}(y, z = 0)e^{i(\beta_{as}-\beta_s)L_{beat}}) \tag{7.2.15}$$

$$= e^{i\beta_s L_{beat}}(\mathbf{E}_s(y, z = 0) - \mathbf{E}_{as}(y, z = 0)) \tag{7.2.16}$$

$$= e^{i\beta_s L_{beat}}\mathbf{E}_b(y, z = 0). \tag{7.2.17}$$

Expressed in the basis of the modes of each waveguide, this means that an input mode localised in the first waveguide, is transformed into an output mode localised in the second waveguide after a length L_{beat}. This is the so-called ***transfer length***, also ***beating length***.

7.2.2 Beam splitter with 50:50 splitting ratio

Let's now consider a third special case, where the two coupled waveguides stop after a length of $z = L_{beat}/2 = \frac{L_{period}}{4}$. We find

$$\mathbf{E}_{out}(y, z = L_{beat}/2) = \mathbf{E}_s(y, z = 0)e^{i\beta_s L_{beat}/2} + \mathbf{E}_{as}(y, z = 0)e^{i\beta_{as} L_{beat}/2} \tag{7.2.18}$$

$$= e^{i\beta_s L_{beat}/2}(\mathbf{E}_s(y, z = 0) + \mathbf{E}_{as}(y, z = 0)e^{i(\beta_{as}-\beta_s)L_{beat}/2}) \tag{7.2.19}$$

$$= e^{i\beta_s L_{beat}/2}(\mathbf{E}_s(y, z = 0) + i\mathbf{E}_{as}(y, z = 0)) \tag{7.2.20}$$

$$= \sqrt{2}e^{i\beta_s L_{beat}/2}e^{i\frac{\pi}{4}}(\mathbf{E}_a(y, z = 0) - i\mathbf{E}_b(y, z = 0)). \tag{7.2.21}$$

Expressed in the basis of the modes of each waveguide, this means that an input mode localised in the first waveguide is transformed into an output mode localised in both waveguides with equal weight and a mutual phase delay of $\frac{\pi}{2}$ after a length $L_{beat}/2$.

We can follow a similar procedure for an input field

$$\mathbf{E}_{in} = \mathbf{E}_b(y, z = 0). \tag{7.2.22}$$

and find after a length of $z = L_{beat}/2 = \frac{L_{period}}{4}$ in this case

$$\mathbf{E}_{out}(y, z = L_{beat}/2) = \mathbf{E}_s(y, z = 0)e^{i\beta_s L_{beat}/2} - \mathbf{E}_{as}(y, z = 0)e^{i\beta_{as} L_{beat}/2} \tag{7.2.23}$$

$$= e^{i\beta_s L_{beat}/2}(\mathbf{E}_s(y, z = 0) - \mathbf{E}_{as}(y, z = 0)e^{i(\beta_{as}-\beta_s)L_{beat}/2}) \tag{7.2.24}$$

$$= e^{i\beta_s L_{beat}/2}(\mathbf{E}_s(y, z = 0) - i\mathbf{E}_{as}(y, z = 0)) \tag{7.2.25}$$

$$= \sqrt{2}e^{i\beta_s L_{beat}/2}e^{i\frac{\pi}{4}}(-i\mathbf{E}_a(y, z = 0) + \mathbf{E}_b(y, z = 0)). \tag{7.2.26}$$

Overall, we can write the total fields in this case (omitting the $\sqrt{2}$) in matrix form as

$$\begin{pmatrix} \mathbf{E}_{a,out} \\ \mathbf{E}_{b,out} \end{pmatrix} = \begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix} \begin{pmatrix} \mathbf{E}_{a,in} \\ \mathbf{E}_{b,in} \end{pmatrix} \tag{7.2.27}$$

which corresponds exactly to the matrix transfer function of a ***50:50 beam splitter*** you may be more familiar with from free space. This means that waveguide couplers on-chip fulfill the exact same function as beam splitters in free space, as graphically shown in Fig. 7.4 c. Isn't this extremely cool?

7.2.3 Beam splitter with arbitrary splitting ratio

Finally, let's derive what happens for an arbitrary length z and any arbitrary input field, given that $\beta_{as} - \beta_s = 2\kappa$. For this, we assume that any physical field propagating along our two coupled waveguides will always be an arbitrary superposition of the symmetric and antisymmetric modes:

$$\mathbf{E}_{out}(y, z) = a\mathbf{E}_s(y, z = 0)e^{i\beta_s z} + b\mathbf{E}_{as}(y, z = 0)e^{i\beta_{as} z} \tag{7.2.28}$$

$$= e^{i\beta_s z}(a\mathbf{E}_s(y, z = 0) + b\mathbf{E}_{as}(y, z = 0)e^{i2\kappa z}) \tag{7.2.29}$$

$$= e^{i\beta z}(a\mathbf{E}_s(y, z = 0)e^{-i\kappa z} + b\mathbf{E}_{as}(y, z = 0)e^{i\kappa z}) \tag{7.2.30}$$

Replacing now with the guided modes of the isolated waveguides, we find

$$\mathbf{E}_{out}(y, z) = \frac{\mathbf{E}_a(y, z = 0)}{2}e^{i\beta z}((a + b)\cos(\kappa z) - i(a - b)\sin(\kappa z)) + \frac{\mathbf{E}_b(y, z = 0)}{2}e^{i\beta z}((a - b)\cos(\kappa z) - i(a + b)\sin(\kappa z)) \tag{7.2.31}$$

Defining the input fields $\mathbf{E}_{a,in} = \mathbf{E}_a(y, z = 0)\frac{a+b}{2}$ and $\mathbf{E}_{b,in} = \mathbf{E}_b(y, z = 0)\frac{a-b}{2}$ we can write a generalized transfer matrix for any z that relates

the output fields in the two waveguides to the input fields in the two waveguides:

$$\begin{pmatrix}\mathbf{E}_{a,out}\\ \mathbf{E}_{b,out}\end{pmatrix} = e^{i\beta z}\begin{pmatrix}\cos(\kappa z) & -i\sin(\kappa z)\\ -i\sin(\kappa z) & \cos(\kappa z)\end{pmatrix}\begin{pmatrix}\mathbf{E}_{a,in}\\ \mathbf{E}_{b,in}\end{pmatrix} \tag{7.2.32}$$

We can therefore interpret $\cos(\kappa z) = r$ as a reflection coefficient of the electric field into the same waveguide, and $-\sin(\kappa z) = t$ as a transmission coefficient of the electric field from one waveguide into the other:

$$\begin{pmatrix}\mathbf{E}_{a,out}\\ \mathbf{E}_{b,out}\end{pmatrix} = \begin{pmatrix}r & it\\ it & r\end{pmatrix}\begin{pmatrix}\mathbf{E}_{a,in}\\ \mathbf{E}_{b,in}\end{pmatrix} \tag{7.2.33}$$

This interpretation will be very useful in the context of ring resonators that we will discuss in chapter 9.2.

On a final note, we underline that generally it is not a necessary condition that the two modes in the two waveguides are the exact same spatial mode (e.g. corresponding to $u_0(y)$). The same description applies also if $\mathbf{E}_a(\mathbf{r},t) \neq \mathbf{E}_b(\mathbf{r},t)$ (e.g. they could correspond to any $u_n(y)$ and any other $u_m(y)$) and we will have a symmetric and an anti-symmetric eigenstate if the two propagation constants are equal. In this particular case, we can therefore couple modes of distinct transverse order.

7.2.4 Coupling constant κ

In the previous section we introduced the coupling constant κ rather phenomenologically. We could derive that many important parameters such as the beating length L_{beat} depend on it. We could find that the amplitude of the modes in each of the coupled waveguides may change as a function of z, while their shape does not change. This warrants an Ansatz for the corresponding guided fields to look just like guided modes in isolated waveguides ***but*** with a z-dependent amplitude. We write therefore for the field in each of the two waveguides

$$\mathbf{E}_a(y,z) = E_a(y,z)\mathbf{e}_a = e_a(z)u_a(y)e^{i\beta_a z}\mathbf{e}_a \tag{7.2.34}$$

$$\mathbf{E}_b(y,z) = E_b(y,z)\mathbf{e}_b = e_b(z)u_b(y)e^{i\beta_b z}\mathbf{e}_b \tag{7.2.35}$$

where $u_a(y)$ and $u_b(y)$ are the transversal profiles of the corresponding guided modes. $e_a(z)$ and $e_b(z)$ are the normalised amplitudes and describe the evolution as a function of z (i.e. in the case of $\kappa = 0$ the normalised amplitude of the waves would not evolve with z and we'd have trivially $e_a(z) = e_b(z) = 1$). Importantly, each of the fields $E_a(y,z)$ and $E_b(y,z)$ fulfils the Helmholtz equation since they are guided modes of the isolated waveguides:

$$\nabla^2 E_{a,b}(y,z) + k_0^2\epsilon(y)E_{a,b}(y,z) = 0 \tag{7.2.36}$$

With this ansatz we can simplify the coupled equations 7.2.2 to:

$$\frac{\partial}{\partial z}e_a(z) = -i\underbrace{\kappa\int dy u_b(y)u_a^*(y)}_{\tilde{\kappa}}e^{i(\beta_b-\beta_a)z}e_b(z) \tag{7.2.37}$$

$$\frac{\partial}{\partial z}e_b(z) = -i\underbrace{\kappa\int dy u_a(y)u_b^*(y)}_{\tilde{\kappa}^*}e^{-i(\beta_b-\beta_a)z}e_a(z) \tag{7.2.38}$$

In the following, we will understand how κ and $\tilde{\kappa}$ depend on the exact geometry of our system. For this, let's consider the geometry of two coupled slab waveguides with core refractive indices $\sqrt{\epsilon_1}$ and $\sqrt{\epsilon_{1,b}}$ and cladding refractive index $\sqrt{\epsilon_2}$ in Fig. 7.5. For the reminder, we assume $\epsilon_{1,a} = \epsilon_{1,b} = \epsilon_1$. Guided light along waveguide a with corresponding electric field E_a will introduce in waveguide b a ***perturbative polarisation field*** $\delta\mathbf{P}_a = \epsilon_0\delta\epsilon_a\mathbf{E}_a$ with $\delta\epsilon_a = \epsilon_1 - \epsilon_2$ the perturbation of the refractive index effected by bringing waveguide a into the vicinity of waveguide b. This polarisation field ***will act as a source*** towards the generation of a new field inside waveguide b according to the following formula that we derive in section 7.3:

$$\nabla^2\mathbf{E}_b + k_0^2\epsilon(y)\mathbf{E}_b = -\omega^2\mu_0\delta\mathbf{P}_a \tag{7.2.39}$$

Through the symmetry of the problem we can also write an equivalent equation for $\mathbf{E}_a$, keeping in mind that $\delta\epsilon_b = \delta\epsilon_a = \epsilon_1 - \epsilon_2$:

$$\nabla^2\mathbf{E}_a + k_0^2\epsilon(y)\mathbf{E}_a = -\omega^2\mu_0\delta\mathbf{P}_b \tag{7.2.40}$$

and find by assuming $\frac{\partial^2}{\partial z^2}e_{a,b} << \beta_{a,b}\frac{\partial}{\partial z}e_{a,b}$ and the polarisation of the two fields to be the same $\mathbf{e}_a \parallel \mathbf{e}_b$

$$2i\beta_b u_b(y)\frac{\partial}{\partial z}e_b(z) = -\omega^2\delta\epsilon_a u_a(y)e_a(z)e^{i(\beta_a-\beta_b)z}| \times \int dy u_b^*(y) \tag{7.2.41}$$

$$2i\beta_a u_a(y)\frac{\partial}{\partial z}e_a(z) = -\omega^2\delta\epsilon_b u_b(y)e_b(z)e^{i(\beta_b-\beta_a)z}| \times \int dy u_a^*(y) \tag{7.2.42}$$

$$\frac{\partial}{\partial z}e_b(z) = -i\underbrace{\frac{\omega^2\delta\epsilon_a}{-2\beta_b}\int dy u_a(y)u_b^*(y)}_{\tilde{\kappa}}e_a(z)e^{i(\beta_a-\beta_b)z} \tag{7.2.43}$$

$$\frac{\partial}{\partial z}e_a(z) = -i\underbrace{\frac{\omega^2\delta\epsilon_b}{-2\beta_a}\int dy u_b(y)u_a^*(y)}_{=\tilde{\kappa}^* \text{ if } \epsilon_{1,a}=\epsilon_{1,b}\&\beta_a\approx\beta_b}e_b(z)e^{-i(\beta_a-\beta_b)z} \tag{7.2.44}$$

We see clearly that the coupling constant $\tilde{\kappa}$ depends on a few parameters, notably on the ***overlap integral*** between the two guided modes in the two waveguides. The overlap integral itself depends on the mode numbers, on the distances between the two waveguides, on the polarisation of the guided modes, among others. We also see that ***phase matching*** plays a role through the term $e^{i(\beta_b-\beta_a)z}$, not discussed in this lecture.

7.3 Wave equation with perturbative polarisation fields

In this section, we derive briefly the wave equation in the case of perturbative polarisation fields. Imagine you have an electric field incident into a material of permittivity $\chi^{(1)}$ and that the polarisation field generated in this case is

$$\tilde{\mathbf{P}}(\mathbf{r},\omega) = \epsilon_0\tilde{\chi}^{(1)}(\mathbf{r},\omega)\tilde{\mathbf{E}}(\mathbf{r},\omega) + \delta\tilde{\mathbf{P}}(\mathbf{r},\omega) = \tilde{\mathbf{P}}^{(1)}(\mathbf{r},\omega) + \delta\tilde{\mathbf{P}}(\mathbf{r},\omega) \quad (7.3.1)$$

Consequently, $\tilde{\mathbf{D}}(\mathbf{r},\omega) = \epsilon_0\tilde{\mathbf{E}}(\mathbf{r},\omega) + \tilde{\mathbf{P}}(\mathbf{r},\omega)$. In addition to the standard terms, another polarisation field is generated. Using this definition, we find by using the same procedure as in section 4.2 that

$$\begin{aligned}\nabla^2\tilde{\mathbf{E}}(\mathbf{r},\omega) &= -\omega^2\mu_0\tilde{\mathbf{D}}(\mathbf{r},\omega)\\ &= -\omega^2\mu_0(\epsilon_0\tilde{\mathbf{E}}(\mathbf{r},\omega) + \tilde{\mathbf{P}}^{(1)}(\mathbf{r},\omega) + \delta\tilde{\mathbf{P}}(\mathbf{r},\omega)) \quad (7.3.2)\end{aligned}$$

and hence

$$\begin{aligned}\nabla^2\tilde{\mathbf{E}}(\mathbf{r},\omega) + k_0^2\tilde{\mathbf{E}}(\mathbf{r},\omega) &= -\omega^2\mu_0(\tilde{\mathbf{P}}^{(1)}(\mathbf{r},\omega) + \delta\tilde{\mathbf{P}}(\mathbf{r},\omega))\\ \nabla^2\tilde{\mathbf{E}}(\mathbf{r},\omega) + k_0^2\epsilon(\omega)\tilde{\mathbf{E}}(\mathbf{r},\omega) &= -\omega^2\mu_0\delta\tilde{\mathbf{P}}(\mathbf{r},\omega)\end{aligned}$$

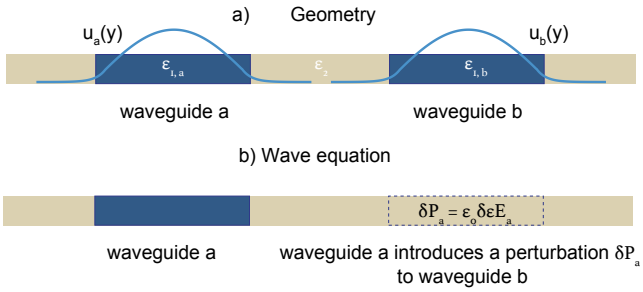

Figure 7.5: Coupling constant $\tilde{\kappa}$. To derive the coupling constant $\tilde{\kappa}$, the following train of thoughts may be applied. Waveguide a can be treated as a perturbation to waveguide b. This means that any light E_a propagating through waveguide a can be regarded as generating a polarisation field $\delta\mathbf{P}_a = \epsilon_0\delta\epsilon\mathbf{E}_a$ that acts as a source field to a guided mode inside waveguide b.

Multimode interference devices

This chapter introduces the notion of multimode interference in wide waveguides. It is used for the realisation of broadband x- and y-couplers, crossed waveguides and, more generally $M \times N$ couplers.
Key questions: Overview:
• What is an MMI?
• What is self-imaging?
• What are mirror images?
• Which considerations are important for the design of an MMI?
• How does a x-coupler from an MMI work and how does it compare to a directional evanescent coupler?
Key concepts: mode overlap, self-imaging, beating length
Key equations: beating length, effective refractive index, phase delay
Literature: Soldano et al. "Optical Multi-Mode Interference Devices Based on Self-Imaging: Principles and Applications", JLT 13, 1995 paper

8.1 Introduction

So far, we have mostly considered integrated structures that are single mode in nature - only the fundamental mode (TE or TM) propagates without loss inside the waveguide. Many optical components that you will encounter are single mode in this sense, and they benefit from the fact that the intensity distribution in the transversal plane is well-defined at any position of the waveguide.

In this chapter, we will increase the number of guided modes from 1 to N - and enter the realm of multi-mode photonics. We will show that this allow us to do two different things. On one hand, we can launch any linear combination of all guided modes into a waveguide. This can be beneficial, for example, to pack as much information as possible into a single channel.

On the other hand, we can exploit the fact that the N guided modes propagate with different phase velocities $v_{ph,m} = \frac{c_0}{n_{eff,m}}$ although they have the same angular frequency ω. This leads to a spatial interference pattern that depends on the exact location along the waveguide. We can then exploit this to our advantage - and realise devices that concentrate light in a chosen region of the waveguide itself.

8.2 Self-imaging principle

To begin, we will consider the most generic case of a multimode step-index slab waveguide of width w_{wg} and refractive index contrast $\Delta n = n_1 - n_2$, with n_1 the refractive index of the core and n_2 the refractive index of the cladding. We assume this waveguide to support N transverse modes. Without loss of generality, we consider these modes to be TE-polarised.

Self-imaging is a property of multimode waveguides that allows to reproduce an input field $\mathbf{E}(x, y, z = 0, t)$ at multiple periodic locations $z = pL$ along the multi-mode waveguide, down-stream from the input. In the following, we will demonstrate mathematically which conditions need to be satisfied to achieve self-imaging. For this, we start from equation 5.1.18 that shows that an arbitrary guided field can be written as a linear combination of all TE modes with arbitrary amplitudes a_m and corresponding propagation constants β_m:

$$\mathbf{E}(y, z, t) = \sum_{m=1}^{N} a_m u_m(y) e^{-i(\omega t - \beta_m z)} \mathbf{e}_x \tag{8.2.1}$$

and impose that:

$$\mathbf{E}(y, z = pL, t) = \mathbf{E}(y, z = 0, t) \tag{8.2.2}$$

Clearly, the equation above is satisfied if for all β_m supported in the multimode waveguide, there exists an integer number q_m (which can be different for the different β_m), such that $\beta_m pL = q_m 2\pi,\ \forall m$.

In the most general case, we have seen that the propagation constants of guided modes depend in a non-trivial way on the refractive indices n_1 and n_2. This also means that it can be quite challenging to satisfy the condition above if many modes are involved. We have seen in chapter 5 that the propagation constant must satisfy the condition that $(\frac{(m+1)\pi}{d_m})^2 + \beta_m^2 = n_1^2 k_0^2$. $m = 0$ corresponds to the fundamental transverse mode, even m correspond to symmetric modes, and uneven m correspond to antisymmetric modes with respect to $y = 0$. If we now further assume that $d_m = d$ is mode-independent and equal to the geometrical waveguide width (which is reasonable for high index contrast, wide waveguides). In

addition, we can approximate that the propagation constant $n_1 k_0 >> \frac{(m+1)\pi}{d_m}$ and find

$$\beta_m = \sqrt{n_1^2 k_0^2 - \frac{(m+1)^2\pi^2}{d^2}} = n_1 k_0 (1 - \frac{(m+1)^2\pi^2}{2d^2 n_1^2 k_0^2}) \tag{8.2.3}$$

and from here we find that

$$\beta_m - \beta_{m-1} \approx -\frac{(2m+1)\pi^2}{2d^2 n_1 k_0}. \tag{8.2.4}$$

and

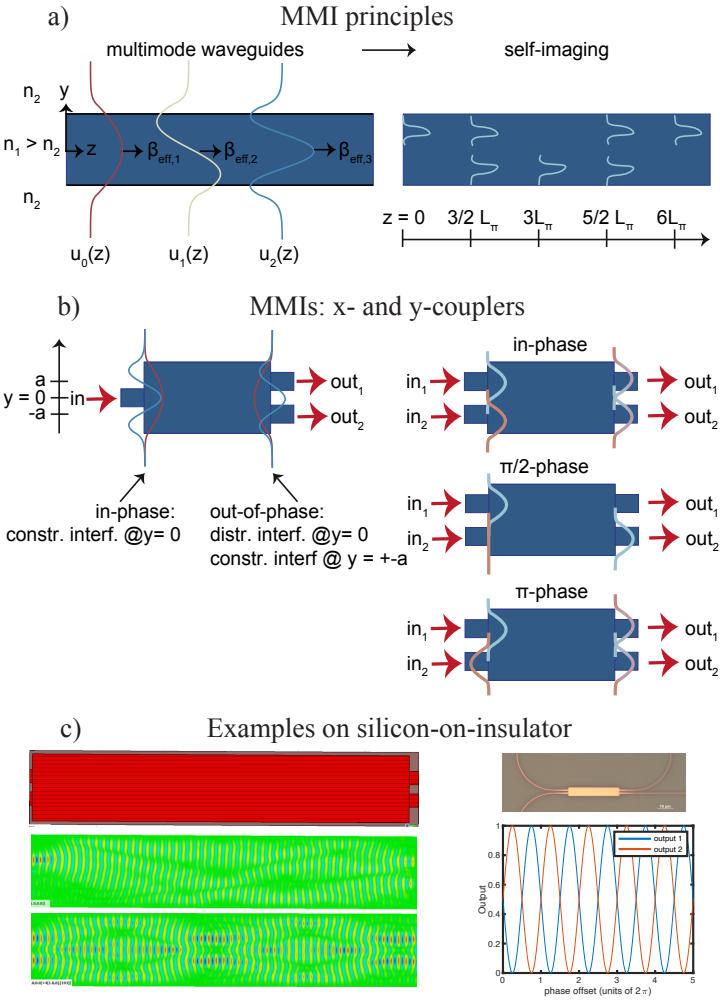

Figure 8.1: Multi-mode interference devices.

$$\beta_m - \beta_0 \approx -\frac{m(m+2)\pi^2}{2d^2 n_1 k_0}. \tag{8.2.5}$$

To simplify the math, we define L_π as the beating length $L_\pi = \frac{\pi}{\beta_0 - \beta_1} = \frac{2d^2 n_1 k_0}{3\pi}$ and find that

$$\beta_m - \beta_0 \approx -\frac{m(m+2)\pi}{3L_\pi}. \tag{8.2.6}$$

We can now rewrite the total field in the multimode waveguide at positions $z = pL$ as

$$\mathbf{E}(y, z = pL, t) = \sum_{m=1}^{N} a_m u_m(y) e^{-i(\omega t - \beta_0 pL + \frac{m(m+2)\pi}{3L_\pi} pL)} \mathbf{e}_x \tag{8.2.7}$$

The overall phase $e^{i\beta_0 pL}$ plays no role in the interference pattern and can be discarded. The periodic locations along the propagation direction in the waveguide where the input field is reproduced depends on the properties of the phase term

$$e^{-i\frac{m(m+2)\pi}{3L_\pi} pL} \tag{8.2.8}$$

In the following, we will analyze this phase term. We consider a first case where

$$e^{-i\frac{m(m+2)\pi}{3L_\pi} pL} = 1, \ \forall m$$

and find that $\mathbf{E}(y, z = pL, t) = \mathbf{E}(y, z = 0, t)$. At these locations pL, the exact input field is reconstructed.

Let's first deeply understand what this math means. We consider the first direct replica of the input field along propagation direction and thus $p = 1$. In this case, we find that

$$m = 0, \ e^{-i\frac{0\pi}{3L_\pi} L} = 1 \tag{8.2.9}$$

$$m = 1, \ e^{-i\frac{3\pi}{3L_\pi} L} = 1 \tag{8.2.10}$$

$$m = 2, \ e^{-i\frac{8\pi}{3L_\pi} L} = 1 \tag{8.2.11}$$

$$m = 3, \ e^{-i\frac{15\pi}{3L_\pi} L} = 1 \tag{8.2.12}$$

$$... \tag{8.2.13}$$

must all be satisfied. The smallest value of L for which these conditions are satisfied are $L = 6L_\pi$. Further direct images are then constructed at distances $pL = 6pL_\pi$.

We note that all of the above conditions are satisfied regardless of the exact properties of $\mathbf{E}(y, z = 0, t)$! Isn't that cool? This means that the existence of these direct self-images ***are entirely independent of the input field***.

8.3 Mirrored images

We then consider a second case where

$$e^{-i\frac{m(m+2)\pi}{3L_\pi}pL} = (-1)^m,\ \forall m.$$

In this case, all symmetric modes are imaged into themselves at positions pL (acquire no phase shift), whereas all anti-symmetric modes acquire a π phase shift, or a minus sign. They are thus flipped with respect to the z-axis. This can be easily verified by writing out the conditions that need to be satisfied where the first flipped self-image will occur ($p = 1$):

$$m = 0,\ e^{-i\frac{0\pi}{3L_\pi}L} = 1 \tag{8.3.1}$$
$$m = 1,\ e^{-i\frac{3\pi}{3L_\pi}L} = -1 \tag{8.3.2}$$
$$m = 2,\ e^{-i\frac{8\pi}{3L_\pi}L} = 1 \tag{8.3.3}$$
$$m = 3,\ e^{-i\frac{15\pi}{3L_\pi}L} = -1 \tag{8.3.4}$$
$$... \tag{8.3.5}$$

The smallest value of L for which these conditions are satisfied are $L = 3L_\pi$. Further direct images are then constructed at distances $pL = 6pL_\pi + 3L_\pi = 3L_\pi(2p + 1)$.

Overall, this results into the input field being images onto a mirrored replica of itself with respect to the z-axis $\mathbf{E}(y, z = pL, t) = \mathbf{E}(-y, z = 0, t)$.

8.4 Mirrored and self-images

We can now investigate one last case of $L = \frac{3L_\pi}{2}$ and find

$$m = 0,\ e^{-i\frac{0\pi}{3L_\pi}\frac{3L_\pi}{2}} = 1 = \frac{1}{2}((1+i)+(1-i)) \tag{8.4.1}$$
$$m = 1,\ e^{-i\frac{3\pi}{3L_\pi}\frac{3L_\pi}{2}} = i = \frac{1}{2}((1+i)-(1-i)) \tag{8.4.2}$$
$$m = 2,\ e^{-i\frac{8\pi}{3L_\pi}\frac{3L_\pi}{2}} = 1 = \frac{1}{2}((1+i)+(1-i)) \tag{8.4.3}$$
$$m = 3,\ e^{-i\frac{15\pi}{3L_\pi}\frac{3L_\pi}{2}} = i = \frac{1}{2}((1+i)-(1-i)) \tag{8.4.4}$$
$$... \tag{8.4.5}$$

In this particular case, we can then summarise that $\mathbf{E}(y, z = \frac{3L_\pi}{2} + 6pL_\pi, t) = \frac{1+i}{2}\mathbf{E}(y, z = 0, t) + \frac{1-i}{2}\mathbf{E}(-y, z = 0, t)$. This means that both an exact replica and a mirrored replica of the input field appears at these locations, albeit each one with their distinct phases. Importantly, note that the phase between the exact and mirrored image is $\frac{\pi}{2}$.

8.5 X- and Y-splitters

The above concepts of multi-mode interference will be extremely useful to create beam splitters. While, in general, beam splitters take N arbitrary input beams and split them into M arbitrary output beams, in this lecture, we will only study 1×2 and 2×2 beam splitters. The former, also called a Y-splitter takes one input and splits it into two outputs. The latter, also called an X-splitter, takes two inputs and splits it into two outputs.

8.5.1 Y-splitters

Let's first consider Y-splitters. One very straight-forward implementation of a y-splitter that you will also see during the exercise session is discussed at the end of the previous section. For a lenth $L = \frac{3L_\pi}{2}$, we have two images forming, an exact replica and a mirrored replica with respect to the z-axis. This is exactly a Y-splitter. Note also the interesting fact that the phase difference of the two output beams is $\frac{\pi}{2}$, exactly the same as one would have in the most basic beam splitter. This means that the physics of a Y-splitter can be exactly mapped into a single-input beam splitter.

We will now consider a second type of beam splitter, where the input beam is symmetric around $y = 0$. In this case, the original and mirrored replica overlap. Nevertheless, also in this case, it is possible to find a condition for which the input beam is split into two identical output beams. For example, in the exercise we will consider an input field where the fundamental and second order excited modes ($m = 0$ and $m = 2$) are primarily excited.

8.5.2 X-splitters

Finally, we will consider X-splitters, which can trivially be implemented through a superposition of two input fields, symmetrically arranged around $y = 0$, into a multi-mode interference device of the kind we discuss in the previous section.

Resonators

This chapter introduces the notion of chip-scale resonators.
Key questions: Overview:
• What is a resonator? Explain the dynamics intuitively.
• How does the reflectivity influence the various properties of a resonator?
• How does the FSR change when the cavity is in vacuum, or in a material with dispersive refractive index?
• How do the various properties of a resonator depend on the wavelength of resonance?
Key concepts: round-trip, Fabry-Perot, ring resonator, various coupling regimes, power build-up, field enhancement, free spectral range, linewidth
Key equations: Finesse, linewidth, free spectral range, quality factor, cavity modes, reflectivities and transmission
Literature: Ch. 11.1 in Saleh and Teich, Laser Photonics Rev. 6, No. 1, 47–73 (2012)

A resonator is a photonic structure that can confine light to a well-defined volume. Unlike plane waves which fill up the entire space and propagate, resonator modes ***are confined and localized*** to a well-defined part of space.

The most simple resonator is a ***Fabry-Perot resonator***. This resonator consists of two parallel, highly reflective mirrors that are separated by a distance L, as shown in Fig. 9.1. The resonator may be filled with a dielectric material or can be empty (vacuum). We first consider the idealized case, where both the medium is lossless and where the two mirrors have complex reflection coefficients $r_{1,2} = |r_{1,2}|e^{i\phi_{1,2}}$ and transmission coefficients $t_{1,2} = |t_{1,2}|e^{i\xi_{1,2}}$. Furthermore, we consider $r_0 = |r_0|e^{i\phi_0}$ the reflection of an externally incident wave off the left mirror.

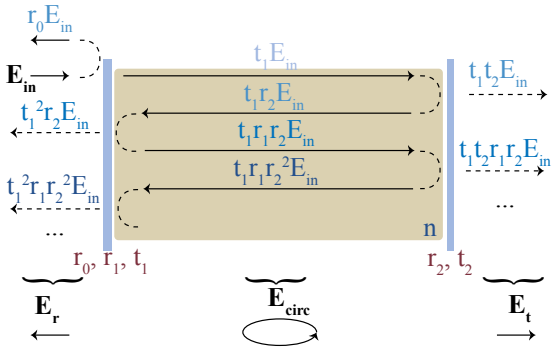

Figure 9.1: A Fabry-Perot resonator consists of two parallel mirrors and a dielectric medium in between. The total reflected, transmitted and internally circulating fields (E_r, E_t, E_{circ}) can be retrieved by considering successive transmission and reflection of the incident wave (E_{in}) at the two mirrors. In the sketch above we omitted the propagation phases for clarity (but they should of course be taken into account!).

9.1 Fabry-Perot resonators

We will now understand the properties of a Fabry-Perot resonator by treating an incident wave with amplitude E_{in} that is traveling through the system in a step-by-step fashion. After passing the first mirror, the field has an amplitude $t_1 E_{in}$. The reflected wave has an amplitude $r_0 E_{in}$. A wave traveling inside the lossless medium acquires a phase $e^{ink_0 d}$ while propagating forward and $e^{ink_0 d}$ while propagating backwards after its reflection at the right mirror. After one full ***round-trip*** we have that the total leftwards propagating beam on the left of the left mirror is $r_0 E_{in} + t_1^2 r_2 E_{in} e^{i2nk_0 L}$ and the total rightwards propagating beam inside the resonator at the right of the left mirror $t_1 E_{in} + t_1 r_1 r_2 E_{in} e^{i2nk_0 L}$.

From these two examples, we find that the total transmitted wave through the resonator after an infinity of round-trips is:

$$E_t = t_1 t_2 e^{ink_0 L} E_{in} + t_1 t_2 r_1 r_2 E_{in} e^{i3nk_0 L} + t_1 t_2 (r_1 r_2)^2 E_{in} e^{i5nk_0 L} + \quad (9.1.1)$$

$$E_t = t_1 t_2 e^{ink_0 L} E_{in} \sum_{l=0}^{\infty} (r_1 r_2)^l e^{i2lnk_0 L} \quad (9.1.2)$$

$$E_t = E_{in} \frac{t_1 t_2 e^{ink_0 L}}{1 - r_1 r_2 e^{i2nk_0 L}} \quad (9.1.3)$$

We find in this case that the total transmission of power in this case is

$$T = \frac{I_t}{I_{in}} = \frac{|E_t|^2}{|E_{in}|^2} = \frac{(1-R_1)(1-R_2)}{(1-\sqrt{R_1R_2})^2 + 4\sqrt{R_1R_2}\sin^2(nk_0L)} \tag{9.1.4}$$

where $R_1 = |r_1|^2$, $R_2 = |r_2|^2$ and the phases $\phi_{1,2} = 0$. The total reflection of power is consequently $R = 1 - T$. From the formula above, we find that the transmitted power is maximized when $nk_0L = m\pi$ and hence

$$\nu_m = \frac{mc_0}{2nL}. \tag{9.1.5}$$

The fields oscillating at frequencies ν_m are ***Fabry-Perot cavity modes***, and they are spaced in frequency domain by the so-called ***free spectral range (FSR)***

$$FSR = \nu_m - \nu_{m-1}. \tag{9.1.6}$$

The transmission is then

$$T(\nu_m) = \frac{(1-R_1)(1-R_2)}{(1-\sqrt{R_1R_2})^2} \overset{R_1=R_2}{=} 1. \tag{9.1.7}$$

Following the same procedure, we find that the total reflected beam from

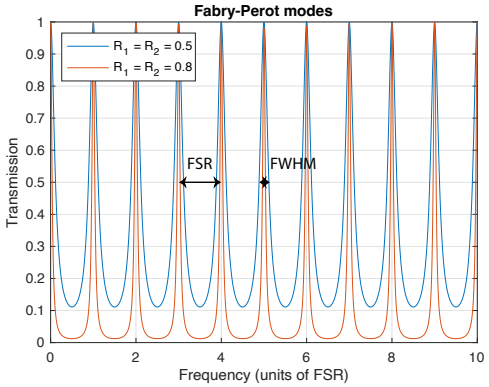

Figure 9.2: A Fabry-Perot resonator has a typical transmission as shown here for a fixed cavity length and two different reflectivities $R_1 = R_2 = R$. The higher the reflectivity, the smaller the FWHM.

the Fabry-Perot is

$$E_r = r_0 E_{in} + t_1^2 r_2 E_{in} e^{i2nk_0L} + t_1^2 r_2 r_1 r_2 E_{in} e^{i4nk_0L} + ... \quad (9.1.8)$$

$$E_r = r_0 E_{in} + t_1^2 r_2 E_{in} e^{i2nk_0L} \sum_{l=0}^{\infty} (r_1 r_2)^l e^{i2lnk_0L} \quad (9.1.9)$$

$$E_r = r_0 E_{in} + \frac{t_1^2 r_2 E_{in} e^{i2nk_0L}}{1 - r_1 r_2 e^{i2nk_0L}} \quad (9.1.10)$$

$$E_r = E_{in} \frac{r_0 - r_0 r_1 r_2 e^{i2nk_0L} + t_1^2 r_2 e^{i2nk_0L}}{1 - r_1 r_2 e^{i2nk_0L}} \quad (9.1.11)$$

Finally, we find that the circulating field inside the resonator is:

$$E_{circ} = t_1 E_{in} + t_1 r_1 r_2 e^{i2nk_0L} E_{in} + t_1 (r_1 r_2)^2 E_{in} e^{i4nk_0L} + ... \quad (9.1.12)$$

$$E_{circ} = t_1 E_{in} \sum_{l=0}^{\infty} (r_1 r_2)^2 e^{i2lnk_0L} \quad (9.1.13)$$

$$E_{circ} = E_{in} \frac{t_1}{1 - r_1 r_2 e^{i2nk_0L}} \quad (9.1.14)$$

Consequently, we find for the case where $R_1 = |r_1|^2$, $R_2 = |r_2|^2$ and the phases $\phi_{1,2} = 0$

$$\frac{|E_{circ}|^2}{|E_{in}|^2} = \frac{1 - R_1}{(1 - \sqrt{R_1 R_2})^2 + 4\sqrt{R_1 R_2} \sin^2(nk_0L)} \quad (9.1.15)$$

For the special case where $R_1 = R_2 = R$ we find the ***power build-up*** or ***maximal circulating field*** when $nk_0L = m\pi$ amounting to

$$\frac{|E_{circ}|^2}{|E_{in}|^2} = \frac{1 - R}{(1 - R)^2} = \frac{1}{1 - R} \quad (9.1.16)$$

Consequently, the intra-cavity power scales with R: the more R approaches unity, the larger the intracavity power.

The typical transmission of a Fabry-Perot resonator is shown in Fig. 9.2. It features equidistant modes spaced by the FSR and of a full-width half-maximum linewidth that is dependent on the reflectivities of the modes. We will show this in our discussion below.

9.1.1 Free spectral range, linewidth and finesse

In the following, we will consider the case where the cavity is filled up with a refractive index $n(\nu)$ that depends on frequency. We then have

$$\nu_m = \frac{mc_0}{2n(\nu_m)L}. \quad (9.1.17)$$

In this case, the free spectral range will depend on the group refractive index as follows:

$$FSR = \nu_m - \nu_{m-1} = \frac{c_0}{2n_g L}. \tag{9.1.18}$$

Let's demonstrate this rigorously below. Starting from eq. 9.1.17 and using the approximation that $n(\nu_{m-1}) = n(\nu_m) - \Delta\nu \frac{dn}{d\nu}|_{\nu_m}$ we find that

$$\Delta\nu = \nu_m - \nu_{m-1} = \frac{mc_0}{2n(\nu_m)L} - \frac{(m-1)c_0}{2n(\nu_{m-1})L} \tag{9.1.19}$$

$$= \frac{mc_0}{2L}\left(\frac{1}{n(\nu_m)} - \frac{1}{n(\nu_{m-1})}\right) + \frac{c_0}{2n(\nu_{m-1})L} \tag{9.1.20}$$

$$= \underbrace{\frac{mc_0}{2L}}_{\nu_m n(\nu_m)} \frac{-\Delta\nu \frac{dn}{d\nu}|_{\nu_m}}{n(\nu_m)n(\nu_{m-1})} + \frac{c_0}{2n(\nu_{m-1})L} \tag{9.1.21}$$

and therefore we find

$$\Delta\nu \underbrace{\left(n(\nu_{m-1}) + \nu_m \frac{dn}{d\nu}|_{\nu_m}\right)}_{n_g} = \frac{c_0}{2L} \tag{9.1.22}$$

$$\Delta\nu = \frac{c_0}{2n_g L} \tag{9.1.23}$$

The inverse of the free spectral range is the ***cavity round-trip time***. One can easily confirm that light will take an amount of time $\frac{1}{FSR}$ to travel a distance equal to twice the cavity length.

The Fabry-Perot modes have a linewidth that can be quantified by the ***full-width at half maximum*** $\delta\nu_{FWHM}$ (FWHM). To find how the linewidth depends on the various parameters of the cavity, let's consider, without loss of generality, the simple case of a cavity with $n = 1$ and $R_1 = R_2 = R$. We then have a transmission equal to

$$T = \frac{I_t}{I_{in}} = \frac{|E_t|^2}{|E_{in}|^2} = \frac{(1-R)^2}{(1-R)^2 + 4R\sin^2(\phi)} \tag{9.1.24}$$

with $\phi = k_0 L$. The half-maximum is when $T = \frac{1}{2}$, and hence

$$4R\sin^2(\phi) = (1-R)^2 \tag{9.1.25}$$

$$\sin\phi = \pm\frac{1-R}{2\sqrt{R}} \tag{9.1.26}$$

For large reflectivities, $\frac{1-R}{2\sqrt{R}} << 1$ we can approximate $\phi = k_0 L = \pm\frac{1-R}{2\sqrt{R}}$ and hence we find the linewidth

$$\frac{2\pi\delta\nu_{FWHM}}{c_0} L = 2\frac{1-R}{2\sqrt{R}} \tag{9.1.27}$$

$$\delta\nu_{FWHM} = \frac{c_0(1-R)}{2\pi L\sqrt{R}} \tag{9.1.28}$$

With this linewidth we can estimate the time it takes for light to escape the cavity. It will be on the order of $\frac{1}{\delta\nu_{FWHM}}$.

Now that we have the linewidth, we can easily compute the ***quality factor Q***:

$$Q = \frac{\nu_m}{\delta\nu_{FWHM}} = \frac{m\pi\sqrt{R}}{1-R} \tag{9.1.29}$$

The quality factor essentially tells you how many oscillation cycles the light does before it decays from the cavity.

And we can also compute the ***Finesse*** $\mathcal{F}$

$$\mathcal{F} = \frac{\Delta\nu}{\delta\nu_{FWHM}} = \frac{\pi\sqrt{R}}{1-R} \tag{9.1.30}$$

The Finesse essentially tells you how many round-trips the light does before it leaves the cavity. This is clearly linked to the power build-up in the cavity. In other words, the higher the Finesse, the higher the build-up because, in this way, the light can interfere constructively with itself. This leads to an increased intensity inside the cavity.

We find generally a couple of very important outcomes. The linewidth of a Fabry-Perot resonator does not depend on the mode number m, nor does its Finesse. On the other hand, these two strongly depend on R - the higher R, the higher the Finesse and the smaller the linewidth. On the other hand, the Q-factor depends on the mode number m!

9.1.2 On-chip Fabry-Perot resonators

While Fabry-perot resonators are mostly prevalent in free-space applications, they can be implemented also on-chip, with flexibility at the moment limited by the available fabrication techniques.

There are a few important differences between free-space and on-chip Fabry-Perot resonators, stemming from the fact that the phase velocity is altered on-chip. In particular, we can now further engineer the effective index of guided modes through geometry. The resonant condition is fulfilled when $n_{eff}k_0L = m\pi$ and hence

$$\nu_m = \frac{mc_0}{2n_{eff}(\nu_m)L}. \tag{9.1.31}$$

This also alters the free spectral range as discussed in the previous section to

$$FSR = \nu_m - \nu_{m-1} = \frac{c_0}{2n_gL}. \tag{9.1.32}$$

now with $n_g = n_{eff} + \nu\frac{dn_{eff}}{d\nu}$. Note the important distinction that while the absolute position of the resonances depends on the effective refractive

index n_{eff}, the spacing between any two adjacent resonances depends on the group index n_g. Note that both generally depend on frequency.

We will treat one such example of waveguide-based Fabry-Perot resonators in the simulation class and is shown in Fig. 9.3. We will however conclude that their usefulness for high-finesse cavities is limited in planar geometries that are accessible in integrated photonic structures.

9.2 Ring resonators

In contrast to Fabry-Perot resonators, ring resonators are made of one looped waveguide that forms a single ring of radius R and geometric length $L_{cav} = 2\pi R$. More generally, ring resonators refer to any looped resonator and there are several variations of ring resonators, such as racetrack or

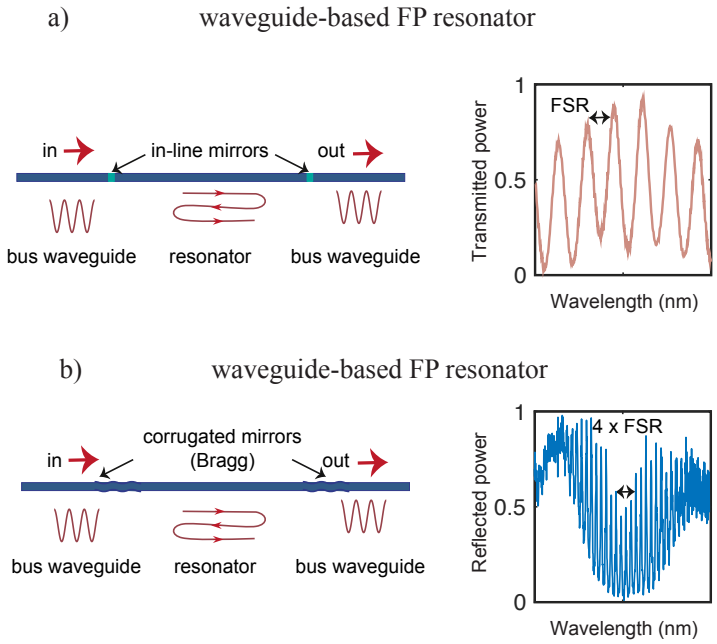

Figure 9.3: On-chip Fabry Perot resonators. The cavity is made of a waveguide, which allows control over the effective refractive index. The mirrors are however somewhat more difficult to do on-chip: in-line mirrors from another dielectric material require a fabrication step where the waveguide material is replaced by another material. Another, more elegant option, is to do Bragg mirrors from corrugated waveguides.

folded spiral resonators. In these cases, the radius of curvature changes as a function of the position along the waveguide but - on the positive side - they can be much more compact at the same total cavity length L_{cav}, or have a long portion of the ring oriented along one single direction of the constituent waveguide material, which is favourable for nonlinear applications where the crystallographic orientation matters. We showcase a few such examples in Fig. 9.4.

At this point, it is worthwhile stressing out the fundamental aspect that total internal reflection is not perfect at curved interfaces, leading to radiative losses inside curved resonators. These are typically referred to as ***bending losses***. While we do not specifically treat this in this lecture, generally, the smaller the bending radius, the larger the losses. In addition, high-confinement waveguides, where the light is efficiently confined to its core, have lower bending losses at the same radius of curvature than low-confinement waveguides.

The coupling of light from the outside into the ring resonator is typically not done via these radiative channels that intrinsically exist all along the curved waveguide of ring resonators (one could think of coupling in light through the side walls, but this would be extremely inefficient). Instead, a much better controlled way is by placing a **bus waveguide** in the immediate vicinity of ring resonators that couple light from the bus to the ring

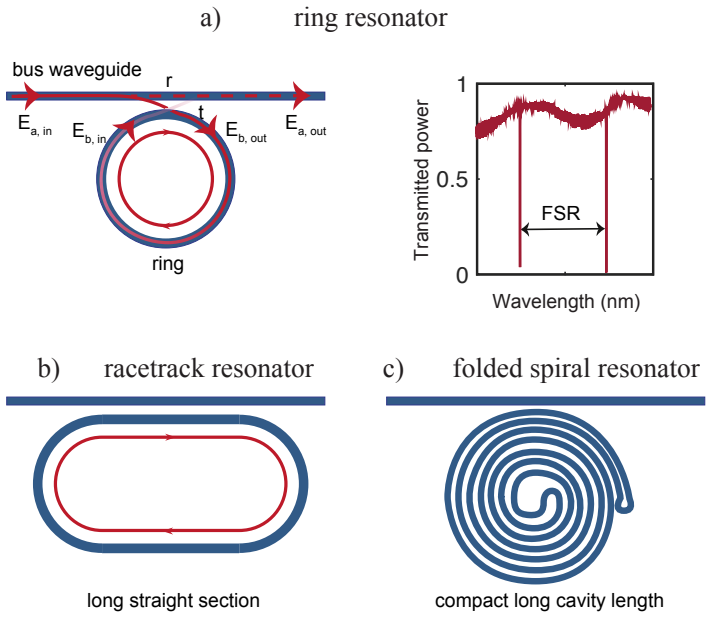

Figure 9.4: On-chip ring resonators can take many shapes .

evanescently. As we have studied in chapter 7 about coupled waveguides, the total amount of power that can be coupled from the bus waveguide to the ring depends on the distance between the two and the total interaction length. Both can be freely chosen by design.

Much like in the case of Fabry-Perot resonators, also ring resonators have a free spectral range that is given by

$$FSR = \frac{c_0}{n_g L_{cav}} = \frac{c_0}{n_g 2\pi R} \tag{9.2.1}$$

Let's now derive the most important properties of ring resonators, noting that the coupling from the bus waveguide to the ring resonator resembles strongly the case of two coupled waveguides where the end of waveguide b is closed onto itself. In addition, the length of the coupling region is fixed by design. For us, this means that the electric fields in the bus waveguide and inside the ring resonator can be described by the relationship:

$$\begin{pmatrix} \mathbf{E}_{a,out} \\ \mathbf{E}_{b,out} \end{pmatrix} = \begin{pmatrix} r & it \\ it & r \end{pmatrix} \begin{pmatrix} \mathbf{E}_{a,in} \\ \mathbf{E}_{b,in} \end{pmatrix} \tag{9.2.2}$$

We further note that the field $\mathbf{E}_{b,in}$ is the propagated $\mathbf{E}_{b,out}$ that 1. acquired a propagation phase that depends on the radius R of the ring resonator and 2. experienced an attenuation due to absorption and bending losses:

$$\mathbf{E}_{b,in} = e^{-\frac{\alpha}{2}2\pi R} e^{ik_0 n_{eff} 2\pi R} \mathbf{E}_{b,out} = a e^{i\phi} \mathbf{E}_{b,out} = a e^{i\phi} (r\mathbf{E}_{b,in} + it\mathbf{E}_{a,in}) \tag{9.2.3}$$

where α is the intensity absorption coefficient ($I(z) = e^{-\alpha z} I(0)$), and n_{eff} the effective index of the mode inside the ring resonator. Hence, a is the single-pass absorption coefficient of the mode and $\phi = k_0 n_{eff} 2\pi R$ is the single-pass phase accumulation.

In conclusion, we find

$$\mathbf{E}_{b,in} = \frac{ita e^{i\phi}}{1 - ra e^{i\phi}} \mathbf{E}_{a,in} \tag{9.2.4}$$

The ***power build-up (field enhancement FE)*** in the cavity therefore depends on ϕ as follows:

$$FE = |\frac{\mathbf{E}_{b,in}}{\mathbf{E}_{a,in}}|^2 = \frac{t^2 a^2}{1 - 2ra\cos\phi + r^2 a^2}. \tag{9.2.5}$$

We find that the total output field in the bus waveguide is:

$$\begin{aligned} \mathbf{E}_{a,out} &= r\mathbf{E}_{a,in} + it\mathbf{E}_{b,in} && (9.2.6)\\ &= (r - \tfrac{t^2 a e^{i\phi}}{1 - ra e^{i\phi}})\mathbf{E}_{a,in} && (9.2.7)\\ &= \tfrac{r - r^2 a e^{i\phi} - t^2 a e^{i\phi}}{1 - ra e^{i\phi}} \mathbf{E}_{a,in} && (9.2.8)\\ &= \tfrac{r - a e^{i\phi}}{1 - ra e^{i\phi}} \mathbf{E}_{a,in} && (9.2.9) \end{aligned}$$

The ***power transmission*** in the cavity depends on ϕ as follows:

$$T = |\frac{\mathbf{E}_{a,out}}{\mathbf{E}_{a,in}}|^2 = \frac{r^2 - 2ra\cos\phi + a^2}{1 - 2ra\cos\phi + r^2a^2}. \tag{9.2.10}$$

We can solve for $T = 0$ which is fulfilled when $\phi = 2\pi m$ and $a = r$, which is called the ***critical coupling*** condition. Here, we have that $1 - a^2 = 1 - r^2 = t^2$ which signifies that the coupling from the bus waveguide to the ring is equal to the single-pass absorption loss in the ring. In this case, no power is transmitted further than the ring and a maximal amount of power is coupled into the ring. We can calculate the intensity build-up for the case of critical coupling $a = r$:

$$|\frac{\mathbf{E}_{b,in}}{\mathbf{E}_{a,in}}|^2 = \frac{r^2}{1 - r^2} \tag{9.2.11}$$

In the condition where $a > r$ and hence $1 - a^2 < t^2$, the single round-trip loss is smaller than the coupling rate and the ring resonator is ***over-coupled.*** In the condition where $a < r$ and hence $1 - a^2 > t^2$, the single

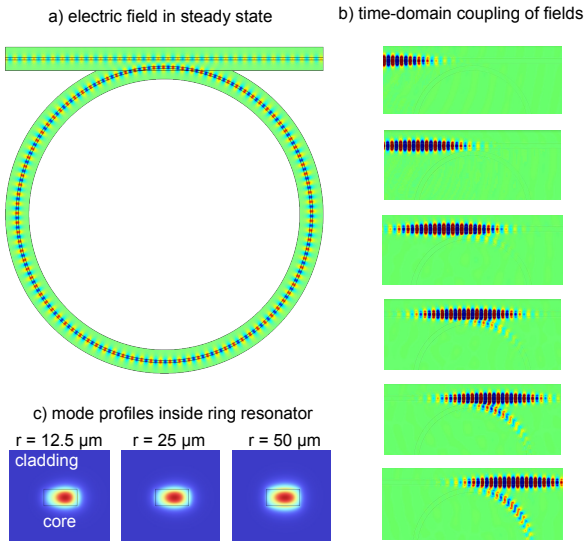

Figure 9.5: Fields as they couple into ring resonators. a) On-resonance, the cavity field confined inside the ring resonator is enhanced compared to the one present in the bus waveguide. b) Time-domain picture of optical fields coupling into a ring resonator via directional couplers. c) The mode profile in a ring resonator is slightly pushed outwards and not as symmetrical anymore compared to the guided modes of a ridge waveguide. This effect becomes stronger, the smaller the ring radius is.

round-trip loss is larger than the coupling rate and the ring resonator is ***under-coupled.*** Light can not efficiently build up in the cavity as a significant proportion is lost to absorption. We show in Fig. 9.6 that only in critically coupled resonators the transmission is zero and power build-up is maximal at the resonant wavelength.

Let's now consider the more realistic case where $a \neq r$, and $\phi = 2\pi m$. In this case, we find that the transmission reaches a minimum of

$$T_{min} = \frac{(r-a)^2}{(ra-1)^2}. \tag{9.2.12}$$

To find the full with at half maximum in this case we search for ϕ where $T(\phi) = \frac{1}{2}(1 + T_{min})$ and hence:

$$\frac{r^2 - 2ra\cos\phi + a^2}{1 - 2ra\cos\phi + r^2a^2} = \frac{1}{2}(1 + \frac{(r-a)^2}{(ra-1)^2}). \tag{9.2.13}$$

An analytical expression can be found for any a and r but for critical coupling we find

$$\cos\phi = \frac{4r^2 - r^4 - 1}{2r^2} \tag{9.2.14}$$

and consequently $\phi = \pm arccos\frac{4r^2-r^4-1}{2r^2}$. With this, we find $\Delta\phi = 2arccos\frac{4r^2-r^4-1}{2r^2}$ and the full width at half max by solving

$$\frac{2\pi\delta\nu_{FWHM}}{c_0} 2\pi Rn = 2arccos\frac{4r^2 - r^4 - 1}{2r^2} \tag{9.2.15}$$

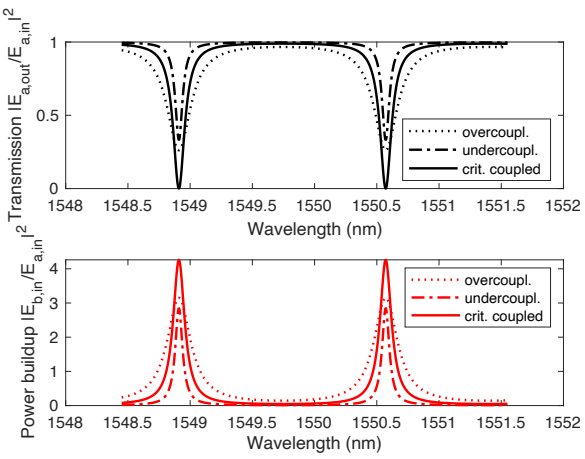

Figure 9.6: Transmission and power build-up properties of ring resonators. Optimal power build-up is found in the critical coupling condition.

We find that if $r \sim 1$ and hence $\Delta\phi$ is very small, then $\cos\phi = 1 - \frac{\phi^2}{2}$ and hence $\phi = \pm\frac{r^2-1}{r}$ Finally, the ***Finesse*** of a critically coupled ring resonator, defined as the free spectral range over the full width at half maximum of a resonance

$$\mathcal{F} = \frac{2\pi}{2arccos\frac{4r^2-r^4-1}{2r^2}} \underbrace{=}_{r\sim 1} \frac{2\pi r}{2(1-r^2)} = \frac{\pi r}{1-r^2} = \frac{\pi\sqrt{R}}{1-R} \tag{9.2.16}$$

We find the Finesse of a ring resonator to be formally equivalent to the one of a Fabry-Perot resonator.

Antennas

This chapter introduces the notion of printed antennas.
Key questions:
• What is a Hertzian dipole?
• How does the emitted power scale with wavelength and length of antenna?
Key concepts: Radiation pattern, antenna gain, power flow.
Key equations: Radiation pattern of a dipolar antenna
Literature: Balanis, Antenna theory and design

Antennas are very important components that can radiate energy into free space. They can also be operated in reverse and collect radiation from propagating fields into a well-defined space.

The most popular antennas you may know are ***wire antennas***, akin to those that you may have seen connected to old radios or to cars. They essentially consist of a metallic wire that, as we shall see later, can efficiently irradiate electromagnetic energy from a generator into free space.

10.1 Hertzian dipole antennas

For the beginning, we will treat the simple case of a ***Hertzian dipole***. It consists of two metallic wires as shown in Fig. 10.1 connected to a signal generator (e.g. a voltage source) oscillating at frequency ω. The length l of the two metallic wires is by definition much smaller than the wavelength for a Hertzian dipole. In this case, we can approximate the current density $\mathbf{J}$ in the wire by

$$\mathbf{J}(\mathbf{r},t) = Il\delta(\mathbf{r})\mathbf{e}_z = \frac{dq}{dt}l\delta(\mathbf{r})\mathbf{e}_z \tag{10.1.1}$$

where $q = Ne$ is the total charge carried by N electrons that oscillate and $\delta(\mathbf{r})$ is defined so that $\int dv\delta(r) = 1$ and dv is a volume integral.

We can now take the Fourier transform of the above equation and analyze it in the frequency domain and find that

$$\mathbf{J}(\mathbf{r},\omega) = -i\omega ql\delta(\mathbf{r})\mathbf{e}_z = -i\omega p\delta(\mathbf{r})\mathbf{e}_z \tag{10.1.2}$$

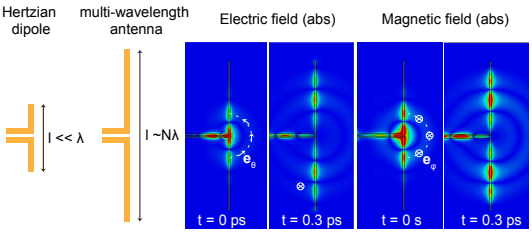

Figure 10.1: Antennas and their electric and magnetic field emission patterns.

where $p = ql$ is the dipole moment (polarisation) introduced by the n charges. From the current density, we can compute the vector potential $\mathbf{A}(\mathbf{r}, \omega)$:

$$\mathbf{A}(\mathbf{r}, \omega) = \mu_0 \int\int\int dv' \mathbf{J}(\mathbf{r'}, \omega) \frac{e^{i\beta|\mathbf{r}-\mathbf{r'}|}}{4\pi|\mathbf{r}-\mathbf{r'}|} \tag{10.1.3}$$

$$= -i\omega p \mu_0 \frac{e^{i\beta|\mathbf{r}|}}{4\pi|\mathbf{r}|}\mathbf{e}_z = A_z \mathbf{e}_z \tag{10.1.4}$$

where $|r| = \sqrt{\rho^2 + z^2}$. From the vector potential we can then compute the magnetic field using cylindrical coordinates:

$$\mathbf{H}(\mathbf{r}, \omega) = \frac{1}{\mu_0} \nabla \times \mathbf{A}(\mathbf{r}, \omega) \tag{10.1.5}$$

$$= \frac{1}{\mu_0}\left(\frac{1}{\rho}\frac{\partial}{\partial\phi} A_z \mathbf{e}_\rho - \frac{\partial}{\partial\rho} A_z \mathbf{e}_\phi\right) \tag{10.1.6}$$

$$= -\frac{1}{\mu_0}\frac{\rho}{r}\frac{\partial}{\partial r} A_z \mathbf{e}_\phi \tag{10.1.7}$$

$$= -i\omega p \frac{\rho}{r} \frac{i4\pi r\beta e^{i\beta r} - 4\pi e^{i\beta r}}{(4\pi r)^2}\mathbf{e}_\phi \tag{10.1.8}$$

We also remind that $\rho = r\sin\theta$ and find in spherical coordinates that:

$$\mathbf{H}(\mathbf{r}, \omega) = -\omega q l \frac{e^{i\beta r}}{4\pi r^2}(\beta r + i)\sin\theta \mathbf{e}_\phi = H_\phi \mathbf{e}_\phi \tag{10.1.9}$$

We can easily pass into the time domain and find

$$\mathbf{H}(\mathbf{r}, t) = -iIl \frac{e^{i\beta r}}{4\pi r^2}(\beta r + i)\sin\theta \mathbf{e}_\phi = H_\phi \mathbf{e}_\phi \tag{10.1.10}$$

The magnetic field therefore points at any position into the direction of $\mathbf{e}_\phi$. From here, we can compute the electric field associated with the Hertzian

dipole:

$$\mathbf{E}(\mathbf{r},\omega) = \frac{i}{\omega\epsilon_0}\nabla\times\mathbf{H}(\mathbf{r},\omega) \tag{10.1.11}$$

$$= \frac{i}{\omega\epsilon_0}(\frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}\sin\theta H_\phi\mathbf{e}_r - \frac{1}{r}\frac{\partial}{\partial r}rH_\phi\mathbf{e}_\theta) \tag{10.1.12}$$

$$= -\frac{ql}{\epsilon_0}\frac{e^{i\beta r}}{4\pi r^3}(\beta r+i)2\cos\theta\mathbf{e}_r \tag{10.1.13}$$

$$-\frac{iql}{\epsilon_0}\frac{e^{i\beta r}}{4\pi r^3}(-i\beta^2r^2+\beta r+i)\sin\theta\mathbf{e}_\theta \tag{10.1.14}$$

We find therefore the electric field to have both an r and a θ component.

We can now compute the far-field patterns, where we assume $\beta r >> 1$ and find

$$\mathbf{H}(\mathbf{r},\omega) = -\frac{ql\omega^2}{c_0}\frac{e^{i\beta r}}{4\pi r}\sin\theta\mathbf{e}_\phi \tag{10.1.15}$$

and

$$\mathbf{E}(\mathbf{r},\omega) = -\frac{ql\omega^2}{\epsilon_0 c_0^2}\frac{e^{i\beta r}}{4\pi r}\sin\theta\mathbf{e}_\theta \tag{10.1.16}$$

We can again represent this in time domain

$$\mathbf{H}(\mathbf{r},t) = -i\frac{Il\omega}{c_0}\frac{e^{i\beta r}}{4\pi r}\sin\theta\mathbf{e}_\phi \tag{10.1.17}$$

and

$$\mathbf{E}(\mathbf{r},t) = -i\frac{Il\omega}{\epsilon_0 c_0^2}\frac{e^{i\beta r}}{4\pi r}\sin\theta\mathbf{e}_\theta \tag{10.1.18}$$

and find that

$$\frac{E_\theta}{H_\phi} = \frac{1}{\epsilon_0 c_0} = \sqrt{\frac{\mu_0}{\epsilon_0}} \tag{10.1.19}$$

the vacuum impedance.

From here we can compute the power flow as

$$\langle S\rangle = \frac{1}{2}\mathcal{R}[\mathbf{E}\times\mathbf{H}^*] = \frac{(Il\omega)^2}{2\epsilon_0 c_0^3(4\pi r)^2}\sin^2\theta\mathbf{e}_r \tag{10.1.20}$$

and can get the total power by integrating over the entire area

$$P = \int dA\langle S\rangle = \int_0^{2\pi} d\phi\int_0^\pi d\theta r^2\sin\theta\langle S\rangle \tag{10.1.21}$$

and with $\int_0^\pi d\theta \sin^3\theta = \frac{4}{3}$ we get

$$P = \frac{4}{3} 2\pi \frac{(Il\omega)^2}{2\epsilon_0 c_0^3 (4\pi)^2} = \frac{(Il\omega)^2}{12\pi\epsilon_0 c_0^3} \tag{10.1.22}$$

And from here we can compute the antenna gain:

$$G(\theta, \phi) = \frac{\langle S \rangle}{\frac{P}{4\pi r^2}} = \frac{3}{2} \sin^2\theta \tag{10.1.23}$$

which relates the enhanced emission by the dipole compared to an isotropic emitter with the same power.

10.2 Printed bow-tie antennas

In integrated photonic circuits, printed antennas may take various shapes and flavours. We will study the case of a bow-tie antenna in the exercise class.

Conclusions

In this lecture we have explored how patterned dielectric and metallic materials are very useful to control the properties of light on-chip. We have encountered various material platforms and understood - from basic principles - why they have dispersive properties, either due to their intrinsic properties or due to the artificially introduced patterning.

We got acquainted with components that guide light (waveguides), confine light (resonators) or efficiently collect and emit light (antennas). Complex chip-scale architectures can be assembled from this basic toolbox to implement an overall functionality of the entire photonic circuit.

Bibliography