THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Controlling light:
Methods for inverse design of nonlinear nanophotonic waveguides on chip

Albin Jonasson Svärdsby

Department of Physics
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2026

Controlling light:
Methods for inverse design of nonlinear nanophotonic waveguides on chip
Albin Jonasson Svärdsby
isbn:978-91-8103-356-4

Acknowledgements, dedications, and similar personal statements in this thesis, reflect the author's own views.

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5813
issn: 0346-718X

doi:https://doi.org/10.63959/chalmers.dt/5813

Department of Physics
Chalmers University of Technology
se-412 96 Göteborg, Sweden
Telephone +46 (0)31 772 10 00

Cover: Illustration of the inverse design process.

Chalmers digitaltryck
Göteborg, Sweden 2026

Controlling light: Methods for inverse design of nonlinear nanophotonic waveguides on chip

Albin Jonasson Svärdsby
Department of Physics
Chalmers University of Technology

Abstract

The nonlinear response of optical materials is in general very weak. In order to make use of these nonlinear interactions for optical processes in integrated optical devices, such as optical parametric amplifiers, we often need very long waveguides to achieve appreciable efficiency. These waveguides can have lengths of the order of metres and need to fulfil multiple requirements over a range of wavelengths.

Modern advances in manufacturing and nanophotonics have made possible a high degree of tailoring of waveguide geometries for near-field light enhancement to meet these requirements. However, simulating and designing these nonlinear integrated optical devices is challenging.

In this thesis, I will present methods for simulating periodic optical waveguide structures with nontrivial unit cells, and how we can use knowledge of the physics to tailor the mesh adaptation in finite-element simulations to electrodynamic problems. I will also present how we can combine machine learning and physics for scattering problems and how we can use inverse design to suggest waveguide cross sections that fulfil multiple design requirements on dispersion characteristics.

Keywords: Nanophotonics, nonlinear optics, FEM, parametric amplifiers, four wave mixing, OPA, FWM, Inverse design, AI, machine learning, conditional variational autoencoder, metamaterials, waveguide

LIST OF APPENDED PAPERS

This thesis is partially based on work presented in the following papers:

I **Adaptive meshing strategies for nanophotonics using a posteriori error estimation**
Albin J. Svärdsby and Philippe Tassin
Optics Express **32**, 14, pp. 24592-24602 (2024)

II **General framework for knowledge integration in machine learning for electromagnetic scattering using quasinormal modes**
Viktor A. Lilja, Albin J. Svärdsby, Timo Gahlmann, and Philippe Tassin
Under review.
preprint on arXiv, https://arxiv.org/abs/2509.06130

III **Determining the dispersion and nonlinear characteristics of 3D periodic waveguides using finite-element eigenmode simulations**
Albin J. Svärdsby and Philippe Tassin
Manuscript.

IV **Inverse design of optical waveguides for phase sensitive amplifiers using machine learning**
Albin J. Svärdsby and Philippe Tassin
Manuscript.

The author's contribution to the papers:

I Performed implementation of the method, evaluated the results, and wrote the paper.

II Provided technical expertise regarding machine learning and assistance with writing.

III Performed implementation of the method, evaluated the results, and wrote the paper.

IV Performed implementation of the method, evaluated the results, and wrote the paper.

Contents

List of abbreviations

ACO ant colony optimizations.

AE autoencoder.

AI artificial intelligence.

BPSK binary-phase-shift keying.

CNN convolutional neural network.

cVAE conditional variational autoencoder.

CW continuous wave.

DOF degrees of freedom.

EDFA erbium-doped fiber amplifier.

FD finite differences.

FDTD finite-difference time-domain.

FE finite element.

FEM finite element method.

FWM four-wave mixing.

GA genetic algorithm.

GAN generative adversarial network.

GPT generative pre-trained transformer.

HNLF highly nonlinear fiber.

LLM large language model.

LSTM long short-term memory.

ML machine learning.

NF noise figure.

NLSE nonlinear Schrödinger equations.

NN neural network.

OPA optical parametric amplifier.

PCF photonic crystal fibre.

PDE partial differential equations.

PEC perfect electrical conductor.

PhC photonic crystal.

PIA phase-insensitive amplifier.

PMC perfect magnetic conductor.

PML perfectly matched layers.

PSA phase-sensitive amplifier.

PSO particle swarm optimization.

QNM quasinormal mode.

QPM quasi-phase matching.

SBC scattering boundary condition.

SHG second harmonic generation.

SLM small Language Model.

SSFM split-step Fourier method.

TE transverse electric.

THG third harmonic generation.

TM transverse magnetic.

VAE variational autoencoder.

1 Introduction

1.1 What are we pursuing?

Perhaps one of the most common day-to-day appliances relying on nonlinear optics that you have come into contact with are green laser pointers. Here, nonlinear optical effects are used to convert a 1064 nm (infrared) wavelength laser source into 532 nm (green) by means of a second-order nonlinear effect. This thesis build towards towards the ability to simulate these kinds of effects for the purpose of amplifying optical signals through optical parametric processes. Amplification through nonlinear optical effects is interesting because it enables the amplification of an optical signal with a very low added noise which can be achieved with an on-chip design [1]. Common methods that achieve this rely on rather simple geometrical layouts, often consisting of some homogeneous materials or systems such as Bragg gratings [2].

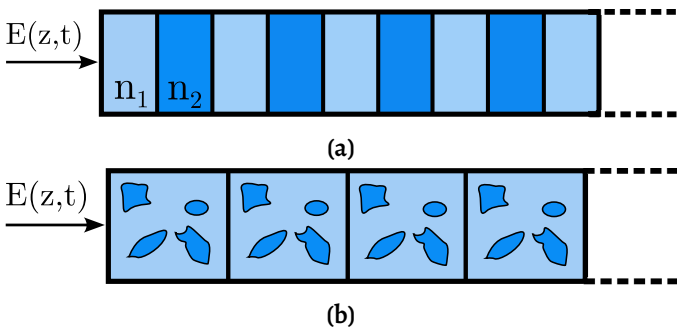


Figure 1.1: Simple vs complex geometries. Fig. 1.1a shows a uniform Bragg grating, alternating between different dielectric materials. In Fig. 1.1b, we see an example of a more complicated geometrical shape.

This geometrical constraint, although it simplifies the modelling, may also exclude better designs. However, complex geometries are more costly and complicated to simulate and cannot rely on analytical solutions, which can assist in the case well-defined and symmetric geometries. In this thesis, we will develop the tools needed to simulate these kinds of structures in the context of a periodic optical waveguide structures where each unit cell can have an arbitrary pattern, as illustrated in Fig. 1.1.

1.2 Structure of this thesis

Before presenting Papers I-IV, we will go through the relevant theory that we need to understand and simulate nonlinear optical effects in waveguides, with the intention of using them for designing optical parametric amplifiers. What this means will become clearer over the course of this thesis. My assumption will be that the reader at a point in time have read an undergraduate course in electromagnetism or equivalent. I will repeat some core concepts of linear electromagnetics in Ch. 2, so we are up to speed, and refresh the terminology that we will use on this journey. Once we are on board, we will move over to nonlinear effects in Ch. 3. Having the theory set up, we will move to how we concretely simulate these effects in Ch. 4. In order to utilise these methods, we need to understand under what limitations we operate and what kinds of structures we actually can manufacture. To that end, we will build this intuition through multiple examples of nanophotonic constructions in Ch. 5. Once we have a grasp of what structures are possible, we will traverse into the process of how one can come up with designs for devices to suit our needs, so-called *inverse design*. We will approach this subject utilizing machine learning (ML) techniques to speed up calculations in Ch. 6 and finally solving for the inverse problem in Ch. 7. Finally, I will present four research papers, papers I-IV.

I hope you as a reader will find the journey enlightening, and perhaps invoke some internal thoughts over the progress that is being made in the field and the interesting applications these nonlinear effects have.

2

Electromagnetic wave propagation

In this Chapter, we will summarize concepts of electromagnetic wave propagation that are needed to understand the nonlinear optics and simulations that we will treat in Ch. 3 and 4. This will not be an exhaustive introduction, but should be seen more as a reminder. For those looking for a more thorough introduction to electromagnetics, see for example Griffith [3], Jackson [4] or Yariv & Yeh [5].

In Sec. 2.1, we will treat the concept of dispersion and discuss material dependencies on frequency. This will then be followed in Sec. 2.2 with a discussion on modes of propagation in waveguides. Finally, we will discuss periodic structures in Sec. 2.3.

2.1 Understanding dispersion

The framework for studying the propagation of electromagnetic waves has been known for a long time, starting with the work by d'Alembert [6–8] in the 18th century and continued by Euler [9] and Lagrange [10], resulting in the wave equation. The next significant description came when Maxwell presented his work on electromagnetism in 1865 [11], verified by Hertz in 1893 [12]. The Maxwell equations that we are familiar with today were, however, formulated by Heaviside based on the work of Maxwell [13, 14]. The macroscopic differential versions of the equations can be written as:

$$\begin{aligned} \nabla \cdot \mathbf{D} &= \rho_f, & \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t}, \\ \nabla \cdot \mathbf{B} &= 0, & \nabla \times \mathbf{H} &= \mathbf{J}_f + \frac{\partial \mathbf{D}}{\partial t}. \end{aligned} \tag{2.1}$$

Any time an electromagnetic wave interacts with a dielectric material there will be an interaction where the electric field induces electric dipole moments in the material.

These dipole moments then cause an added term to the electric field, **E**, in the form of a polarisation, **P**, and together they form what is called the displacement field [3, 4]:

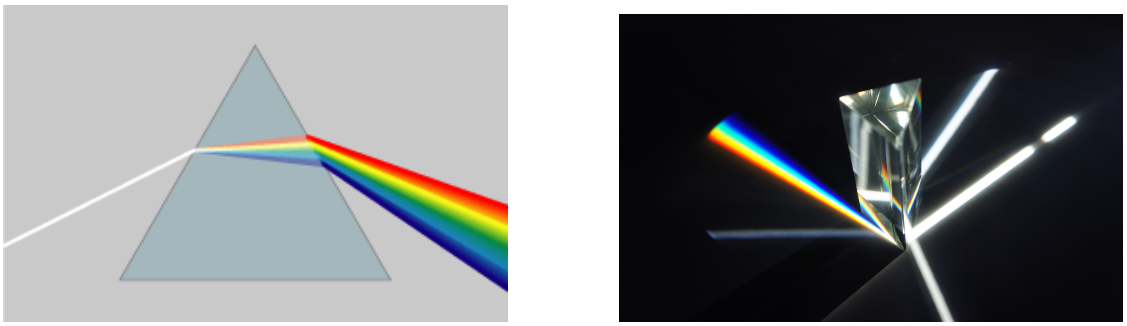
$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P} = \varepsilon \mathbf{E}, \tag{2.2}$$

where the permittivities ε and ε_0 quantify the polarizability in the material and in free space, respectively. What gives rise to this effect is that a material contains both positive and negative charges, and when an electric field is applied they are pulled apart, or displaced, causing the charge distribution to be polarised. If a molecule already has a polarised distribution then the molecule also has the possibility to rotate in order to align with the applied field.

If the polarisation is the result of a field interacting with the material on a macroscopic scale, it is not unreasonable to think that this effect should be dependent on the field strength in the material and that indeed holds true. The amount of polarisation that occurs in a material when it interacts with an electric field is expressed with the electric susceptibility χ of the material:

$$\mathbf{P} = \varepsilon_0 \chi \mathbf{E}. \tag{2.3}$$

The susceptibility has a frequency dependence $\chi(\omega)$, a phenomenon called dispersion. The effect of this is that waves propagating through the material have different amount of attenuation and different phase velocities depending on frequency. This give rise to different frequencies travelling with different speeds in the material, which results in things like refraction, usually illustrated with a prism like in Fig. 2.1.



(a) (b)

Figure 2.1: Illustration of dispersion in a prism. To the left we see a 2D representation and to the right we see a version in 3D.

When we are designing devices, waveguides, these linear effect have big consequences, because the design needs to operate with multiple frequencies. But since we have this frequency dependence, it also opens up the possibility to tailor systems and materials, so that it only favours desirable frequencies and counteracts undesired ones.

2.2 Modes of propagation in waveguides

When we talk about dielectric optical waveguides, we are referring to structures that consists of an optical medium with high refractive index, called a core, surrounded by a material of lower refractive index, called a cladding. The purpose of the core is to confine light and guide it along the longitudinal direction. Waveguides come in many different shapes and configurations, and readers interested in optical waveguides will do well to read Liu[15], Yariv & Yeh [5], or Saleh & Teich [16].

When we have a structure where a wave can propagate with a defined field pattern in the structure that is confined, we refer to it as a mode. If the wave leaks out of its confinement as it propagates in the waveguide, then it is a leaky mode. To illustrate what a mode is, we can take the example of a planar waveguide stretching to infinity as depicted in 2D in Fig. 2.2. This type of waveguide has a fixed set of stable modes, which have analytical solutions for the transverse field. If the solution has no electric field in the propagation direction, but only in the transverse direction, it is called a transverse electric (TE) mode. If the solution instead has no magnetic field in the propagation direction, but only in the transverse direction, then it is called a transverse magnetic (TM) mode. When we have multiple modes then we refer to them as TE0, TE1 etc. For more complex geometries, solutions involving both electric and magnetic field, so called *hybrid modes*, are also allowed [15].

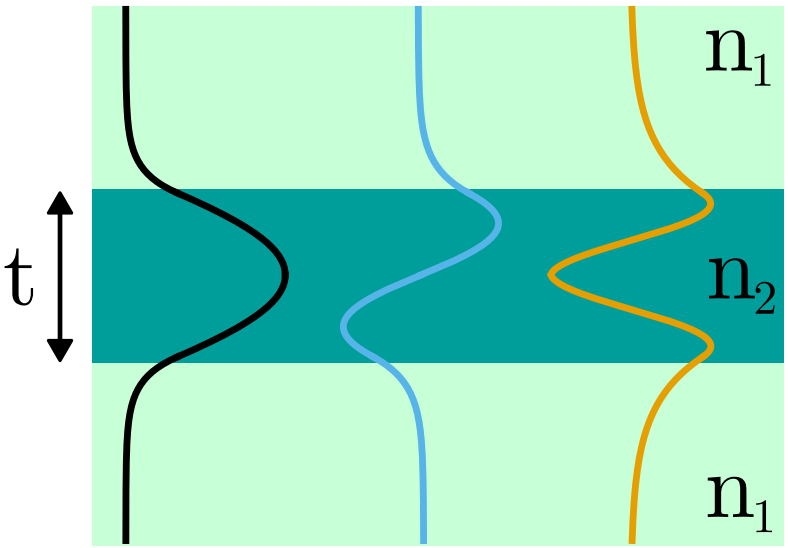


Figure 2.2: The first three propagating TE modes of a symmetric planar waveguide with dielectrics with refractive index n_1 and n_2. From left to right we see TE0, TE1, and TE2.

To quantify the propagation of a wave in a media, we can use the propagation constant, γ, defined as the ratio of the complex amplitude at the source to the wave at a

distance z:

$$\frac{A_0}{A_z} = e^{\gamma z}; \quad \gamma = \alpha + i\beta, \tag{2.4}$$

where α is referred to as the *attenuation constant* and β as the *phase constant*.

The different modes of the waveguide each have their own dispersion. Therefore, we can express the different modes on a dispersion plot, as demonstrated in Fig. 2.3. We can think of the modes as a change in the phase velocity of the wave due to the wave bouncing more frequently against the walls of the waveguide, as it propagates in the longitudinal direction, as a result of different angles of incidence.

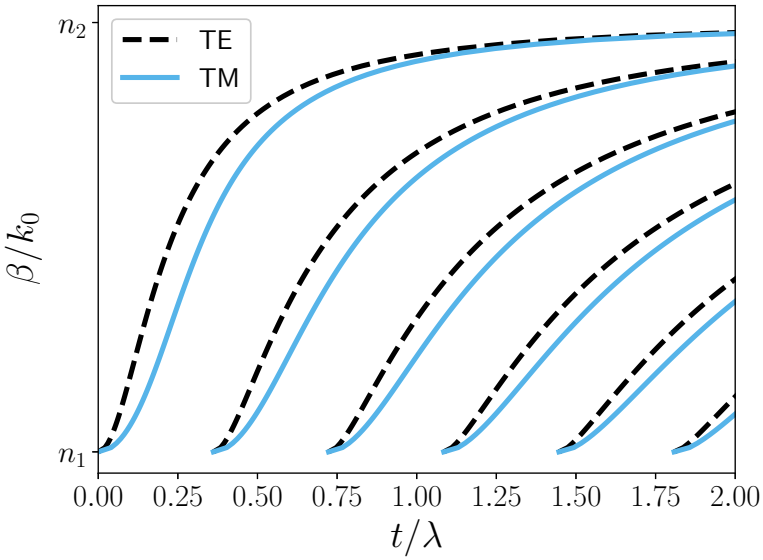
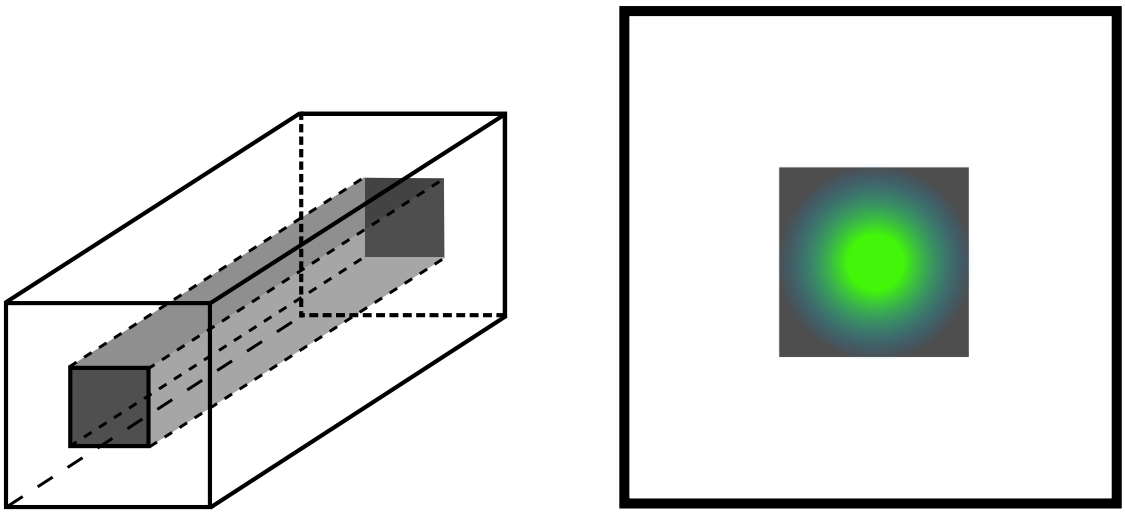


Figure 2.3: A dispersion plot showing the allowed TE and TM modes for a symmetric planar waveguide as shown in Fig. 2.2. Dispersion on the y-axis is scaled to effective index by wave number in vacuum and x-axis shows width, t, of the centre material scaled by the wavelength, λ.

The dispersion relation becomes important when we want to design waveguides that can operate over multiple frequencies. In Ch. 3, the dispersion in the waveguide will play a role in how nonlinear effects emerge.

Although analytical solutions exist for planar dielectric waveguides, once we enclose the side of the core and deal with non-planar waveguides, analytical solutions generally do not exist and we are left with numerical simulations to calculate the modes. How the mode confinement in an embedded waveguide can appear is illustrated in Fig. 2.4. Techniques for how we can calculate the dispersion numerically will be treated in Ch. 4.



(a) (b)

Figure 2.4: Depiction of the fundamental mode in an embedded waveguide. To the left in Fig. 2.4a, we have a schematic waveguide with a centre core surrounded by cladding material. To the right in Fig. 2.4b, we see an illustration of the electric field of the fundamental mode.

2.3 Bloch waves and periodic structures

We often find ourselves in a situation where we have a need to look at periodic structures. Though much of the research is focused on being able to predict the behaviour of periodic homogeneous materials, such as crystals, the same results can be used for waveguides, where we design patterning of materials with a given periodicity. It is convenient to design materials like this, since we can then analyse and categorise a smaller segment or unit cell, simplifying the problem and decreasing the computational cost that might hinder the design process. This can have a big impact on the simulation of nanophotonic structures that might need to resolve the structure on a subwavelength scale. This can lead to situation in which we have devices on the order of metres that needs to be simulated with a resolution on the order of nanometres. These types of simulations are usually unfeasible and in the best case very slow and costly, which is why everything greatly simplifies if we can use techniques like periodicity to perform calculations on a small part of the structure. Readers interested in wave propagation in different kinds of photonic structures are advised to read Markos and Soukoulis [17]. Knowledge of periodic crystal structures can also be found in solid-state physics books like Refs. [18, 19].

Let us look at the periodic structure of a waveguide depicted in Fig. 2.5. In order for us to take advantage of periodicity, we want to design the system such that the incoming and outgoing waves are equal, save for a phase rotation or amplitude decay, as expressed

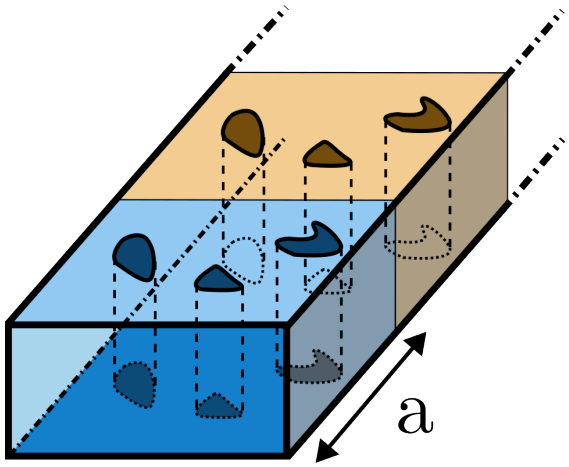


Figure 2.5: Example of a periodic patterned waveguide structure where the structure repeats after a distance a. The unit cell is coloured differently for easier distinction.

in Equation (2.5)

$$\psi_k(\mathbf{r} + \mathbf{a}) = e^{i\mathbf{k}\cdot\mathbf{a}}\psi_k(\mathbf{r}), \tag{2.5}$$

where $\mathbf{a}$ denotes the lattice vector of the periodic unit cell. Waves that behave like this are called Bloch waves, so named after Bloch who coined the requirement in 1929 [18, 20, 21]. An understanding of the concept of Bloch periodic structures will become relevant when we want to tailor waveguides by changing the geometry through patterning.

3

Nonlinear optics

In this chapter, we present the theory of nonlinear optics. We start with a basic description of what kind of nonlinearity we refer to with nonlinear optics. Thereafter the theory needed to understand the concept of optical parametric amplification (OPA) is presented. To reach this goal, we will treat nonlinear effects of materials as a perturbation and show how these combined effects can be used to understand the intermodulation phenomenon four-wave mixing (FWM), which is used to describe the concept of OPA. Readers on the hunt for general books on the subject of nonlinear optics are advised to read the books (all named "Nonlinear optics") by Moloney and Newell [22], Robert Boyd [23], and Nicolaas Bloembergen [24]. I can also recommend the video series by the International School on Parametric Nonlinear Optics (ISPNLO), held in 2015 [25].

3.1 Nonlinear effects in materials

With the term "nonlinear optics", we refer to effects that stem from the permittivity and permeability of a material depending on the field strength:

$$\begin{aligned} \mathbf{D} &= \varepsilon(\mathbf{E})\mathbf{E}, \\ \mathbf{B} &= \mu(\mathbf{H})\mathbf{H}. \end{aligned} \tag{3.1}$$

Here we will focus on the electric field. Recall that we can express the displacement field as the electric field in free space plus a polarisation term:

$$\mathbf{D} = \varepsilon(\mathbf{E})\mathbf{E} = \varepsilon_0\mathbf{E} + \varepsilon_0\chi\mathbf{E}. \tag{3.2}$$

We then express the polarisation as a power series expansion, starting with scalar notation to simplify the introduction:

$$P = \varepsilon_0 \left[\chi^{(1)} E + \chi^{(2)} E^2 + \chi^{(3)} E^3 + ... \right] \\ = P^{(1)} + P^{(2)} + P^{(3)} + ... \tag{3.3}$$

From Equation (3.3), we see that the nonlinear dependence emerges from the higher orders in the expansion. $\chi^{(1)}$ is the normal linear susceptibility recognized from Equation (2.3), while $\chi^{(2)}$ and $\chi^{(3)}$ are the second- and third-order susceptibilities etc. From the expansion, we see that higher orders are more sensitive to field strength and the scale of the higher-order contributions to the polarisation depend on the susceptibilities.

Let us put some order of magnitude on these constants, so we can get a feeling of how they affect the fields. We can estimate these values by calculating the characteristic atomic field strength $E_{\text{atom}} = e/(4\pi\varepsilon_0 a_0^2)$, where e is the elementary charge and $a_0 = 4\pi\varepsilon_0\hbar^2/(me^2)$ is the Bohr radius of hydrogen and m is the electron mass. Plugging in the values yields $E_{\text{atom}} = 5.14 \times 10^{11}$ V/m. The first-order susceptibility, $\chi^{(1)}$, is of the order of unity for condensed matter, and we can estimate the orders of $\chi^{(2)}$ and $\chi^{(3)}$ with $\chi^{(1)}/E_{\text{atom}}$ and $\chi^{(1)}/E_{\text{atom}}^2$, respectively [23]. The orders of magnitude are presented in Table 3.1.

Table 3.1: Table showing order of magnitudes for the first three orders of susceptibilities [23].

Susceptibility $\chi^{(n)}$	Typical order of magnitude
$\chi^{(1)}$	10^0
$\chi^{(2)}$	10^{-12} m/V
$\chi^{(3)}$	10^{-24} m^2/V^2

From Table 3.1 and Equation (3.3), we can deduce that for these effects to appear, we clearly need to either have large field strengths or letting the light pass through large volumes of material, so it can accumulate enough nonlinear phase.

Before we go into a more thorough description of the nonlinear effects, it is worth pointing out that what we are talking about here is electronic polarisation, effects induced on the electrons in the materials under the influence of an electric field. In fact, the nonlinear effects in materials can be caused by different physical mechanisms, as summarised by Boyd [23, 26] and presented in Table 3.2. The focus of this thesis is on electronic polarisation, sometimes referred to as instantaneous polarisation due to its short response time, but there are other nonlinear effects that operate on longer time scales, the extremes being heating effects in materials that have response times on the order of milliseconds.

Moving back to electronic polarisation, we can describe the general polarisation field as a combination of multiple electric fields, each with its own frequency, which may or

Table 3.2: Values of the nonlinear refractive index for linear polarised light [23]. The response time is the time it takes to reach the steady state after we 'turn on our laser'.

Nonlinear mechanism	n_2 [cm^2/W]	$\chi^{(3)}$[m^2/V^2]	Response time [s]
Electronic polarisation	10^{-16}	10^{-22}	10^{-15}
Molecular orientation	10^{-14}	10^{-20}	10^{-12}
Electrostriction	10^{-14}	10^{-20}	10^{-9}
Saturated atomic absorption	10^{-10}	10^{-16}	10^{-8}
Thermal effects	10^{-6}	10^{-12}	10^{-3}

may not be the same,

$$\mathbf{P} = \varepsilon_0 \left[\chi^{(1)}\mathbf{E}_1 + \chi^{(2)}\mathbf{E}_1\mathbf{E}_2 + \chi^{(3)}\mathbf{E}_1\mathbf{E}_2\mathbf{E}_3 + ... \right], \tag{3.4}$$

which opens up for interaction between different frequencies, such as intermodulation.

To understand this intermodulation, we need to start from the time dependent expression for the linear polarisation field where the polarisation is expressed as a convolution of the time-dependent susceptibility and the time-dependent electric field.

$$\mathbf{P}(t) = \varepsilon_0 \int_{-\infty}^{\infty} \chi(t-\tau)\mathbf{E}(\tau)d\tau. \tag{3.5}$$

From causality we require that $\chi(t-\tau) = 0$ for $\tau > t$, and since complex values of χ are nonphysical, the reality conditions require χ to be real [27]. For higher-orders, the susceptibility becomes a tensor and similarly to the expression in Equation (3.5), we can express the polarisation in Equation (3.4) for its j:th component as:

$$\begin{aligned} P_j(t) = {} & \varepsilon_0 \int_{-\infty}^{\infty} \chi^{(1)}_{jk}(t-\tau_1)E_k(\tau_1)d\tau_1 \\ & + \varepsilon_0 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \chi^{(2)}_{jkl}(t-\tau_1, t-\tau_2)E_k(\tau_1)E_l(\tau_2)d\tau_1 d\tau_2 \\ & + \varepsilon_0 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \chi^{(3)}_{jklm}(t-\tau_1, t-\tau_2, t-\tau_3)E_k(\tau_1)E_l(\tau_2)E_m(\tau_3)d\tau_1 d\tau_2 d\tau_3 \\ & + ... \end{aligned} \tag{3.6}$$

where we are using the Einstein summation convention [28]. This expression may seem a bit confusing at first, but some interesting things happen when we view this expression in the frequency domain. Before presenting this expression, recall that the exponential representation of a Dirac delta function δ can be written as [29, 30]

$$\delta(x - x_0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ik(x-x_0)}dk. \tag{3.7}$$

With this, we can express the polarisation by Fourier transforming Equation (3.6) yielding

$$\begin{aligned}\hat{P}_j(\omega) = \varepsilon_0 \hat{\chi}^{(1)}_{jk}(\omega)\hat{E}_k \\ + \frac{\varepsilon_0}{2\pi}\iint_{-\infty}^{\infty} \hat{\chi}^{(2)}_{jkl}(\omega_1,\omega_2)\hat{E}_k(\omega_1)\hat{E}_l(\omega_2)\delta(\omega_1+\omega_2-\omega)d\omega_1 d\omega_2 \\ + \frac{\varepsilon_0}{(2\pi)^2}\iiint_{-\infty}^{\infty} \hat{\chi}^{(3)}_{jklm}(\omega_1,\omega_2,\omega_3)\hat{E}_k(\omega_1)\hat{E}_l(\omega_2)\hat{E}_m(\omega_3)\delta(\omega_1+\omega_2+\omega_3-\omega)d\omega_1 d\omega_2 d\omega_3 \\ + ...\end{aligned} \tag{3.8}$$

From this expression, we can deduce that we have requirements on the frequencies to match for second-and higher-order effects from the delta functions. For example, for third-order susceptibilities, we have the requirement that

$$\omega_1 + \omega_2 + \omega_3 - \omega = 0, \tag{3.9}$$

for the effects to appear. This is something that we will use in the context of FWM, which we will treat in Sec. 3.2. We can also deduce multiple symmetry requirements of the tensors which is very useful since $\chi^{(3)}$, for example, has 3^4 different elements. For details regarding symmetry derivations, see, for instance, Refs. [23, 27].

3.2 Four-wave mixing

From Equation (3.8) in Sec. 3.1, we could see that we had requirements on the frequencies. For the third-order process, we have 4 different frequencies that need to obey Equation (3.9). The interaction between these four frequencies is what is called four-wave mixing (FWM). If we assume we have a material in which we can express the susceptibility as a scalar quantity, we can write out all the possible complex amplitudes for third-order nonlinear interactions. This will help with the intuition of the effects. Negative frequencies are allowed and we interpret them as the complex conjugate of the field. We end up with the following expressions where we combine all possible combinations that result in a polarisation contribution at one frequency [23]:

$$\begin{aligned} P(\omega_1) &= \varepsilon_0\chi^{(3)}\left(3E_1E_1^* + 6E_2E_2^* + 3E_3E_3^*\right)E_1, & P(3\omega_1) &= \varepsilon_0\chi^{(3)}E_1^3, \\ P(\omega_2) &= \varepsilon_0\chi^{(3)}\left(6E_1E_1^* + 3E_2E_2^* + 3E_3E_3^*\right)E_2, & P(3\omega_2) &= \varepsilon_0\chi^{(3)}E_2^3, \\ P(\omega_3) &= \varepsilon_0\chi^{(3)}\left(6E_1E_1^* + 6E_2E_2^* + 3E_3E_3^*\right)E_3, & P(3\omega_3) &= \varepsilon_0\chi^{(3)}E_3^3, \end{aligned}$$

$$
\begin{aligned}
P(\omega_1+\omega_2+\omega_3) &= 6\varepsilon_0\chi^{(3)}E_1E_2E_3, & P(\omega_1+\omega_2-\omega_3) &= 6\varepsilon_0\chi^{(3)}E_1E_2E_3^*,\\
P(\omega_1-\omega_2+\omega_3) &= 6\varepsilon_0\chi^{(3)}E_1E_2^*E_3, & P(-\omega_1+\omega_2+\omega_3) &= 6\varepsilon_0\chi^{(3)}E_1^*E_2E_3,
\end{aligned}
$$

$$
\begin{aligned}
P(2\omega_1+\omega_2) &= 3\varepsilon_0\chi^{(3)}E_1^2E_2, & P(2\omega_1+\omega_3) &= 3\varepsilon_0\chi^{(3)}E_1^2E_3,\\
P(2\omega_2+\omega_1) &= 3\varepsilon_0\chi^{(3)}E_2^2E_1, & P(2\omega_2+\omega_3) &= 3\varepsilon_0\chi^{(3)}E_2^2E_3,\\
P(2\omega_3+\omega_1) &= 3\varepsilon_0\chi^{(3)}E_3^2E_1, & P(2\omega_3+\omega_2) &= 3\varepsilon_0\chi^{(3)}E_3^2E_2,\\
P(2\omega_1-\omega_2) &= 3\varepsilon_0\chi^{(3)}E_1^2E_2^*, & P(2\omega_1-\omega_3) &= 3\varepsilon_0\chi^{(3)}E_1^2E_3^*,\\
P(2\omega_2-\omega_1) &= 3\varepsilon_0\chi^{(3)}E_2^2E_1^*, & P(2\omega_2-\omega_3) &= 3\varepsilon_0\chi^{(3)}E_2^2E_3^*,\\
P(2\omega_3-\omega_1) &= 3\varepsilon_0\chi^{(3)}E_3^2E_1^*, & P(2\omega_3-\omega_2) &= 3\varepsilon_0\chi^{(3)}E_3^2E_2^*.
\end{aligned}
\tag{3.10}
$$

So how do we interpret this? What we see is a manifestation of a parametric process that can be described by an energy diagram presented in Fig. 3.1. Here the initial and final energy states are the same, so if we have two photons with frequency ω_1 and ω_2 entering the system, they can briefly excite a virtual energy level. Conservation of energy means that when the energy state relaxes, the energy needs to radiate out again. The result is a combination of photons that contains the same total energy entering and leaving the system, forming a parametric process.

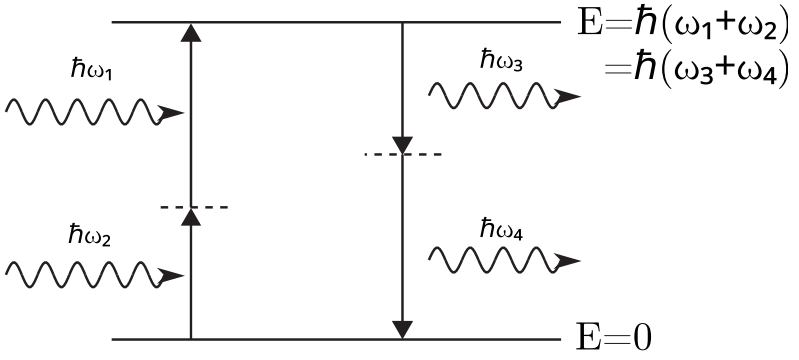


Figure 3.1: Energy levels of a four wave mixing process. The energy of the incoming photons (ω_1 & ω_2) adds up to a virtual energy level that later relaxes into two new photons with different frequencies (ω_3 & ω_4).

Let us, for instance, take the first term in the expression for $P(\omega_1)$ in Equation (3.10). The only way for the system to produce a photon with frequency ω_1 is to excite the energy state with energy from ω_1 and then add another frequency that returns during the relaxation. This would result in a situation where the photons interacting in Fig. 3.1 would obey $\omega_2 = \omega_3$ and $\omega_1 = \omega_4$.

A similar argument can be made for second-order effects and we would then end up with another set of relations like in Equation (3.10), but instead of the four frequencies

we treat here we would have three different frequencies. For details, see for example Ref. [23].

The idea of mixing different frequencies in a nonlinear material was known early on, but the field of nonlinear optics really took off in the 1960s when Franken, Hill, Peters and Weinreich experimentally demonstrated the generation of optical harmonics in 1961 [31]. They used a second-order nonlinearity to generate a harmonic of a pump frequency, where two photons of frequency ω are converted into one photon with frequency 2ω, so-called second harmonic generation (SHG). The same year, Kaiser and Garett also experimentally showed the nonlinear optical phenomenon "two-photon absorption" in CaF_2[32], where 2 photons through a virtual energy level could reach a higher energy state in the material and produce a higher-frequency photon. The theory of this operation was already put forth by Göppert-Mayer in 1931 [33], but it took 30 years before it could be measured. Once SHG was demonstrated, the concept of third harmonic generation (THG) was quickly demonstrated. In THG, we combine three photons to create one photon with thrice the frequency, 3ω, through a third-order nonlinear process. This was demonstrated in 1962 by Terhune, Maker and Savage [34]. The same year, Bloembergen, Armstrong, Ducuing, and Pershan used quantum mechanical perturbation theory to first theoretically describe the process that Franken et al. had demonstrated in 1961 [35].

3.2.1 Phase matching

We saw in the previous section that we got a requirement on energy conservation for FWM to work, which resulted in a relation between frequencies in the mixing. We also have another requirement on the waves, with different frequencies, that interact and that is referred to as phase matching. This means that for the waves to interact and properly mix they need to be close in phase.
For a four wave mixing process, like in Fig. 3.1 we have a phase matching condition of

$$\Delta\beta = (\beta_1 + \beta_2 - \beta_3 - \beta_4)\,. \tag{3.11}$$

Ideally the mismatch should be zero, which in turn would mean that we have conservation of momentum for the photons. There are different ways to handle this depending on material properties. Midwinter and Warner defined two types of phase matching for second-order processes for birefringent materials in 1962 [36], which uses the crystal orientation to give different dispersion for different frequencies to counteract unwanted dispersion. Another approach to phase matching is to change the propagation constant by changing the angle at which the beams enter the material, as demonstrated by Hobden [37]. The above approaches uses slabs of materials but we can also achieve phase matching over a periodic structure. This approach was originally proposed by Bloembergen, Armstrong, Ducuing, and Pershan in 1962 [35] and is named quasi-phase

matching (QPM). The use of QPM has many applications and opens up many possibilities in combination with better materials engineering [38, 39]. For a review of different quasi-matched materials, see for instance Ref. [40]. The idea is that one engineers the dispersion of the material so that at the end of each periodic segment the phase mismatch is negated. For this approach to work efficiently, the length of these unit cells in the periodic structure needs to be sufficiently smaller than the coherence length L_C:

$$L_C = \frac{\pi}{\Delta k}. \tag{3.12}$$

We can tailor the QPM over longer distances than L_c but at the expense of weaker nonlinear interaction. This can be relevant in situations where it is hard to manufacture the material thin enough and we can then simplify the manufacturing at the expense of weaker nonlinear interaction per distance. Another reason might be that we have unwanted nonlinear processes with other frequency components that we want to remove, then we can design the material such that phase matching for the unwanted process is unfavourable. The process of manufacturing these periodic materials often results in distributed noise that covers a wide frequency spectrum where basically everything is a little bit phase matched due to manufacturing tolerances [41]. A good resource for QPM tuning is Fejer and Byer [42]. If one wants to utilise FWM using pulsed waves, the phase matching with first-order dispersion is usually not enough – we also need to take the higher-order dispersion, like group velocity into account, but if we are operating in a continuous wave (CW) or quasi-CW regime, these effects can often be neglected [43].

3.3 Optical parametric amplifiers

In practical terms, we can use the concept of FWM described in Section 3.2 to amplify a desired signal, by means of difference frequency generation, also called optical parametric amplification (OPA). What we want is to use the energy of a pump laser and transfer it to a desired signal. To achieve this, we need to ensure that when the excited state, as described in Fig. 3.1, decays into two different photons, we want to tailor it so that one of the outgoing photons has the frequency of the signal. The way to handle this most easily is to ensure that the mixing allows the pump to interact with itself to form a higher energy state and then relax down into the signal frequency. This in turn means that the frequency of the signal must be lower than twice the frequency of the pump. As a result of the mixing process, we will also have another photon from the relaxation that has a frequency different from the pump and signal. This is called the idler frequency. The concept of OPA is depicted in Fig. 3.2.

The effect of OPA was first measured by Harris, Oshman and Byer in 1967 [44], but the optical effect was treated theoretically by Siegman, Yariv and Louisell earlier in 1961 [45]. The OPA has an advantage regarding its amplification, compared to other techniques,

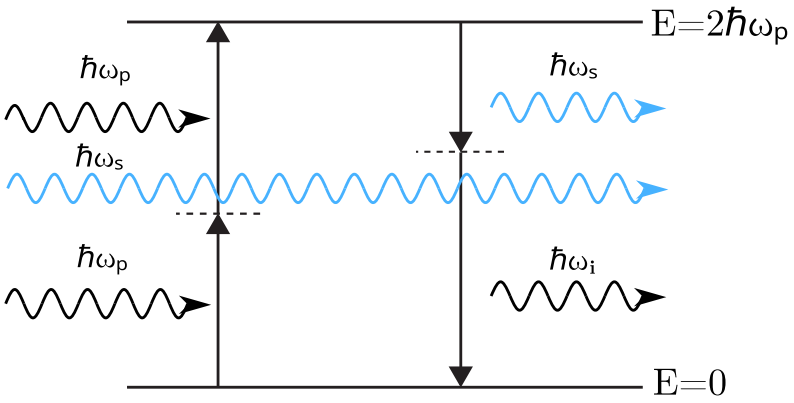


Figure 3.2: Energy levels for a four wave mixing process in a optical parametric amplifier. Two photons of the pump laser, with frequency ω_p, form an energy state that relaxes into two photons for the signal (ω_s) and idler (ω_i) frequency. The signal photon amplifies the existing signal photons that passes through the system, indicated in blue.

due to the fact that it has very low noise [45, 46]. The downside of the OPA, as we have seen from expression of $\chi^{(3)}$, is that the interaction is very weak, and we therefore need to propagate the light through a large volume of material to accumulate nonlinear phase. Since the light needs to interact with a large volume it also attenuates due to losses in the material. For amplification to occur we therefore must ensure that the losses in the material are low enough so it doesn't negate the amplification. Due to the length of material required for OPA, they are usually constructed in the form of fibres and on chip waveguides [47].

3.3.1 Phase-sensitive amplifier

When using optical parametric amplification, there are two types of amplifiers that are common and that is phase-insensitive amplifiers (PIAs) and phase-sensitive amplifiers (PSAs). As the name suggests, the difference between the two is that the gain in phase-sensitive amplifier (PSA) is dependent on the absolute optical phase of the incoming wave, while this is not important for PIA. A common example of PIA is the erbium-doped fiber amplifier (EDFA) [48]. In general, optical amplification is limited by thermal noise, which results in a lowest possible noise figure (NF) of 2 dB [46], with typical commercial EDFA having NF $\approx$ 5 dB [49]. A PSA, however, can operate under certain conditions where it doesn't generate excess noise, meaning it has a quantum-limited NF = 0 dB [46, 50]. Implementations of different types of PSA can be seen in Refs. [49, 51–60].

In order to build some understanding of how an OPA process looks like when measured, we can look at an example taken from Ref. [56]. Here a PSA is realised using highly nonlinear fiber (HNLF) and a setup where two pumps are used to amplify a signal. Schematically the system behaves as shown in Fig. 3.3, where unwanted signals

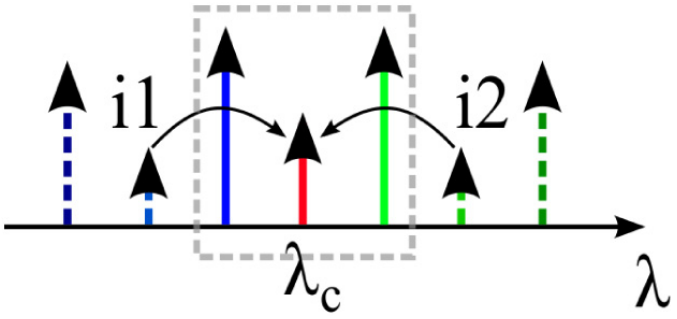


Figure 3.3: Figure depicting a FWM mixing process using two pumps centred round centre signal. The dashed lines represent unwanted FWM signals including the higher order idler, i1, i2. Reprinted from Ref. [56] with permission from Optica.

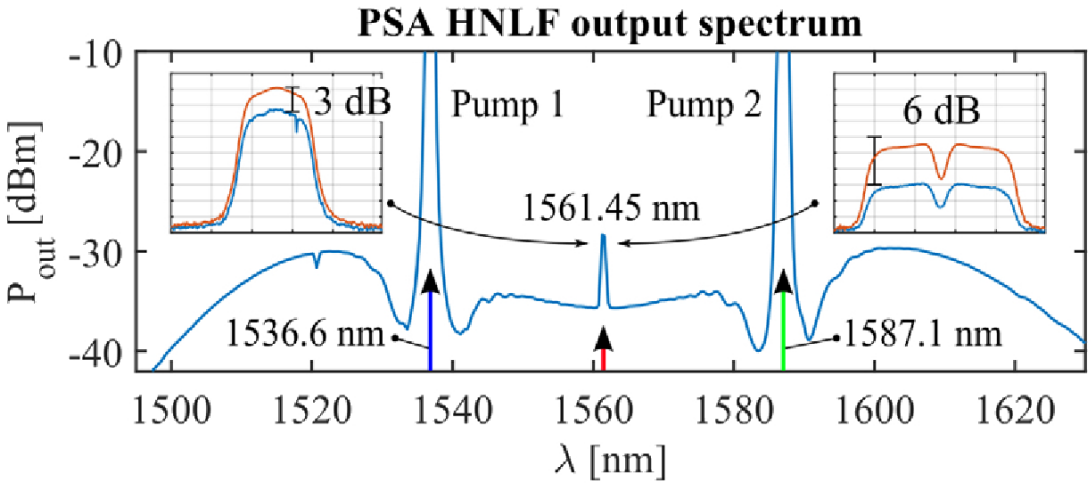


Figure 3.4: Figure depicting the measured optical spectrum of a PSA based on HNLF with two pumps. Line colour on inset to the right highlight the difference between PIA (blue) and PSA (orange) and for the inset on the right it highlights difference between unlocked (blue) and phase-locked (orange) amplified binary-phase-shift keying (BPSK) signal. Reprinted from [56] with permission from Optica.

from the FWM process is located outside the pumps.

How this system looks while measured is depicted in Fig. 3.4. As we can see, mixing processes resulting from combinations of all components combined with the fact that signals have a bandwidth and accumulated noise creates regions with higher unwanted signal strength. In order to keep the NF down for the signal, one wants to suppress these unwanted signals. As mentioned in Sec. 3.3, creating favourable circumstances for desired wavelengths and unfavourable circumstances for unwanted wavelengths contribute to a better NF. In the rightmost inset in Fig. 3.4, we can also see an example of the difference between PIA and PSA that can be achieved.

Ensuring that the conditions for this type of process for the purpose of on-chip design waveguide for PSA is part of the goal of papers III & IV. In paper III, we focus on simulation techniques to simulate dispersion and nonlinearity for 3D structures and in paper IV, machine learning and 2D mode simulations of waveguide cross sections are used in order to achieve good phase matching and nonlinearity for power transfer to a signal while also keeping conditions unfavourable for unwanted mixing processes.

4 Electromagnetic simulations

In this chapter, we will go through the different types of numerical methods needed to achieve the goals, presented in Ch. 1, of simulating arbitrary patterned Bloch periodic waveguides using the theory described in Ch. 2 and 3. The chapter will describe the use of nonlinear Schrödinger equations (NLSE) to simulate the propagation over long distances and how we will use finite element methods to determine the constants that we need in the NLSE. For a good overview of numerical methods in the field of photonics, see for instance Obayya [61].

4.1 Simulating the propagation of optical signals in waveguides

In this section, we will discuss how we can simulate nonlinear wave propagation over longer distances. When we are after results of wave propagation over long photonic structures spanning up to metres in length, as used in Ref. [1], full-wave simulations of the entire waveguide structure becomes computationally infeasible without some simplification. In the following sections in this chapter, we will introduce how this can be done by solving a NLSE to get an expression for how the excited wave propagates along the longitudinal direction of the waveguide.

4.1.1 Nonlinear Schrödinger equation

We will here go through the steps taken in order to arrive at the NLSE used to calculate the wave propagation. As the name suggest, it is an extended variation of the Schrödinger equation, first published by Erwin Schrödinger i 1926 [62]. The NLSE was first written down in 1962 by Townes, Garmire and Chiao [63], but is in itself a simplification of the

Ginzburg-Landau equation first published in 1959 by Landau and Ginzburg [64, 65]. For further details of the derivations we refer to the work by Marhic [66] and Agrawal [43]. The derivation will be made with just linear components and the nonlinear contributions will be added afterwards as a perturbation. In Sec. 2.1 we discussed the concept of dispersion and it will, here, play a big role in setting up the NLSE. For the problems we are interested in, we want to address a range of frequencies, ω, for the solution to work we need them centred around a centre frequency, ω_c and the difference

$$\Omega = \omega - \omega_c \tag{4.1}$$

to be sufficiently small. We can the express the dispersion as a power series:

$$\beta(\omega) = \beta(\omega_c + \Omega) = \beta_c + \sum_{n=1}^{\infty} \frac{\beta^{(n)}}{n!}\Omega^n. \tag{4.2}$$

where

$$\beta_c = \beta(\omega_c) \quad \text{and} \quad \beta^{(n)} = \left.\frac{d^n\beta}{d\omega^n}\right|_{\omega=\omega_c}. \tag{4.3}$$

We express the electric field along the longitudinal direction, $\hat{z}$, with the slowly varying envelope $\mathbf{A}(z,t)$.

$$\mathbf{E}(z,t) = \frac{1}{2}\left[\mathbf{A}(z,t)e^{i(\beta_c z - \omega_c t)} + \text{c.c.}\right] \tag{4.4}$$

where c.c. stands for the complex conjugate of the field. This is an assumption that A varies slowly over time and that most change occurs in the propagation direction.

In order to reach the final expression, we will first decompose $\mathbf{A}$ in terms of its Fourier decomposition, $\mathbf{B}(\Omega)$. Then we will differentiate the expression and combine them to obtain our results.

At the start, we have our envelope

$$\mathbf{A}(0,t) = \int_{-\infty}^{\infty} \mathbf{B}(\Omega)e^{-i\Omega t}d\Omega. \tag{4.5}$$

Since we assume linear propagation along the longitudinal direction we can describe the envelope at any position along the propagation axis as

$$\mathbf{A}(z,t) = \int_{-\infty}^{\infty} \mathbf{B}(\Omega)e^{i(\beta-\beta_c)z - i\Omega t}d\Omega. \tag{4.6}$$

We then take the derivative of Equation (4.6), which yields

$$\frac{\partial\mathbf{A}(z,t)}{\partial z} = \int_{-\infty}^{\infty} \mathbf{B}(\Omega)i(\beta-\beta_c)e^{i(\beta-\beta_c)z - i\Omega t}d\Omega. \tag{4.7}$$

From Equation (4.2), we know that we can replace the parts containing $\beta - \beta_c$ and we get

$$\frac{\partial \mathbf{A}(z,t)}{\partial z} = i \int_{-\infty}^{\infty} \mathbf{B}(\Omega) \left(\sum_{n=1}^{\infty} \frac{\beta^{(n)}}{n!} \Omega^n \right) e^{i(\beta-\beta_c)z - i\Omega t} d\Omega. \tag{4.8}$$

If we then were to take the m:th derivative of (4.6) with respect to t, we get

$$\frac{\partial^m \mathbf{A}(z,t)}{\partial t^m} = \int_{-\infty}^{\infty} \mathbf{B}(\Omega)(-i\Omega)^m e^{i(\beta-\beta_c)z - i\Omega t} d\Omega. \tag{4.9}$$

Comparing Equations (4.9) and (4.8), we note that we can replace the integral in Equation (4.8) with a sum of time derivatives, which results in

$$\frac{\partial \mathbf{A}(\mathbf{z},\mathbf{t})}{\partial z} - \sum_{n=1}^{\infty} \frac{i^{n+1}\beta^{(n)}}{n!} \frac{\partial^n \mathbf{A}(\mathbf{z},\mathbf{t})}{\partial t^n} = 0, \tag{4.10}$$

which is our expression for the wave propagation taking only dispersion into account. To reach our final expressio,n we need to add the loss of the propagation constant, α, and the nonlinear part. The loss is incorporated by adding the term $\alpha/2$ to the dispersion. The nonlinear effect is added using a nonlinear parameter γ to achieve the full NLSE:

$$\frac{\partial \mathbf{A}(z,t)}{\partial z} + \frac{\alpha}{2}\mathbf{A}(z,t) - \sum_{n=1}^{\infty} \frac{i^{n+1}\beta^{(n)}}{n!} \frac{\partial^n \mathbf{A}(z,t)}{\partial t^n} = i\gamma |\mathbf{A}(z,t)|^2 \mathbf{A}(z,t). \tag{4.11}$$

When we are operating in a regime where the frequency components are close together we can assume that the nonlinear part has a constant frequency dependence and if we in turn have a system where the nonlinearity has a low spatial dependency, we can simplify the expression of γ to be

$$\gamma = \frac{2\mu_0 \omega \chi^{(3)}}{8\bar{n}^2 A_{\text{eff}}}, \tag{4.12}$$

where $\bar{n}$ is the effective refractive index

$$\beta = \bar{n}\frac{2\pi}{\lambda} \tag{4.13}$$

and the effective area, A_{eff}, is a measure on the transverse mode profile overlap [43]:

$$A_{\text{eff}} = \frac{\left[\iint (\psi^* \psi) dxdy\right]^2}{\iint (\psi^* \psi)^2 dxdy}. \tag{4.14}$$

Here ψ denote the transverse component of the electric field, with dependence on x and y. What Equation (4.12) and (4.14) describe is that we can get a nonlinear effect either

by choosing a material with high $\chi^{(3)}$ or by ensuring that the field is concentrated in the cross section as much as possible. To be able to achieve stable solutions for the NLSE, so called solitons, requires us to either tailor the spatial details of our system or tweak the time dependence, like sending pulses of light. The first soliton solution achieved was a spatial soliton, realised by Garmire, Chiao and Townes in 1964 [67] and the first temporal soliton solution using pulses of light was realised in optical fibres by Tappert and Hasegawa in 1973 [68]. In order for us to use the NLSE we need to have values for the constants present in Eq. (4.11) and in paper III we present methods of how this can be achieved for complex 3D geometries in a computationally efficient way.

4.1.2 Split-step Fourier method

In order to solve the NLSE for the wave propagation, described in Sec. 4.1.1, we can use the split-step Fourier method (SSFM) first developed by Tappert in 1972 at Bell Laboratories [69]. The method is based on that we separate the components of the NLSE (4.11) into a linear and nonlinear part and then assume that they can act separately on the wave function during small steps in the longitudinal direction. To speed this up one then lets the linear part act in the frequency domain and the nonlinear part in the time domain. The process goes as follows [43]:

We take Equation (4.11) and express it like

$$\frac{\partial \mathbf{A}(z,t)}{\partial z} = \left[-\frac{\alpha}{2} + \sum_{n=1}^{\infty} \frac{i^{n+1}\beta^{(n)}}{n!} \frac{\partial^n}{\partial t^n} + i\gamma |\mathbf{A}(z,t)|^2 \right] \mathbf{A}(\mathbf{z},\mathbf{t}) = \left[\hat{D} + \hat{N} \right] \mathbf{A}(z,t), \tag{4.15}$$

where $\hat{D}$ expresses an operation representing the linear part of the NLSE, while $\hat{N}$ represent the nonlinear part. If we then take a small enough step, h, in the longitudinal z direction, we assume that the numerical error induced by separating $\hat{D}$ and $\hat{N}$ is small. This means that we can write the solution for propagation of a small step, h, as

$$A(z+h,t) \approx e^{h\hat{D}} e^{h\hat{N}} A(z,t). \tag{4.16}$$

We then note from Equation (4.15) that the linear part $\hat{D}$ contains all the time derivatives. These are cumbersome to calculate, but we note that the time derivatives become much more simple to calculate in the frequency domain

$$\mathscr{F}\left\{\hat{D}\right\} = \mathscr{F}\left\{ -\frac{\alpha}{2} + \sum_{n=1}^{\infty} \frac{i^{n+1}\beta^{(n)}}{n!} \frac{\partial^n}{\partial t^n} \right\} = -\frac{\alpha}{2} + \sum_{n=1}^{\infty} \frac{i^{n+1}\beta^{(n)}}{n!} (i\omega)^n, \tag{4.17}$$

where $\mathscr{F}$ denotes the Fourier transform and the transform gives $(i\omega)$ in place of the time derivatives. Thus, the easiest way to calculate is to propagate the linear part in the frequency domain and the nonlinear part in the time domain.

Putting it together, we then get the expression for one spatial step, h, to be

$$A(z+h,t) = \mathscr{F}^{-1}\left\{e^{h\mathscr{F}\{\hat{D}\}}\mathscr{F}\left\{e^{h\hat{N}}A(z,t)\right\}\right\}, \tag{4.18}$$

which has an accuracy to the second-order in the step size h [43].

We can further increase the accuracy of the SSFM by using a staggered step where the nonlinear part is allowed to act on A at the midpoint between steps, $h/2$. This is known as the symmetrized split-step Fourier method, illustrated in Fig. 4.1, first used by Feit, Fleck, and Morris in 1976 [70].

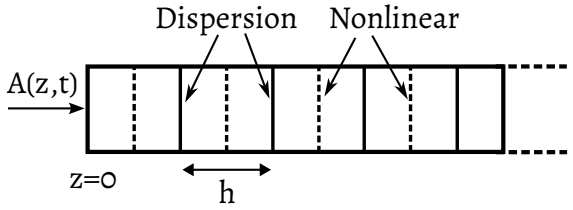


Figure 4.1: Illustration of symmetrized split-step Fourier method, showing the placement in the geometry when dispersion and nonlinearity are added.

The symmetrized SSFM instead expresses Equation (4.16) as two half linear propagation steps on the boundaries and one full nonlinear step in the segment

$$A(z+h,t) \approx e^{\frac{h}{2}\hat{D}}\exp\left(\int_z^{z+h}\hat{N}(z')dz'\right)e^{\frac{h}{2}\hat{D}}A(z,t). \tag{4.19}$$

When we use a small step size, the integral in Eq. (4.19) can be approximated with

$$\exp\left(\int_z^{z+h}\hat{N}(z')dz'\right) \approx e^{h\hat{N}}. \tag{4.20}$$

We can then use a trick to speed up the calculations. By combining Equations (4.20) and (4.19), we can rewrite them as

$$A(z+h,t) \approx e^{\frac{h}{2}\hat{D}}e^{h\hat{N}}e^{\frac{h}{2}\hat{D}}A(z,t) = e^{-\frac{h}{2}\hat{D}}\left(e^{h\hat{D}}e^{h\hat{N}}\right)e^{\frac{h}{2}\hat{D}}A(z,t). \tag{4.21}$$

From Equation (4.21) we note that the middle section contains a full step length. This means that for the propagation over a longer length containing M steps we can write the final expression of A as

$$A(z+Mh,t) \approx e^{-\frac{h}{2}\hat{D}}\left(\prod_{m=1}^{M}e^{h\hat{D}}e^{h\hat{N}}\right)e^{\frac{h}{2}\hat{D}}A(z,t). \tag{4.22}$$

This means that, apart from the beginning and end, we can evaluate the linear and nonlinear contribution with the same step length, cutting the number of Fourier transforms down by almost a factor of 2 [43].

4.2 Determining propagation and dispersion constants

This section will describe the use of the finite element method to determine the dispersion for a Bloch periodic waveguide. Recall from Sec. 4.1 that we need these values if we want to describe the behaviour of the nonlinear effects and solve the NLSE. By utilizing the methods described in this section we will be able to use full-wave linear simulations to determine nonlinear respones of systems by combining the numerical results with the theory in Sec. 4.1.

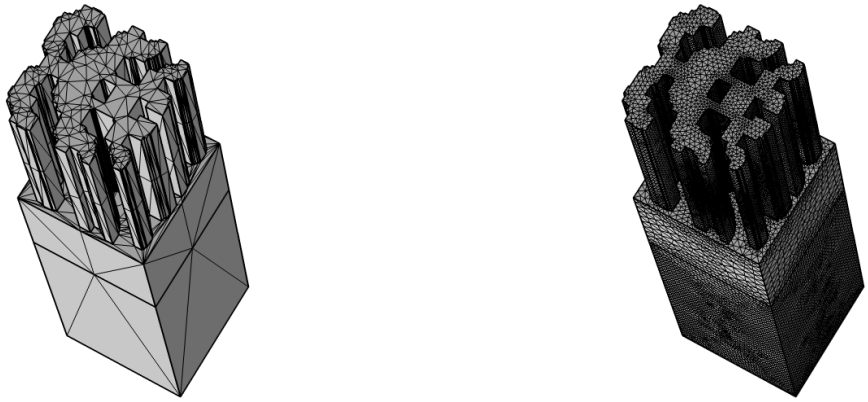
We will introduce the finite element method and how we can use it to solve for attenuation dispersion by the use of eigenmode simulations. The overarching goal with this section is to present the tools needed to simulate these waveguide and provide an understanding of what different steps in the process do, while providing references for further reading on the subject.

4.2.1 Finite element method

Since it was first conceived in 1943 by Courant [71], the finite element method (FEM) has grown to be a very popular and useful numerical tool to solve partial differential equations (PDE). The method works by subdividing the geometry into smaller mesh cells and approximate the solution with a finite number of parameters, referred to as degrees of freedom (DOF). The process of subdividing the geometry is called meshing, where we partition the geometry with a mesh consisting of polygons and polyhedrons. The physics calculations are then performed on these elements and the size of the elements determine how accurately the physics can be described. A good reference on the use of FEM in electromagnetism is Ref. [72]. This thesis will not go into details of the exact workings of the finite element method, but will treat subjects needed to understand and use FEM for solving for the dispersion in optical waveguides. For a thorough understanding of finite element solvers, read for example Ottosen and Petersson [73], but also manuals from commercial FEM packages usually have a good explanation, see for example COMSOL [74–76].

Having many elements results in more calculations to be performed since we need to determine more DOF. It is always of interest to keep the DOF down, both to cut down the simulation times and also to be able to utilise available memory better. In paper I [77], we presented a meshing strategy based on *a posteriori* information of the calculated electric fields on the domain, implemented in COMSOL [78]. By starting with a coarse mesh and then refining mesh elements that exhibit larger errors, we can ensure that we only refine elements that needs refining to resolve the electromagnetic behaviour, as demonstrated in Fig. 4.2.

In the following subsections, we will go through how we convert the strong form of the



(a) (b)

Figure 4.2: Demonstration of a mesh refinement in a 3D structure were regions with smaller features are resolved with a denser mesh because of the need to resolve the electromagnetic fields.

PDE of our problem into the weak form (Sec. 4.2.1.1), how we deal with different boundaries for our simulation domain (Sec. 4.2.2) and how we rewrite our electromagnetic expression in order to solve for for eigenmodes at a specified frequency (Sec. 4.2.3).

4.2.1.1 Weak expression

When solving for a physics based PDE with FEM, we do not need to use the original form of the PDE, called the strong form. Instead, we express the variational form of the PDE, called the weak form.

The concept revolves around expressing the original PDE as an integral equation. The solutions of interest are then approximated with a set of basis functions on small sections of the integral. We then multiply the original PDE expression with a test function, this test function can then be tweaked and chosen in a way that we can can approximate the value of our solution over the integrated area by computing smaller elements of the integral individually.

To arrive at the weak expression needed to solve for the electric field, we start with the wave equation for the electric field:

$$\nabla \times \left(\frac{1}{\mu} \nabla \times \mathbf{E} \right) - \varepsilon \frac{\omega^2}{c^2} \mathbf{E} = 0. \tag{4.23}$$

We then get the weak formulation by multiplying the expression in Equation (4.23) with a test function (sometimes called weight function), $\mathbf{v}$, to yield:

$$F_E(\mathbf{v}, \mathbf{E}) = (\nabla \times \mathbf{v}) \cdot \frac{1}{\mu} (\nabla \times \mathbf{E}) - \varepsilon \frac{\omega^2}{c^2} \mathbf{v} \cdot \mathbf{E} = 0, \tag{4.24}$$

which then is solved for in the context of an integral equation over our domain.

$$\iiint F_E(\mathbf{v}, \mathbf{E}) = 0. \tag{4.25}$$

The test function can be any function and Equations (4.24) and (4.25) still hold, because we have multiplied the test function with an expression that was zero and then integrated the expression that was zero.

What we gain from this formulation is not obvious at first glance, but this opens up the possibility to split the integral equations into multiple smaller sections, or "elements". Each element gets its own test function and the value of $\mathbf{E}$ we want to solve for is approximated over the element by a set of basis functions. The basis functions assume different types of shapes over the element and are therefore also named shape functions. The purpose of the shape functions is to be able to describe the solution over the element and therefore also need to form a complete basis over the geometry. The test functions are chosen depending on the implementation, but a common choice is based on the Galerkin method [79] [73], which is why we sometime see this name when people are referring to FEM.

As an example to illustrate this point, we can in Fig. 4.3 see a 1D example of how an arbitrary function $u(x)$ can be approximated with different basis functions for different element sizes. Here the basis functions are chosen to be the simplest form of piecewise linear. This example also highlights that we then can choose elements so they have a smaller size where the solution changes a lot, while we can have a bigger element size where the solution is more constant, which means that we can keep the DOF down and solve for the system faster.

4.2.2 Boundary conditions

When we simulate electromagnetic fields over a domain, we need to be able to confine the size of the geometry. Electromagnetic interactions extends to infinity so there is a need for us to accurately handle wave propagation when it "leaves" the structure so we do not get unwanted numerical artefacts and can save computation time. In the following subsections, we will present the boundary conditions needed to solve for wave propagation in an optical waveguide. We will treat periodic conditions, perfect magnetic/electric conductor, and scattering boundary conditions. The boundary conditions are discussed from the viewpoint of use in COMSOL [78], but the principle is the same for other implementations.

4.2.2.1 Perfect electric/magnetic conductor - PEC/PML

When we want to simulate interactions with an ideal metal, we can simulate a perfect electrical or magnetic boundary. This is useful when we are interested in capturing ef-

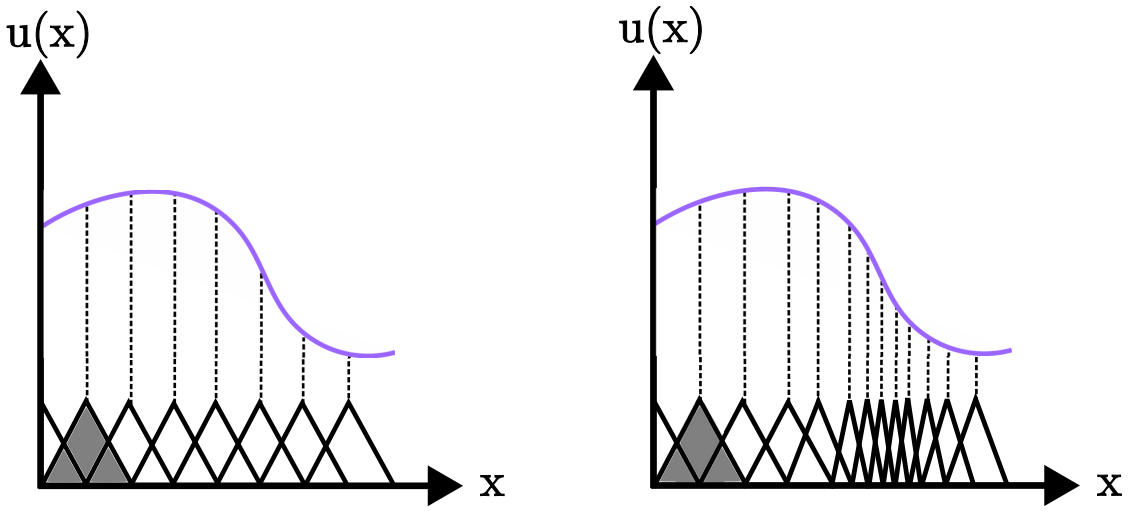


Figure 4.3: Example of how an arbitrary function $u(x)$ is described by means of basis functions over different element partitions. The first basis function is coloured grey. In Fig. 4.3a, the elements have equal size and in Fig. 4.3b the elements are refined in regions where u varies more.

fects of ground planes or metallic effects where we do not need to account for metallic losses.

The boundary condition for a perfect electrical conductor (PEC) simply imposes the boundary condition

$$\mathbf{n} \times \mathbf{E} = 0 \tag{4.26}$$

along the boundary. Here $\mathbf{n}$ is the normal vector to the boundary. Creating a symmetry plane where tangential electric-fields becomes anti-symmetric. In a similar way, the perfect magnetic conductor (PMC) can be used to create a symmetry plane for electric fields by imposing the boundary condition

$$\mathbf{n} \times \mathbf{H} = 0. \tag{4.27}$$

The effect of using different boundary conditions is illustrated in Fig. 4.4.

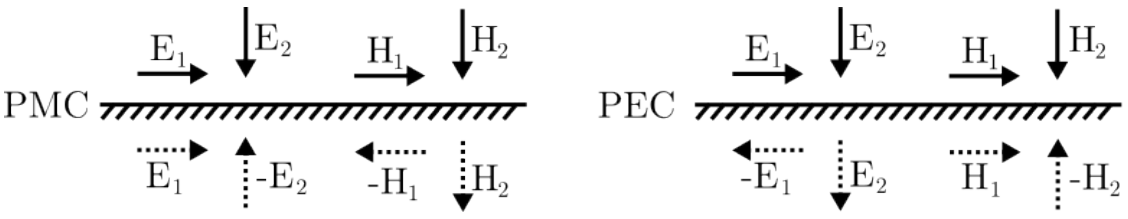


Figure 4.4: Figure depicting the symmetries of fields when using PEC and PMC boundary conditions. Subscripts 1 and 2 denote tangential and normal component of the fields.

This means that if we are interested in simulating modes in a waveguide, we can use these boundaries to create symmetry planes if the device allows for it. So if we want to simulate transverse electric modes in a waveguide that is homogeneous or has a symmetry plane along the longitudinal propagation direction, we only need to simulate half of the geometry if we add a PMC in the middle, as shown in Fig. 4.5. The side effect of this choice is that we then exclude the transverse magnetic modes, since they won't appear due to the added boundary. If we instead were to use PEC, we would calculate the transverse magnetic modes.

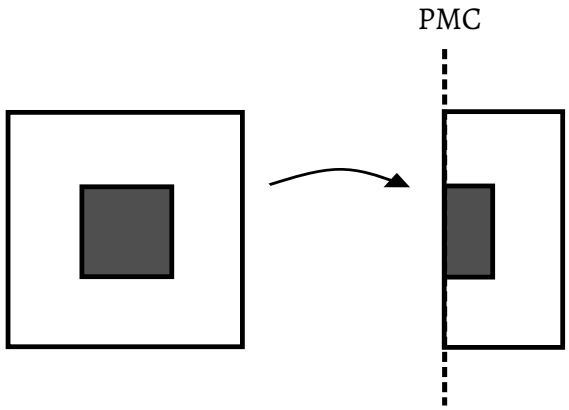


Figure 4.5: Figure showing the use of PMC to create a symmetry plane to cut the simulation domain in half.

4.2.2.2 Periodic Floquet boundary conditions

Floquet boundary conditions are a set of boundary conditions that uses Floquet theory named after Gaston Floquet who studied these systems during the late nineteenth century [80]. It describes systems where we can express a vector field at a destination location $\mathbf{x}$ as a phase shift operation acting on a the vector field from the source location

$$u_{\text{destination}} = u_{\text{source}} e^{-i\mathbf{k}_F \cdot \mathbf{x}}. \tag{4.28}$$

In a normal periodic problem, when we have a unit cell with length

$$L = |\mathbf{r}_{\text{dst}} - \mathbf{r}_{\text{src}}|, \tag{4.29}$$

we would have a boundary condition stating

$$u(L) = u(0). \tag{4.30}$$

If we instead have accrued a change in phase or amplitude, we can express the periodicity like in Equation (4.28) as

$$u(L) = u(0) e^{-i\mathbf{k}_F \cdot (\mathbf{r}_{\text{dst}} - \mathbf{r}_{\text{src}})}. \tag{4.31}$$

This boundary condition is useful when we want to simulate an infinite periodic structure, as described in Sec. 2.3, and we know that the change after the wave has propagated one unit cell results in an added phase shift. This is used and demonstrated in paper III.

4.2.2.3 Scattering boundary conditions

When we have a wave exiting our domain of interest, we want a way to make this wave 'go away' so it does not disturb the rest of the simulation. There are different ways to handle this, such as with perfectly matched layers (PML) [81, 82], which is a well established method to truncate a simulation volume. It does come with a performance penalty and requires an additional region to be meshed to enclose the simulated volume.

Another less costly method is the scattering boundary condition (SBC) of the first-order [78]

$$\mathbf{n} \times (\nabla \mathbf{E}) - ik\mathbf{n} \times (\mathbf{E} \times \mathbf{n}) = 0, \tag{4.32}$$

which is a complex-valued variation of the Robin boundary condition [83, 84]. This can be made more effective by including a second-order term with an added computational cost. The second-order scattering boundary condition adds a tangential derivative term along the boundary which lessens reflection at greater angles of incidence

$$\mathbf{n} \times (\nabla \mathbf{E}) - ik\mathbf{n} \times (\mathbf{E} \times \mathbf{n}) - \frac{1}{2ik_0} \nabla \times (\mathbf{n} \cdot (\nabla \times \mathbf{E})) = 0. \tag{4.33}$$

All of these methods still allow reflections to occur when the incidence angle of the field deviates from the boundary normal. In many cases, the SBC is good enough but it depends on the geometry and scattering properties of the specific problem we try to solve. This boundary condition is used in paper III-IV.

4.2.3 Eigenmode simulations

When we have a periodic Bloch structure, as described in Sec. 2.3 and 4.2.2.2, we often find ourselves in a situation where we want to calculate the dispersion relation with respect to frequency. The conventional way to achieve this numerically, as described in, for instance, Refs. [85, 86], is to solve for the eigenmodes of the system. We fix the wave vector $\mathbf{k}$ and solve for which frequencies the eigenmodes have. Then we perform a parametric sweep of the wave vector to get the full dispersion. In some instances, however, it is more interesting to do this the other way around and fix the frequency instead of $\mathbf{k}$. We might, for example, be interested in a design that can only operate at certain frequencies. Approaches to achieve this in 2D with FEM were first proposed in Refs. [87], but generalised formally in 3D in Ref. [88].

The process to get the wave vector at a specified frequency with finite element eigenmode simulation involves a rewrite of the weak expression presented in Equation (4.24). The way we do this is rather straightforward, we express the electric field as a Bloch wave:

$$\mathbf{E}(\mathbf{x}) = \mathbf{u}(\mathbf{x})e^{i(\omega t - \mathbf{k}\cdot\mathbf{x})}, \tag{4.34}$$

where $u(\mathbf{x})$ is a periodic function. We then insert Equation (4.34) into the wave equation (4.23) and receive a new field equation for $\mathbf{u}$:

$$\frac{k^2}{\mu}\mathbf{u} - \frac{k}{\mu}(\mathbf{k}\cdot\mathbf{u}) - i\mathbf{k}\times\left(\frac{1}{\mu}\nabla\times\mathbf{u}\right) - i\nabla\times\left(\frac{1}{\mu}\mathbf{k}\times\mathbf{u}\right) + \nabla\times\left(\frac{1}{\mu}\nabla\times\mathbf{u}\right) - \varepsilon\frac{\omega^2}{c^2}\mathbf{u} = 0. \tag{4.35}$$

We can then solve for $\mathbf{k}$ like an eigenvalue problem, by writing $\mathbf{k} = \lambda\cdot\hat{\mathbf{r}}$, where $\hat{\mathbf{r}}$ is the normalised direction of periodicity and λ is the eigenvalue we solve for. We need to fix the periodic direction, $\hat{\mathbf{r}}$, in order to decrease the dimension of $\mathbf{k}$ down to 1, so we can solve for it using an eigenvalue solver.

To be able to solve it, we need to use Equation (4.35) to create a new weak formulation for the finite element solver. This is done by multiplying Equation (4.35) with a test function $\mathbf{v}$:

$$\begin{aligned} F_E(\mathbf{v},\mathbf{u}) = {} & \frac{k^2}{\mu}\mathbf{v}\cdot\mathbf{u} - \frac{1}{\mu}(\mathbf{v}\cdot\mathbf{k})(\mathbf{k}\cdot\mathbf{u}) - i\frac{1}{\mu}\mathbf{v}\cdot[\mathbf{k}\times(\nabla\times\mathbf{u})] - i(\nabla\times\mathbf{v})\cdot\frac{1}{\mu}(\mathbf{k}\times\mathbf{u}) \\ & + (\nabla\times\mathbf{v})\cdot\frac{1}{\mu}(\nabla\times\mathbf{u}) - \varepsilon\frac{\omega^2}{c^2}\mathbf{v}\cdot\mathbf{u} = 0. \end{aligned} \tag{4.36}$$

So if we are interested in the propagation constant along the propagation direction, z, along a periodic structure, we use expression (4.36) and set the wave vector to be $\mathbf{k} = \lambda\hat{z}$ and solve the expression for eigenvalue λ.

This method for 3D structures is not readily available in common electromagnetic solvers or packages. However, the 2D equivalent method, where instead of simulating a 3D structure and fix the propagation direction one simulated the 2D cross section of an infinite structure is readily available. Solution for solving for the mode profile is present both in finite differences (FD) [89] and finite element (FE) methods [72] and are generally called mode analysis, boundary mode analysis, or eigenmode solver.

4.2.4 Selecting modes from simulations

When working in nanophotonics with eigenmode solvers like those discussed in Sec.4.2.3, we find ourselves in a situation where we have frequency regions where some modes don't propagate, for example illustrated in Fig. 2.3 in Sec. 2.1. In practice, selecting and

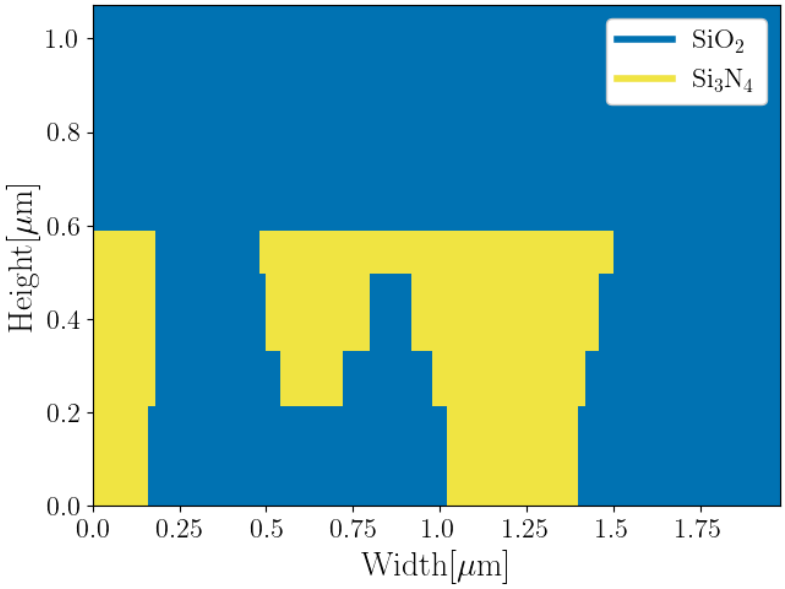


Figure 4.6: Figure depicting the cross section of a simulated waveguide.

discerning different modes based on propagation constants can be tricky. As an example, take the cross section geometry of a waveguide as shown in Fig.4.6.
By performing a sweep over frequency or wavenumber, one finds oneself with a set of eigenvalues as depicted in Fig. 4.7.

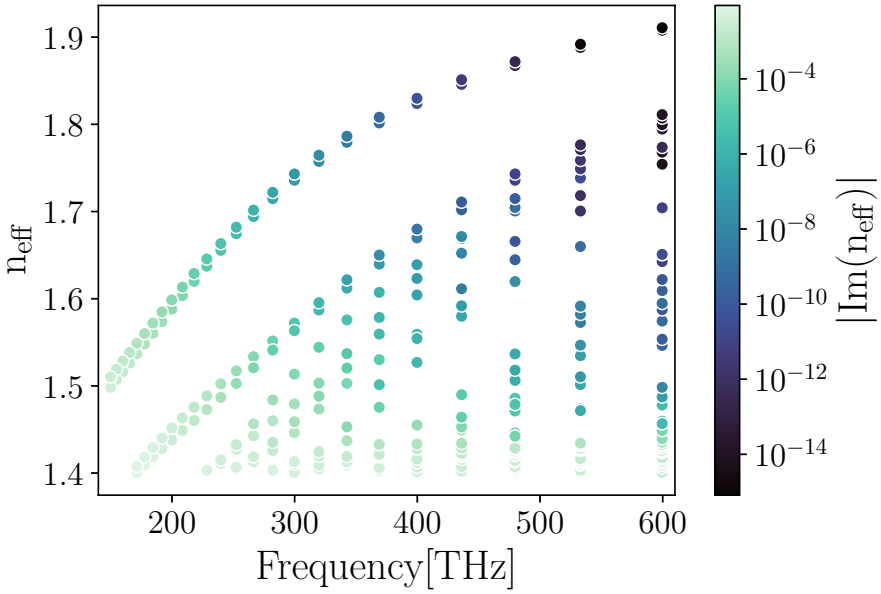


Figure 4.7: Figure depicting the plethora of modes that can be present in a waveguide.

Here we can see that when solving for eigenvalues, we can also get modes that are lossy and therefore not often interesting for the application. A common method to extract relevant modes is to use the fact that since the eigenvalue of the mode is complex, we have one part that determines the loss and the other part that is the propagation constant or, with scaling, effective refractive index, n_{eff}:

$$\beta = n_{\text{eff}} \frac{2\pi}{\lambda}. \tag{4.37}$$

Many times, the effective index is given as a complex value, in which case the imaginary part expresses the attenuation of the system, with the same scaling as in Eq. (4.37).

In order to extract relevant modes, a common way is to simply sort them on loss as employed in for example, Ref. [88]. A disadvantage with this method is that when we have multiple propagating modes or when all are somewhat lossy, sorting on loss can give eigenvalues of different modes for different adjacent frequencies. If we are interested in the same mode over multiple frequencies in order to calculate its derivative, as detailed in Sec. 4.1.1, this can result in discontinuities that negatively impacts the results. If the eigenvalue are all that is available for the problem one has to make do with only these. In many cases, however, one has also access to the mode profile, i.e., the field distribution. This can be utilized to calculate quantities, which can be used to aid in selecting modes. For waveguides, one such metric is the effective area (or effective mode area) (4.14) mentioned in Sec.4.1.1.

Another method that can prove valuable if we are interested in the field concentration of the mode in specific parts of a waveguide is to use the confinement factor[90]:

$$\Gamma = \frac{\int_{core} \mathbf{E}^* \mathbf{E} \, d\mathbf{x}}{\int \mathbf{E}^* \mathbf{E} \, d\mathbf{x}}, \tag{4.38}$$

which gives us a direct measure of the amount of energy that is present in the core of the waveguide. This can also be a good measure when we are interested in modes that concentrate the field to specific regions of the waveguide, then we simply change the region of integration from the core geometry to the desired region. This can, for example, be useful for optimizing a geometry for nonlinear optics where concentration of field to certain materials that have a higher nonlinear interaction is desired.

All of these methods do not guarantee the same mode is selected for different frequency in a scenario like that described in Fig. 4.7. In order to ensure that the same mode is followed over the entire frequency range, the mode profile needs to be tracked over different wavelengths by comparing how close the field distributions are. This can be achieved by calculating the coupling efficiency [91–93]:

$$\eta = \frac{Re\left(\int \mathbf{E}_1^* \mathbf{E}_2 \mathrm{d}\mathbf{x}\right)^2}{\int \mathbf{E}_1^* \mathbf{E}_1 \mathrm{d}\mathbf{x} \int \mathbf{E}_2^* \mathbf{E}_2 \mathrm{d}\mathbf{x}}, \tag{4.39}$$

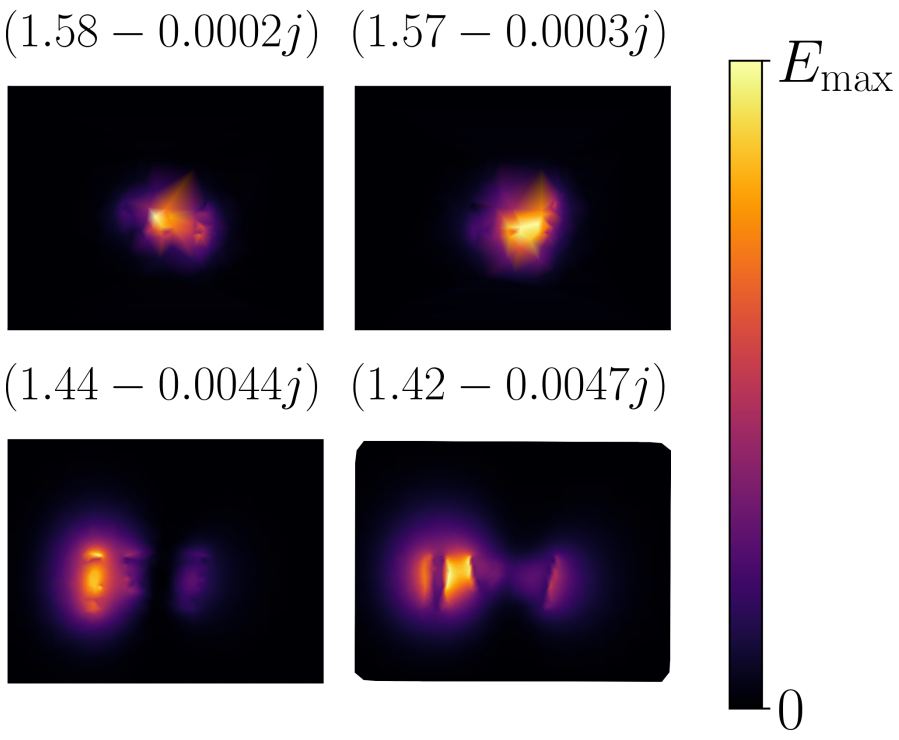


Figure 4.8: Electric field distribution and effective index of the first 4 modes from eigenmode simulations of the geometry in Fig.4.6 at 200 THz.

where $\mathbf{E}_1$ & $\mathbf{E}_2$ are the fields for different modes to compare. If one also have access to the magnetic field $\mathbf{H}$, then the power coupling coefficient [94] can be used instead.

$$\frac{P_1}{P_2} = Re\left\{\frac{\left(\int \mathbf{E}_1^* \mathbf{H}_2 \mathrm{d}\mathbf{x}\right)\left(\int \mathbf{E}_2^* \mathbf{H}_1 \mathrm{d}\mathbf{x}\right)}{\int \mathbf{E}_1^* \mathbf{H}_1 \mathrm{d}\mathbf{x}}\right\} \frac{1}{Re\left(\int \mathbf{E}_2^* \mathbf{H}_2 \mathrm{d}\mathbf{x}\right)} \tag{4.40}$$

But even this can give an undesired result if modes are compared that are far from the effective index of the mode, where the coupling efficiency is sought after. This is illustrated with the example in Fig. 4.8.

Here, the mode profile of all modes at 200 THz is presented and depicts a scenario where modes are distributed close to the adjacent modes and can thus give similar overlap values. This can in turn be mitigated by both filtering away undesired modes based on metrics previously mentioned or by only calculating the coupling efficiency in the near vicinity of the target mode in terms of n_{eff}.

Combining these techniques, we can go from a scenario as depicted in Fig. 4.7 to a

scenario where we have traced all the modes based on field overlap depicted in Fig. 4.9.

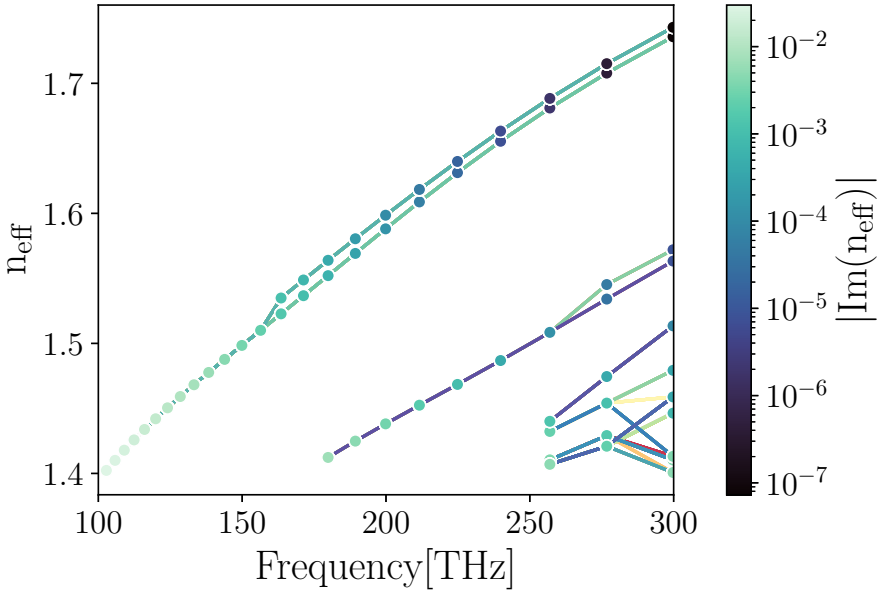
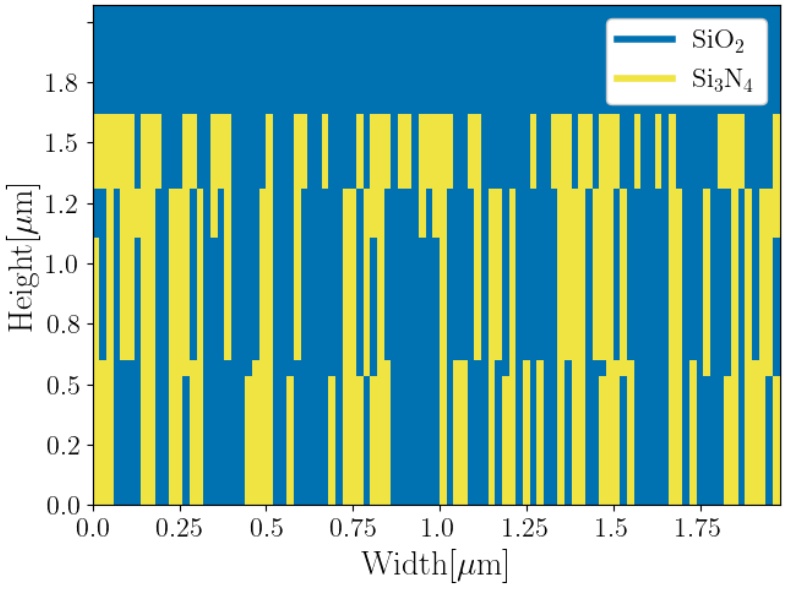


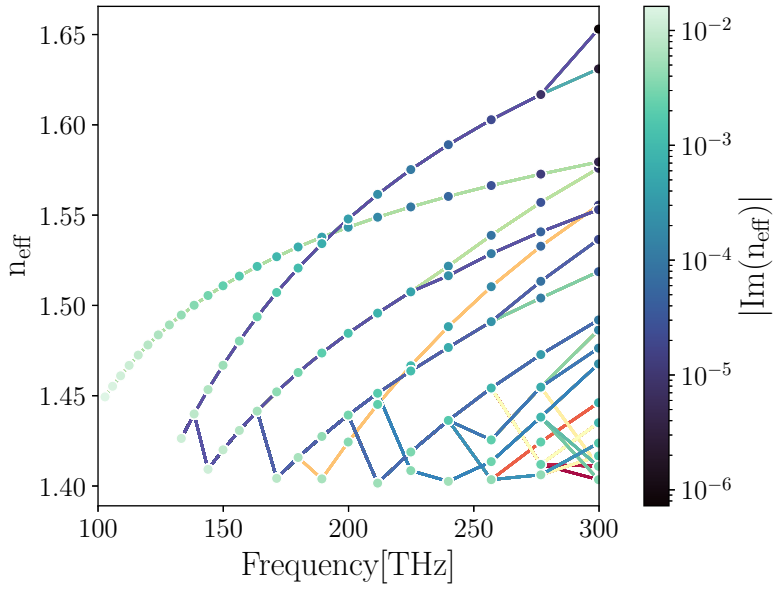
Figure 4.9: Tracking of modes in the geometry of Fig.4.7 using coupling efficiency presented in Eq. (4.39) starting the mode track from 300 THz and going down.

Note that the method of overlap with adjacent modes relies on the fact that there actually exists such a mode. If, for example, there exists a bandgap, we will have modes that are purely imaginary and risk losing the track in the gap. For other problems, where symmetries and reciprocities can be utilized, there exists other methods that can be used to track modes and mode crossings [95–97]. For practical considerations, the use of any field overlap calculations is also dependent on the resolution of the fields and, therefore, on the mesh resolution of the problem. While we can get very accurate values of the effective index, even with a coarse mesh, the resolution of the fields might be less exact. If one is working on a single design, having higher resolution of the mesh might be acceptable, but if it is needed to generate a big dataset, as in the case of work in Paper IV, then one might need to accept a coarser resolution of the fields to keep memory and computational cost down.

Another point for consideration when designing free-form waveguides is that you can get structures that contain "multiple waveguides" or exhibits crossovers between modes. An example of this is seen in Fig. 4.10. For this geometry, we can see a scenario where two fundamental modes are crossing at about 2 THz. The structure is overall rather lossy, which, given the shape of the cross section, might not come as a surprise given the erratic shape of this particular example.



(a)



(b)

Figure 4.10: Tracking of modes in the geometry from Fig. 4.10a using coupling efficiency presented in Eq. (4.39), starting the mode track from 300 THz and going down.

4.2.5 Quasi-normal modes (QNM)

When we are dealing with electromagnetic systems, it is often so that we are interested in a structure's response in terms of a frequency-dependent scattering matrix, $S(\omega)$ [98, 99], which contains the information about the transmission, reflection, and absorption of a system. Here, it can sometimes be advantageous to instead look at the system in terms of active resonances and use the fact that there's a formalized theory that relates the scattering parameters with the resonances in the system. This theory is called *quasinormal modes* (QNMs) [100, 101].

What this means is that we can look at eigenmodes of a lossy system described by the source-free Maxwell's equations (4.41),

$$\begin{aligned} \nabla \times \tilde{\mathbf{E}}_m &= i\tilde{\omega}_m \mu \tilde{\mathbf{H}}_m, \\ \nabla \times \tilde{\mathbf{H}}_m &= -i\tilde{\omega}_m \varepsilon \tilde{\mathbf{E}}_m, \end{aligned} \tag{4.41}$$

for complex mode $\tilde{\omega} = \omega_m + i\gamma_m$ of mode number, m, where we make the assumption that energy leaks away from the resonator, often described as satisfying *outgoing-wave boundary conditions* [102].

If we have a system consisting of N ports that contains P modes, we can describe the scattering matrix as [103, 104]:

$$S(\omega) = e^{i\omega\tau}[C(\omega) + D(i\omega - i\tilde{\Omega})^{-1}M^{-1}D^{\dagger}C(\omega)]e^{i\omega\tau}, \tag{4.42}$$

where $C(\omega)$ is an $N \times N$ matrix representing a slowly varying background, D is an $N \times P$ matrix with columns $\mathbf{d}_m$, $\tilde{\Omega}$ is a diagonal matrix of QNM eigenfrequencies $\tilde{\omega}_m$, and M is a $P \times P$ matrix with elements

$$M_{mp} = \frac{\mathbf{d}_m^{\dagger}\mathbf{d}_p}{i\tilde{\omega}_{\mathrm{r},p} - i\tilde{\omega}_{\mathrm{r},m}^{*}},$$

where $\tilde{\omega}_{\mathrm{r},m} = \omega_m + i\gamma_{\mathrm{r},m}$. The N complex-valued elements of $\mathbf{d}_m$ should be interpreted as the amplitudes of the m:th QNM on the ports. τ is a diagonal $N \times N$ matrix of real numbers, τ_n, that determine the phase accumulated between the scatterer and the ports.

Using this formalism one can think of a scattering spectrum in terms of a set of poles on a complex plane. This formalism was used in Paper II [104] where a neural network was trained to predict the values of the constants in Eq. (4.42).

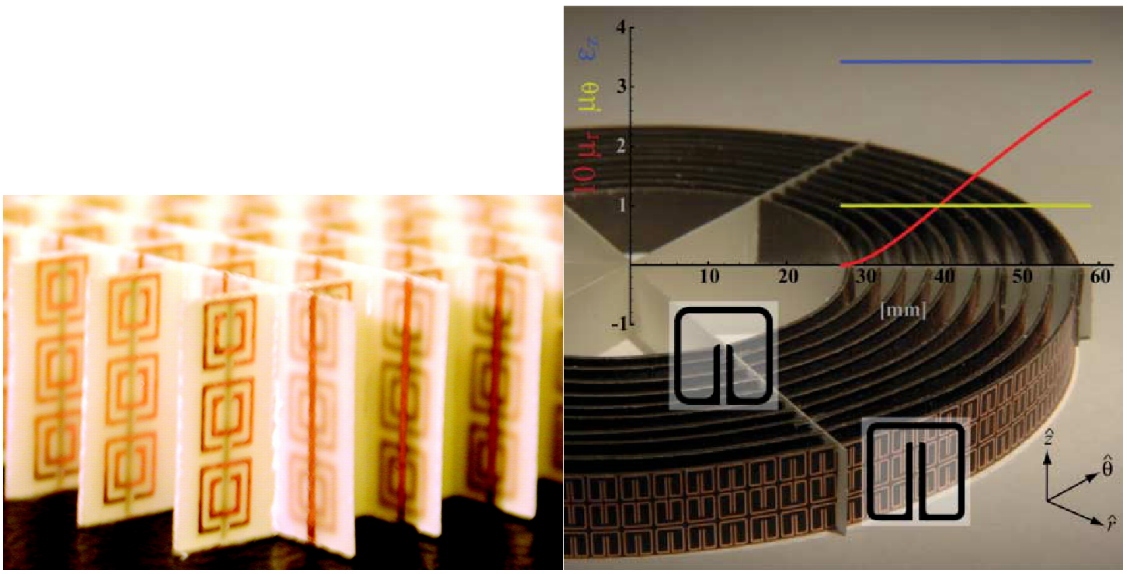
5

Metamaterials

In this chapter, I want to explain what metamaterials are, what the concept builds on and give a sense of the history and the development of the field. The ultimate goal culminating in Sec.5.1 is that we get an understanding about some of the challenges these materials can assist in solving and that we also get a sense of the kind of structures that nowadays can be manufactured to realise these applications. The take away from this, that we will transfer to designs of integrated optics, is the possibilities that subwavelength patterning of structures can present when it comes to controlling light. We will first focus on applications and then in the following chapters we will delve into methods with which to simulate and design these devices, which also is treated in paper II.

You might have come across the term *metamaterials* before and its function is described by the Greek work "meta", meaning beyond. A metamaterial is a material where we purposefully design the material's structure so that it exhibits behaviour different from the bulk material, i.e. achieving performance beyond what the chemical composition of the material can provide. The concept of metamaterials can be applied to many areas of physics such as mechanics [105–107] and acoustics [108, 109], but we will in this thesis focus on metamaterials for electromagnetism. The history of metamaterials begin to some extent in the 19th century, with investigations and measurements on dielectrics by Hertz [110], but only in the middle of the 20th century did structures, such as metallic delay lenses [111], that manipulate the phase velocity of radio waves appear, making it a starting point for development of metamaterials for electromagnetism. Arguably the field really took off round the start of the millennia with the work of Pendry and Smith, when they proposed and realized solutions for achieving negative refraction [112, 113]. This work then resulted in the first invisibility cloak for electromagnetic waves using metamaterials [114], displayed in Fig. 5.1b.

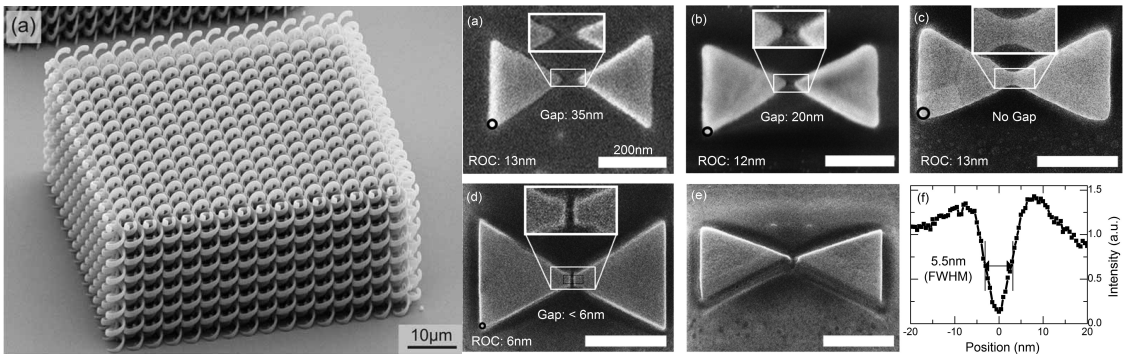
Common for all metamaterials is the need to control geometries on subwavelength scale to be able to efficiently control the wave interactions. Some natural examples do



(a) (b)

Figure 5.1: (a) The experimental verification of a metamaterial with negative index of refraction, from [115]. Reprinted with permission from AAAS. (b) A 2D microwave cloaking structure, from Ref. [114]. Reprinted with permission from AAAS.

exist such as *structural colour*, the phenomenon by which a butterfly has patterning on their wings to create their shade of blue [116, 117], but overall man-made constructions have the potential to outperform natural elements with vast design freedom. This in turn puts two concrete limitations on the field: (1) The resolution we can manufacture a structure with and (2) which structure shapes we predict to work either through theory or numerical calculations. Much of the physics and computational methods needed to predict the behaviour of metamaterial structures have been present for some time and the continued improvements in computational hardware and software have greatly lowered the bar of entry and made it possible to simulate complicated structures with a lower wall-clock time. Although many metamaterials are designed for a specific response, there's also the concept of *tunable* or *reconfigurable* metamaterials, that can alter its response [118, 119]. As to manufacturing, we are now able to manufacture structures on the order of nanometres [120–122], still with increasing accuracy. These numbers can be hard to contextualize in the head, so in Fig. 5.2a we see some of the possibilities we now have of manufacturing, with more examples in Sec. 5.1.



(a) **(b)**

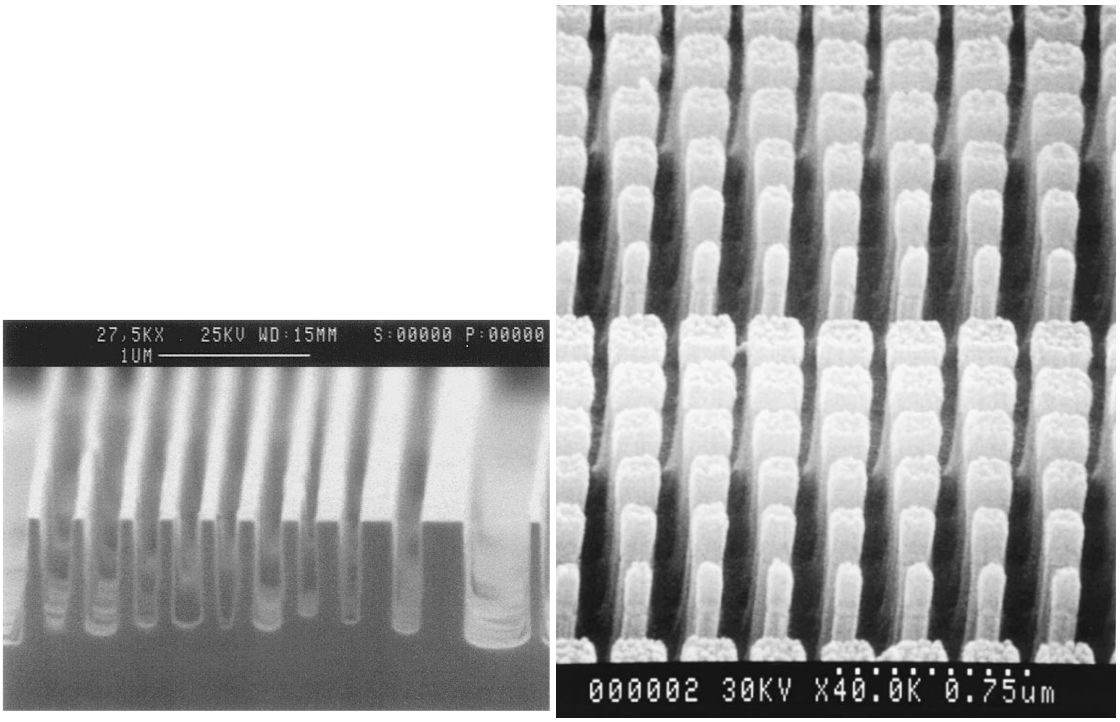
Figure 5.2: (a) An example of a bi-chiral photonic crystal, reprinted with permission from Ref. [123], Copyright 2025 American Chemical Society. (b) illustrate precision of fine details of structures, reprinted with permission from Ref. [124], Copyright 2025 American Chemical Society.

5.1 Metamaterials in nanophotonics

When it comes to metamaterials for photonics, the fact that we are dealing with optical wavelengths results in that we need to manufacture structures that are smaller than micrometres. In the previous section, we touched on some structures from nature at these length scales, but for man-made structures the ability to build sub-micrometre structures only appeared in the late 1980s, following many decades of development, mostly driven by semiconductor development, following Moore's law[125]. First in the 1990s, we could see emergence of optical subwavelength structures [126–128], with example images in Fig. 5.3.

The metamaterials at the time mainly consisted of different combinations of pillars, ridges and holes [129, 130]. Later on, during the 2010s, we started seeing metamaterials emerge that utilized new kinds of shapes as building blocks, such as angled v-shaped rod [131] shown in Fig. 5.4, where Nanfang et al. managed to create abrupt phase changes on the scale of the wavelength to create a desired reflection and refraction. From here, metamaterials with a more freeform design started to appear, such as the work of Piggot and Vučković et al. [132–134], where gradient-based methods were used to generate 2D structures, shown in Fig. 5.5. Here, a demultiplexer has been designed, which redirects incoming light with different wavelengths to separate ports.

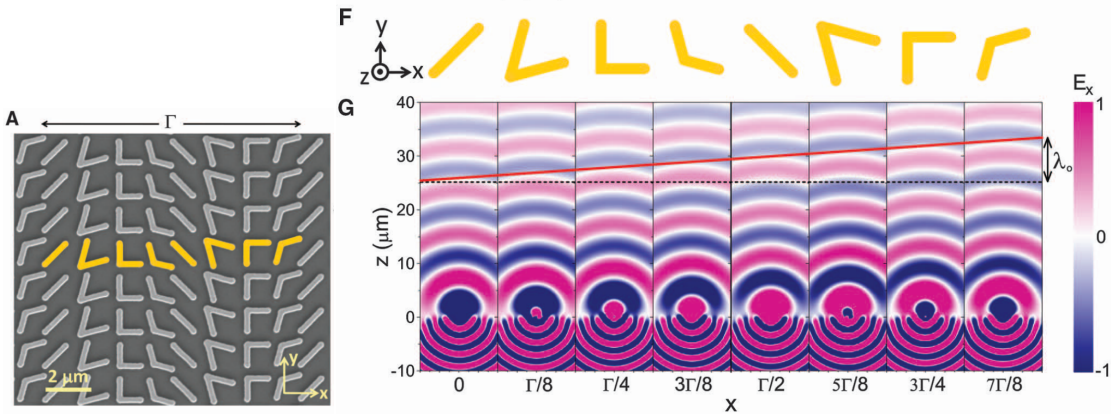
Following these development, we have seen larger metagratings, metalenses, and structures [130, 135–138] and a continuous development of new areas of applications and further free form designs, bringing new advantages and solutions to problems previously unsolved, and with increasingly smaller feature sizes [139–151]. Recently we have also seen metamaterials being used for analogue computing [152, 153], where we can perform matrix multiplications with light directly on chip from multiple inputs [154, 155],



(a) (b)

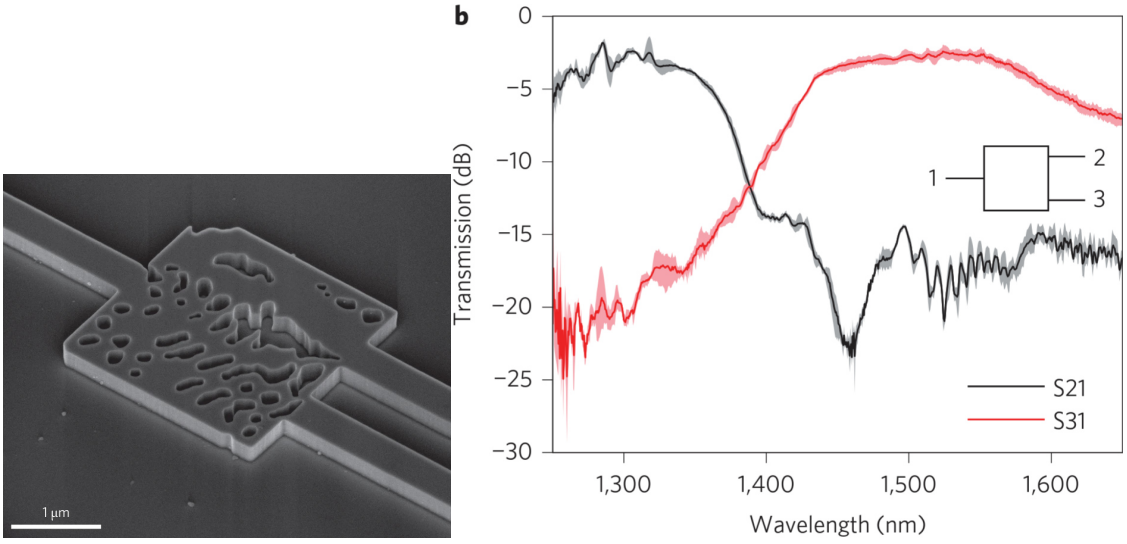
Figure 5.3: (a) A scanning electron microscope (SEM) micrograph of a cross section of a subwavelength transmission grating from Ref. [126], ©2025 Optica Publishing Group. (b) SEM image of a subwavelength grating, from Ref. [128] ©2025 Optica Publishing Group.

as illustrated in Fig. 5.6.



(a) (b)

Figure 5.4: (a) a SEM micrograph of an array of V-shaped gold rods with varying angle, that each introduces a specific phase shift. Combining multiple of these shapes into an array enables the control of the angle of a wavefront as illustrated in (b). Here we see a formed wavefront based on FDTD simulations, from Ref. [131]. Reprinted with permission from AAAS.



(a) (b)

Figure 5.5: (a) a SEM micrograph of a demultiplexer. Light incoming from left (port 1) is directed to either port 2 or 3, shown in (b). Reprinted with permission from Ref. [133]. Copyright © 2015, Springer Nature Limited.

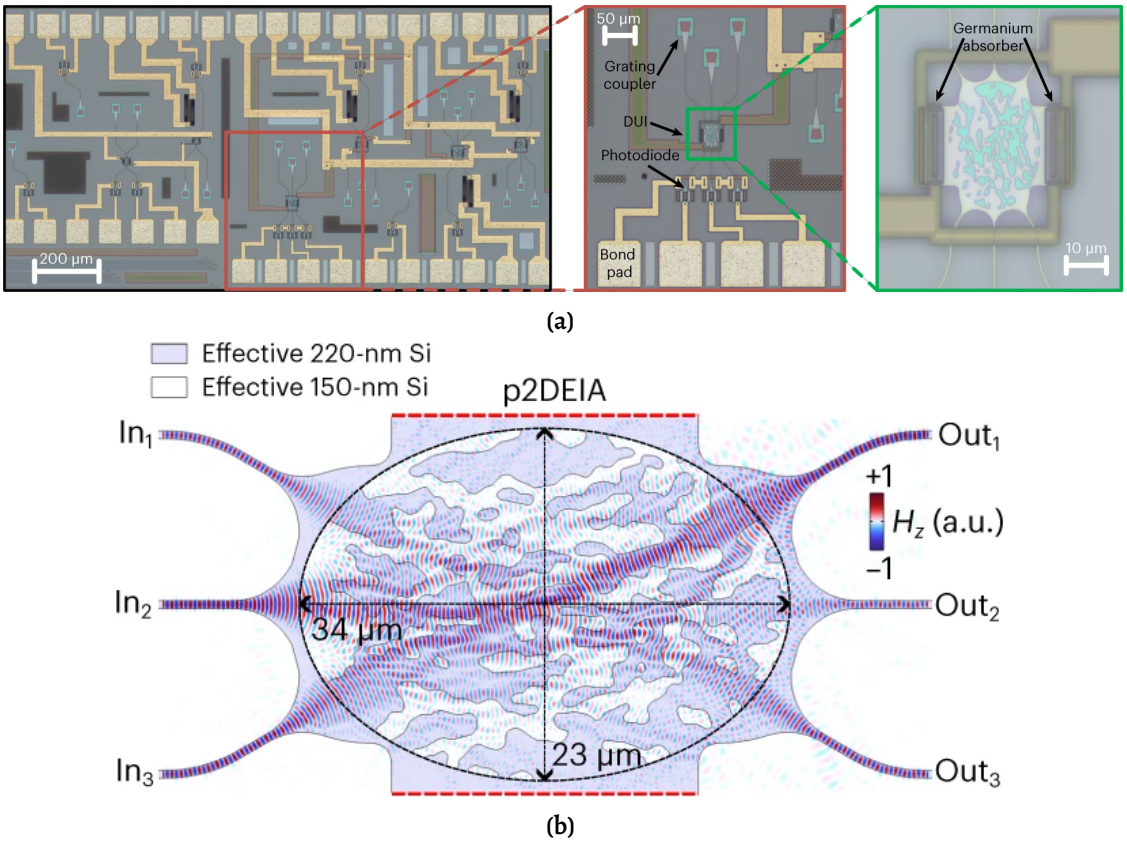


Figure 5.6: (a) A manufactured chip that perform analogue vector-matrix multiplication of 3 input fields. (b) shows a time snapshot of the simulated magnetic field distributions. Reprinted with permission from Ref. [154]. Copyright © 2024, Springer Nature Limited.

6

Machine learning for nanophotonics

In previous chapters, we have delved into fundamental theory and how we use physics and numericas methods to simulate and predict how a material or geometry behaves. This is also referred to as "solving the *forward* problem". That is, we have a very clear physics-based mapping that, given the input like material parameters or geometry, deterministically predicts the behaviour of the system. We will in this chapter treat how we can utilize machine learning to solve these problems and later on in Ch. 7 we will explore how we can use machine learning to solve for the more complicated inverse problem, where we want to predict which input to the forward problem will give us the desired outcome.

Machine learning is nowadays becoming integrated into basic education and the basics of how a neural network works in terms of back-propagation, activation functions and perceptions etc., are assumed known to the reader. I will here instead focus on common practices, methods, and workflows with relevance to nanophotonics and inverse design in particular. For readers interested in introductory materials to machine learning, I would suggest reading *"Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python"* by Sebastian Raschka [156], or for a more statistics focused approach to read *"An Introduction to Statistical Learning"* by Gareth James [157]. For the basic practicalities of working with and developing machine learning code, a very good source is the tutorials included with larger machine learning packages. The most accessible of these are PyTorch [158–160], TensorFlow [161, 162], scikit-learn [163] and Keras [164]. I have mainly used PyTorch during this work, but the difference between using different frameworks is more a matter of syntax or if one has the possibility to piggy-back on someone else's previous work. If you are currently about to start on a new project with no known biases, I would recommend to use PyTorch in this year.

6.1 Why use machine learning for the forward problem?

Now you might be wondering: "Why should I need machine learning to solve a deterministic forward problem?". We often find ourselves in situations where we have some method with which we can correctly simulate our system. Methods like simulating the transmission spectra for some material composition perhaps or predicting the effect of changing the thickness of a dielectric slab. For these cases, there often exist analytical solutions or the possibility to perform a full-wave simulation using methods like finite-difference time-domain (FDTD) or FEM. The advantage in this scenario is that of inference time [165]. Simulation of physical system is often time consuming and more often than not the development and characterization of a system will benefit from having a high simulation speed, so it is easier to iterate on ideas early on in the design process. Neural networks (NNs) can bring considerable improvement in this regard in the form of so called *surrogate* models that have been trained to solve the forward problem. During inference, these NNs are typically very fast, allowing for fast iteration. The downside is that they take time to train in the first place, taking time from our development budget. This means we need to judge if it is worth to train a surrogate model for our specific problem. The more simulations we expect to use during the process, the easier it is to motivate utilizing a surrogate model. Even though we can train networks that have a very high accuracy, we make one sacrifice by using NN, and that is certainty. The black box nature of NN introduces an uncertainty depending on the complexity of the project. In practice, this is seldom a big issue in most projects during the development, since we always have the option to validate design candidates with our more reliable but slower method at the end.

Another case where machine learning is very suitable when solving the forward problem is when we have no accurate model to describe our system. We might for example have a dataset of measurement data and are in need of a surrogate model to describe our system, but we don't have any physics model or correct parametrization of the input/output space due to difficulties in measurements. Here machine learning provides a possibility to describe and predict something that otherwise could not be obtained.

Even though the NN usually is applied like a black box solution as mentioned earlier in this section, there are ways to decrease the "black box level" of the model by incorporating knowledge of physics, which we will treat in Sec. 6.2.

6.2 Incorporating physics and exploiting symmetry in neural networks

When training NN on physics problems, the straightforward way is to generate a dataset connecting an intended input space to the desired output and then start training some model. This approach usually works but given that we already know something about the physics of the problem we can then take steps to incorporate this knowledge directly into the machine learning problem. In order to achieve this we need some form of physics model that at least explains part of the system we are working on. By then letting the network learn the input or output of this model we can decrease the complexity of the problem that the NN needs to learn and thereby also cutting computation time and cost. These types of *physics-informed networks* have been demonstrated in a variety of NNs [104, 166–170] and are also incorporated in paper II, where quasinormal mode (QNM) theory, as described in Sec.4.2.5, was used for the simulation of photonic crystal (PhC) slabs and dielectric free-form metasurfaces [104].

Apart from using specific physics models, one can also leverage symmetries in data to great effect. The typical example of this is when we are working with image based problems, where the input space has the form of a 2D image. In many cases, it is not important where in this image a feature resides as long as it exists somewhere in the space. One reason that convolutional neural networks (CNNs) are efficient at these image-based tasks is that they by their nature are translationally equivariant in vanilla form without stride or pooling [171–174]. The sweeping kernel acts the same on the pixels no matter where they are. So one way to ensure that network architectures are translation invariant is to base them on convolution blocks. These kinds of symmetries in input data are usually exploited in machine learning to enlarge the dataset by means of *data augmentation* [175–177].

An alternative to using symmetries to enlarge the dataset is to encode the symmetries directly into the network, given that one has knowledge of these symmetries, and employ what is called "equivariant networks" as building blocks for a machine learning model. These networks make it possible to specify a group action that the feature space in each layer will respect. If the reader is interested in incorporating equivariance into machine learning models, then some good references can be found in the book by Maurice Weiler, Patrick Forré, Erik Verlinde, Max Welling [178] and in Refs. [179–181]. There currently exists an available python packages called escnn [182].

6.3 Open datasets

The availability of data is one of the limiting factors for the field in terms of training machine learning algorithms. While other more traditional and mature fields in machine learning, such as image classification, has an abundance of datasets to pick from to start

developing and benchmarking models, the entry into the field of physics usually start with the task of creating and curating a dataset.

The datasets usually have an unstandardised formatting in comparison to images, reflecting the variations in parametrization and target parameters that are being investigated. Many datasets are usually available upon "reasonable requests" and the accompanied code and documentation is usually not the topic of prioritization. Nevertheless, there exist some open and readily accessible dataset that can help lower the bar of entry into the field if one wants to check methods against other datasets or lacks the computational resources to generate large enough datasets within reasonable wall-clock time.

Attempts have been made to standardise data sharing in photonics [183, 184] and general physics [183], but have so far gained little traction. A search for readily available datasets resulted in Refs. [184–192]. We also plan to make the dataset used in paper IV available.

7

Inverse design in nanophotonics

In the previous chapter we went through uses of NN in nanophotonics when we want to solve the forward problem and how one can use surrogate models to speed up calculations. The forward problem is typically a well posed problem as first defined by Hadamard [193] i.e., a problem that fulfils the following criteria:

1. The problem has a solution;
2. The solution is unique;
3. The solution's behaviour changes continuously with the initial conditions.

When we have a forward problem we know that everything is well defined and understood and we know that if we solve this problem the solution is uniquely what we want. The problem is that this is more of a verification of the current state that we have. A very common problem is rather that we want a system to exhibit a certain response. We want to focus light in a certain way, interact with different wavelengths differently, and achieve a certain design goal. So we really want to solve for it in reverse and solve the *inverse* problem. This is illustrated in Fig. 7.1. Here we have a nanophotonic example, we want to pattern a dielectric surface such that we have a geometry consisting of two different materials. Given such a pattern/geometry, the physics of the problem is well known from Maxwell's equations and we can numerically solve for the electric field, $\mathbf{E}$, propagating in the structure. We know from physics that we will have a solution, since it will produce some field response. And we know from physics that we always will get the same field response for this time-invariant structure. If we want to do the inverse on the other hand, even if we know the physics of the system, we quickly see that this can be an ill-posed problem. Imagine, for example, that we are interested in the transmission coefficient of this system. We most likely can construct many different geometries that

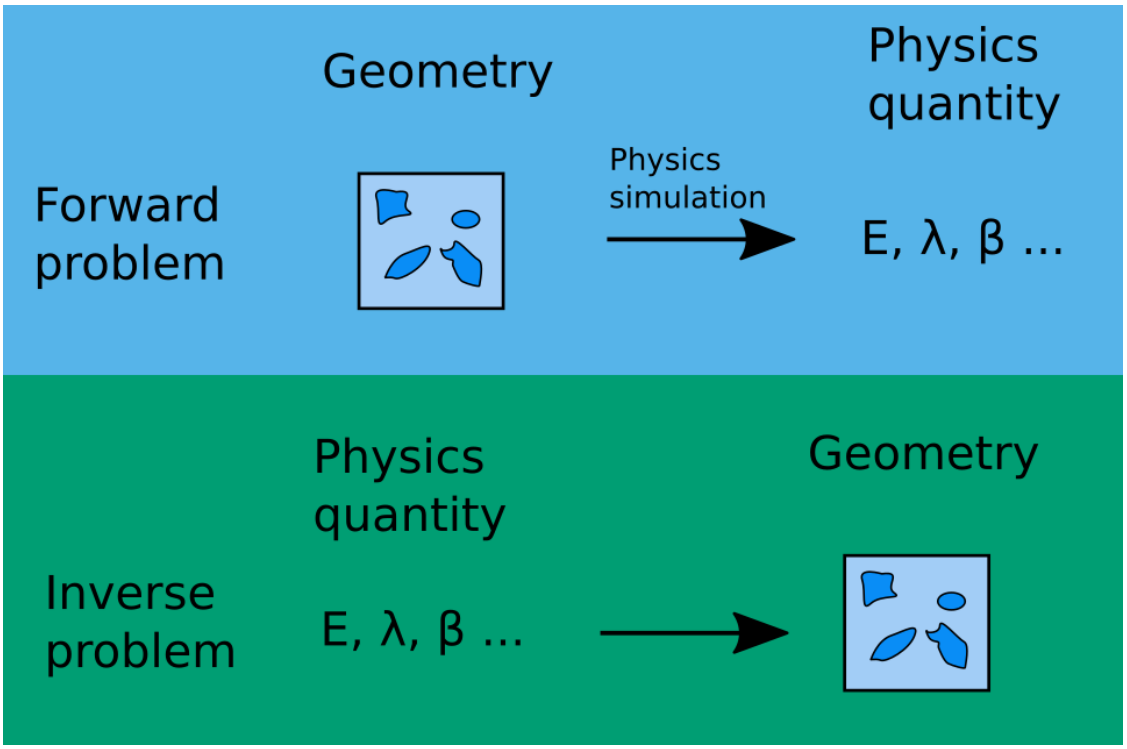


Figure 7.1: Forward vs the inverse design problem.

produce the same transmission, making this a problem without a unique solution. Say we choose a not uncommon way of parametrizing the geometry as a binary image where 1 & 0 denotes different materials, and have a resolution of 256x256 pixels. We then have a system where a space of $2^{(256x256)}$ possible combinations determines 1 value, but in this input space we most likely have many combinations that yield the same output value. If we flip direction to the case of inverse design then using this one value, we would like to recreate as many of these different designs as possible, which is not trivial. Yet, this is still a problem we can solve using a variety of different methods. The field of inverse design in nanophotonics is growing and some great inspirations for inverse design in the field can be found in Refs. [194–201].

Here it might be worth mentioning that there is a difference between finding *one* solution and finding the best solution. In many applications, one solution is good enough, we might design a system with a minimum requirement of some sort and it is *good enough* if it satisfies this requirement. The quest to find the global optimum of the problem is an active research field and currently there is no guarantee of finding the global optimum but there are certain steps that one can take to increase the likelihood of finding the global optimum. It turns out that machine learning is a very good tool to use for this purpose.

When it comes to methods for solving the inverse problem, there are typically two

types of thoughts towards solving the problem. One is to make refinements to one chosen starting design by calculating a gradient or error based on some figure of merit and use this to make alterations to the current design. This process is then iterated multiple times to reach a target. Here we see methods like *topology* and *shape optimization* [202–205]. The other approach, and the focus of this thesis is machine learning based methods, where training data used to construct models that generate the desired designs are generated. Another machine learning based method that is emerging is the use of *agentic systems* which will be further discussed in Sec. 7.2. There is also a class of methods that have been used with success that we will not treat in this thesis and that is what is collectively called *metaheuristic* methods [206–210], that among others, contain stochastic optimizations techniques such as particle swarm optimization (PSO), genetic algorithm (GA), ant colony optimizations (ACO) and more.

Before we further delve into the subject of different machine-learning methods, it is worth discussing the different use cases and when each approach is useful. When it comes to topology optimization, it is by its nature iterative where one repeats a simulation of the structure for each step in the optimization. The efficiency of this method comes down to use the case and how fast one can simulate one iteration, but it has an advantage in the fact that it is easier to understand what the process is doing and how it reaches the final design, while machine learning methods have a black box element to it that obfuscates the inner workings of the method. If multiple designs are sought, then the topology optimization process needs to be repeated for each new design.

When looking at machine learning methods, then multiple simulations need to be executed to build the dataset, a process that costs a lot of computing power. The advantage to the construction of the training dataset is that each sample is independent of the other, meaning that the computation can be made in parallel, given that we have access to sufficient computational resources. Then the process of training the machine learning model also takes time and computing resources. When measuring the total computational cost of both methods, then the machine learning approach is typically more taxing, under the assumption that we are optimizing for few samples. Once the machine learning model is trained, its use is typically very fast and cheap, making it more attractive when we want to generate multiple variations of designs or vary our design targets such as construction of nanophotonic structures with different characteristics such as dispersions as discussed in Ch. 4. When it comes to the work of setting it up and tuning the process, my own experience is that the amount of work needed to spend on either of the approaches is similar.

Machine learning methods employing deeper networks for inverse design, are typically based on one of the following concepts:

1. Iterative inverse methods;
2. Direct inverse method – for example Tandem networks [211] and variational autoencoders (VAE) [212];

3. Generative adversarial networks (GAN) [213, 214];

4. Diffusion models [215].

We will now review the concept of these methods.

For iterative inverse methods, one uses machine learning to train a surrogate model. One then leverages the surrogate model's fast inference time to iterate over the design and evaluate it fast with the surrogate model to reach the desired design target. The success of this approach depends on the speed and accuracy of the surrogate model and how one evaluates changes to the design between iterations. The setup of the surrogate model and its training benefits from lower complexity compared to other machine-learning methods.

The direct inverse methods, as a concept, revolves around training a network structure that once trained can be used directly to retrieve the inverse of the problem. The training of these systems is more complicated than that of a surrogate model but offers the advantage of presenting the inverse solution directly.

Generative adversarial networks (GANs) are a technique that has the potential to generate a higher diversity in the designs [185] and has the potential to generate high-quality designs. GAN training is a type of indirect training that consists of training two models together, one generator that generates designs, and one discriminator that judges how realistic/correct the generated designs are. These two models then compete against each other during training to perform the best. The downside of GANs is that they often are complicated to train and can suffer from mode collapse [216].

Diffusion models have rapidly gained popularity in many fields, largely driven by their success in generating images and video. In this year of 2026 we have seen a huge surge in AI generated content in the form of images and videos with models such as Stable Diffusion [217] and Sora [218] with different derivatives and competitors popping up. If you have seen artificial intelligence (AI) generated visual content online, then it most likely is based on a diffusion model. This interest also means that a lot of work is currently put into its development. The models themselves are trained to denoise a given input and after training it is used to generate designs by replacing the blurred input with random noise.

In the work presented in this thesis, the focus has been on direct inverse methods, specifically conditional variational autoencoder (cVAE), which was used in paper IV. These concepts will be further treated in Sec.7.1.

7.1 Conditional variational autoencoders

In this section, we will explain the concept of a conditional variational autoencoder (cVAE) and the purpose of its design. In order to do it, we will start with the concept of an au-

toencoder (AE) and then, building on this, move on to variational autoencoders (VAEs), and finally to cVAEs.

7.1.1 Autoencoders

The concept of an autoencoder (AE) has its roots in the 1980s [219–225] and is illustrated in Fig. 7.2. The idea is that we want to find a latent space representation that accurately describes our data in a lower dimension than the original. We have an NN that acts as an encoder expressing the information in the image to a few values illustrated by the block in the middle. We then have a decoder NN that, given this latent space representation, reconstructs the image. The use case for these systems is primarily when we want to perform dimensionality reduction [226] in the same way as we might use principal component analysis [227].

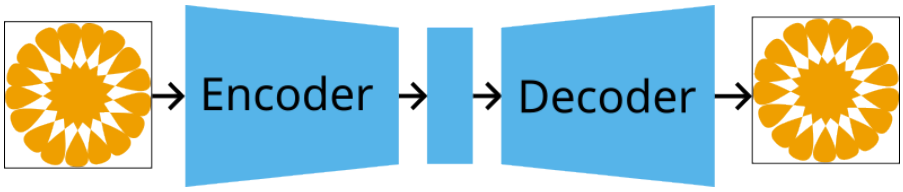


Figure 7.2: Autoencoder structure with a compressed latent representation in the middle. Image to the left is compressed into a latent space and then reconstructed to full size on the right.

7.1.2 Variational autoencoder

An inherent problem with the autoencoder (AE) described in Sec. 7.1.1 is the fact that we have no real control over the latent space that is expressed by the encoder. This makes it hard to sample the latent space and generate new data using the decoder. This is remedied in the variational autoencoder (VAE) [212]. Here, the structure is set up in such a way that the latent space the network learns takes the form of a probability distribution, of which the most common choice is a normal distribution, which we will focus on here. The structure of the variational autoencoder (VAE) is illustrated in Fig. 7.3.

In order to enforce this behaviour, some things need to be adapted from the case of an AE. First we need to add a term to the loss function that ensures that the distribution learned correctly reflects the data. This is achieved by the use of the Kullback Leibler divergence (KL divergence) [228] also known as *information gain* or *relative entropy* [229]:

$$KL\left(p|q\right) = \int p(x) \log\left(\frac{q(x)}{p(x)}\right) dx, \tag{7.1}$$

which is a statistical measure that quantifies how much distribution q differs from distribution p. For illustration purpose the quantity is demonstrated in Fig. 7.4. This has

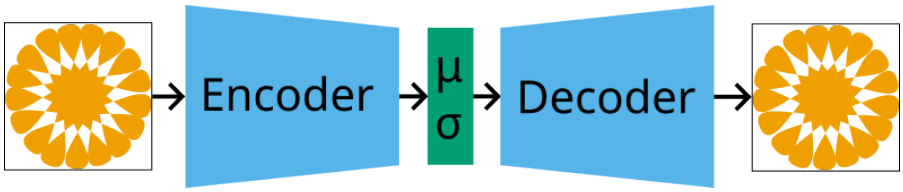


Figure 7.3: VAE structure. Like in Fig. 7.2, the network learns to deconstruct the input to the left. In a VAE this is achieved by learning a distribution here illustrated by a normal distribution with mean, μ, and standard deviation, σ.

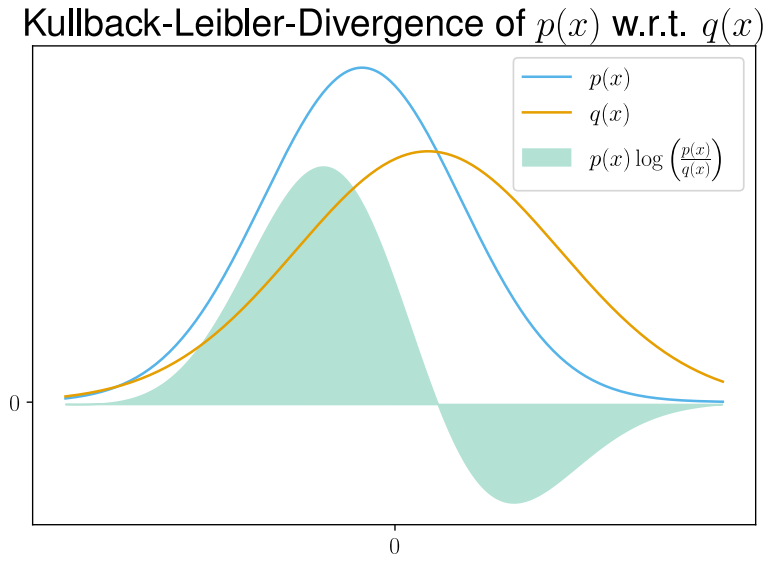


Figure 7.4: Figure illustrating the Kullback Leibler divergence of two distributions as a function of x.

among its properties that it is non-negative, invariant under reparametrizations, and *asymmetric*, meaning that $KL(p|q) \neq KL(q|p)$.

In order to incorporate this into the loss function one makes use of the *variational lower bound* (or *evidence lower bound*, ELBO) and the backward KL divergence to express the loss function as

$$\mathscr{L} = \mathbb{E}\left[\log\left(p(x|z)\right)\right] - KL\left(q(z|x)\|p(z)\right). \tag{7.2}$$

Here the $\mathbb{E}$ is interpreted as the encoder block operating on the value. This can further be extended by adding a weight to the KL part resulting in a β-VAE [230, 231]:

$$\mathscr{L} = \mathscr{L}_{\text{reconstruction}} - \beta\mathscr{L}_{KL}. \tag{7.3}$$

We thus get a loss function that contains a term for the reconstruction loss of the image

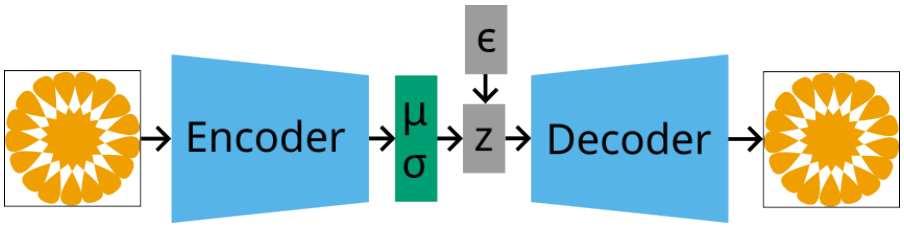


Figure 7.5: Reparametrisation trick in VAE.

(how close the reconstruction is to the original) and one term containing the KL divergence.

It is not sufficient to change the loss function to get the VAE to work, just using the values predicted by the encoder would just result in a normal AE. In order for the system to properly learn a distribution, we need to use the values from the decoder to create a distribution that we can sample from. In order to achieve this, we employ a *reparametrization trick* using stochastic backpropagation [232], illustrated in Fig. 7.5. We are using the values outputted by the encoder as parameters for a distribution from which we draw a new random value, z. If we want the VAE to learn a normal distribution, as illustrated in Fig. 7.5, then we draw this new value using the predicted mean and deviation of the distribution as:

$$z = \mu + \sigma\epsilon;\ \epsilon \sim \mathcal{N}(0, 1), \tag{7.4}$$

where ϵ is a random variable drawn from a normal distribution at the inference step.

This entire structure is then trained end-to-end, and after training one can generate new designs using the distribution by only using the decoder part and sample z as depicted in Fig. 7.6. An example of its use in nanophotonics can be seen in Ref. [233].

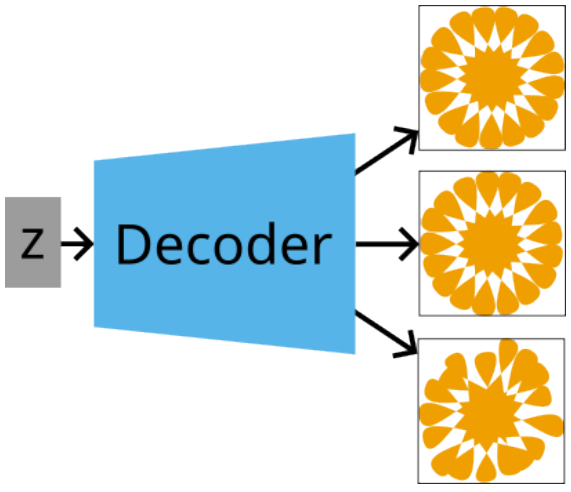


Figure 7.6: VAE inference structure, different designs are generated by varying z.

7.1.3 Conditional variational autoencoder

Even though VAEs described in Sec. 7.1.2 are good for the purpose of generating designs from a specified distribution in the field of physics, we are more often interested in specifying targets more precise than simply drawing from a distribution. This can be achieved by extending the VAE by giving it a conditional expression, which leads to a conditional variational autoencoder (cVAE) [234]. The analogy between a VAE and a cVAE is that, given a dataset of many cats of different labelled breeds as classes we can train a VAE on this system. By then sampling the latent space, z, we can generate many cats, but if we want to generate a "Selkirk Rex" or a "Norwegian Forest" we cannot give this input to the VAE. In some scenarios, the system can encode certain information such that specific parts of the distribution contain different classes, but their location in the latent space is not given in the training. In this case the latent space needs to be investigated in the hope of finding a region that we can sample from to obtain the right result. The concept of cVAE is that we provide the targets as an additional input to both the encoder and the decoder as depicted in Fig. 7.7.

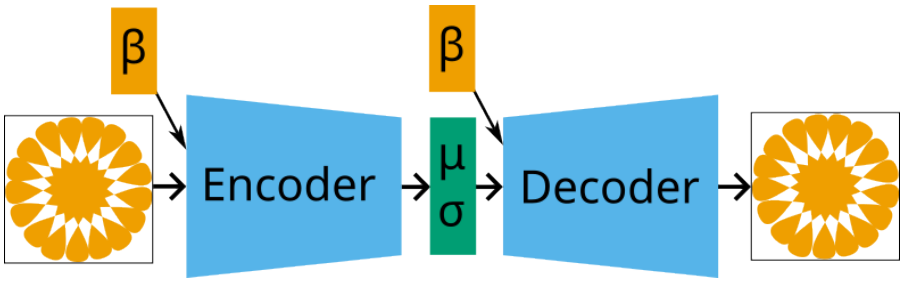


Figure 7.7: CVAE structure, β is the conditional target/label, μ, σ are parameters of the normal distribution learned.

Apart from the inclusion of this data, the training and construction of the cVAE follows the same concept as VAE described in Sec.7.1.2. After training it end-to-end, we similarly discard the encoder part and generate designs, as depicted in Fig. 7.8, using the target β, and vary the results from the distribution by changing the distribution variable z.

For some example of where a cVAE has been used in the field of physics, see Refs. [185, 235–239] as well as its use in paper IV.

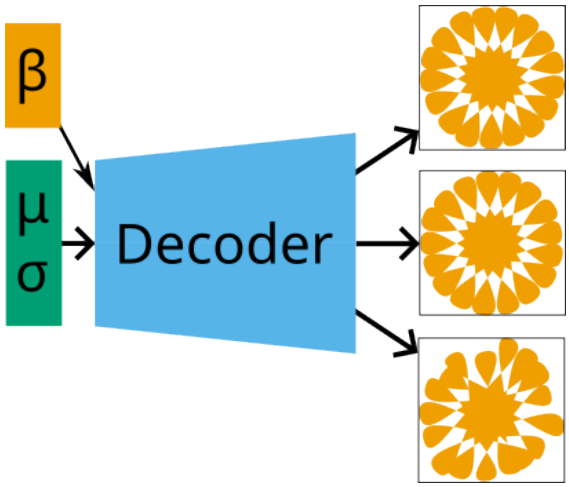


Figure 7.8: CVAE structure that creates new designs.

7.2 Agentic systems

An emerging topic in the world of machine learning is the concept of *agentic systems*. This is a development that has been made possible by the recent advances in the development of large language models (LLMs), arguably popularised by the advent of the first generative pre-trained transformer (GPT) [240] in 2018, utilised by the company OpenAI in its chatbot ChatGPT. This advance was made possible by the development of the transformer network the year before [241], which then took over from the previously dominant long short-term memory (LSTM) models [242] for text and language tasks, much due to their shorter training times.

The availability of these LLMs and their growing performance has opened up an area of automatisation where LLMs are used is to create an autonomous system that performs tasks, called *agentic AI*. The AI is given the target function and tools to simulate/evaluate the designs and then the AI agent iterates until it finds a solution that fulfils the target function. In the field of metamaterials, there now exists work that utilizes these agentic AIs, such as ChatGPT, to solve for problems [243–245]. These systems rely on LLMs for higher-level abstraction of the entire workflow to simulate the designs and interfaces with classical solvers, specialized surrogate models, and also other specialized LLMs.

The concept of creating agents that iterate on a problem is in principle nothing new. In machine learning, there is the concept of *active learning* or *query learning* [246]. Commonly using a Bayesian framework, one forms an iterative process that predicts how to generate new training data to improve the model as optimally as possible. What sets agentic systems apart is the potential for higher-level abstraction that LLMs can provide. In both cases, however, one must take care in formulating the design goals and how to evaluate them. LLMs are fast during inference, however training LLMs is costly,

and readily available LLMs often have the problem of being too general for a problem. One can then refine existing LLMs to be more specialized and for scenarios where the generality isn't needed one can employ what is called small Language Models (SLMs) [247].

8

Summary of papers

Paper I

Adaptive meshing strategies for nanophotonics using a posteriori error estimation

In Paper I, we present a meshing strategy for simulating nanophotonic systems using a posteriori error estimation in the finite element method. As we can recall from the discussion in Ch. 4 & 7, an overall challenge when working with electromagnetism is the computational cost and time associated with simulations. In this paper we introduce a method to utilise finite element method (FEM) more efficiently.

This is achieved by first simulating the electric field on an initial coarse mesh. We can then use this field to evaluate the a posteriori errors for each element in a finite element mesh. Using this information, we can then determine which of these elements exhibit the largest errors. By having a scheme where we iteratively only refine a fraction of the total elements and only choosing the elements that show the largest errors we provide a h-FEM method that reliably and predictably converges to a result while keeping the degrees of freedom (DOF) down as much as possible.

We implement this strategy in COMSOL Multiphysics for arbitrary 3D geometries and benchmark the strategy against other conventional adaptive meshing approaches. We demonstrate that the use of adaptive meshing leads to faster convergence with lower memory footprint for complex three-dimensional nanophotonic structures. Furthermore we discuss an issue with mesh refinement where propagation of mesh refinement can result in poor convergence for situations where the starting mesh is too coarse.

Paper II

General framework for knowledge integration in machine learning for electromagnetic scattering using quasinormal modes

In paper II, we present QNM-Net, a universal physics-informed neural network for electromagnetic scattering based on the quasinormal mode (QNM) expansion of the scattering matrix. In Ch. 6, we elaborated on how we could remove some of the "black box" aspects of a neural network (NN) by incorporating physics models into the system. This approach has multiple advantages, one being that it increases insight into what the NN is learning and the other is that it simplifies the relation that the network needs to learn from data, thereby reducing the amount needed for training data.

We show that incorporating physics knowledge into the network significantly improves data efficiency for training models on photonic crystal (PhC) slabs and all-dielectric free-form metasurfaces. Our implementation works by, instead of allowing a neural network to predict scattering parameters directly from a given device geometry, having it predict the parameters of a QNM physics model. Through this approach, we simplify the problem that the network needs to learn, while also satisfying the constraints of the model and being more robust to noise.

Paper III

Determining the dispersion and nonlinear characteristics of 3D periodic waveguides using finite-element eigenmode simulations

In paper III, we present a method for determining the dispersion and nonlinearity coefficients for 3D waveguides that are periodic in the propagation direction. Recall from Ch. 4 that we can solve for the nonlinear response of a waveguide by means of the nonlinear Schrödinger equations (NLSE), which in turn contain constants that we can calculate from linear full-wave simulations.

Reducing simulation times is always desirable, and in this paper we report a method using the finite element method (FEM) and an eigenvalue solver to calculate the dispersion of waveguide modes at a specified wavelength, making it possible to use this method only at frequencies of interest without having to sweep over a range of wave vectors. This facilitates the simulation of periodic integrated waveguides with patterned unit cells, where simulation time often is the major challenge when using alterative methods. We benchmark the method against systems with known analytical solutions in the form of a slab waveguide and a Bragg reflector. Furthermore, we compare our method with conventional boundary mode analysis for a photonic crystal fibre (PCF).

Paper IV

Inverse design of optical waveguides for phase-sensitive amplifiers using machine learning

In paper IV, we present our work on inverse design of suitable waveguide cross sections for an on-chip Si_3N_4 platform that favours nonlinear four-wave mixing (FWM) for use with optical parametric amplifiers (OPAs).

As we have treated continuously throughout this thesis, the process of predicting how a nanophotonic structure should look to fulfil its design targets is a process consisting of many different parts. In this work, where the cross section of an optical waveguide is optimised through inverse design to produce dispersion that benefits a phase-sensitive amplifier (PSA), incorporates aspects of every chapter so far in this thesis. The approach used in this paper is based on using machine learning, through the training of a conditional variational autoencoder (cVAE), in order to predict waveguide geometries that yield a desired dispersion. We present how we parametrise the problem and generate a dataset for the training of the NN, based on mode analysis, the finite element method (FEM), and mode selection.

9

Conclusions and outlook

We have now reached the final vantage point of this thesis where I will pause briefly and allow myself some reflections on the progress of this field and turn the looking glass towards the horizon to attempt to catch a glimpse of what lies ahead.

In the grand scheme of things, the subject of nonlinear physics is a rather new field, yet also quite old, emerging during the 1900s resting comfortably on the shoulders of electromagnetism and quantum physics. The basic physics has been well studied now for several generations. Yet the field is thriving and expanding at the moment. Our technology in lasers and manufacturing of photonic devices has now improved to such a point that we now can manufacture many things that were previously unthinkable. The possibility to use 2020s lithography techniques (now referred to as "modern") developed much for the purpose of current electronics manufacturing opens up the door for amazing materials design. We can now construct optical amplification below conventional quantum limits [1] with on-chip manufacturing 57 years after Miller coined the term integrated optics [248], a testament to the quality of today's manufacturing prowess.

We are now in a regime where, for many applications, the manufacturing part is no longer the limiting factor when designing a device, but rather our ability to come up with the designs themselves. With more complicated geometries and designs we also see the numerical calculations taking a larger chunk of the work time associated with designing devices. Improvement in computational methods will not necessary change this, since if we have access to more compute power, that also means we can tackle bigger and more complex systems. This, in turn, creates a need for more computational power, leading to us simulating more complex systems, completing the circle.

The process of setting up simulations, however, is labour intensive. Depending on the problem, there are different methods to choose from and many possible parametrisations. Solutions that attempt to address this, through automated pipelines, such as

agentic systems, still relies on detailed domain knowledge to set up and monitor, costing an equivalent amount of work time as alternative methods. Here I believe that we will see a great impact in the future decades. I think our increased access to computational resources will enable us to create more optimized and polished software pipelines that ultimately decrease the work time the researchers and designers needs to spend on software implementation, time which then can be focused on defining good goals for the automated process. The challenges we face are that of fragmentation. Each of our physics problems are highly specialized with huge variations and a pipeline for one problem is seldom usable for another. I think we still have a long way to go until we see a scenario where an experimentalist or system designer that wants some manufacturable photonic crystal design that meets the design requirement can get such a design at the press of a button without a lengthy design process, but I think we are approaching such a future, maybe within our lifetime. Truly, science is progressing and paints a bright future.

Acknowledgments

Many people really deserve to be thanked for the feedback and support they have given me during this work, and I would like to thank you all for your assistance and encouragement. I would like to thank my supervisor Philippe Tassin in particular for the insights and light-hearted discussions over the years that led up to this thesis. Likewise, I would like to thank my examiner Andreas for the support over the years and for contributing to and ensuring a good and friendly working environment. A big thank you to all the members of our group and collaborators for support and interesting discussions over the years.

My work was funded by the Swedish Research Council under Grant No. 2020-05284 and relied heavily on the resources provided through the Swedish National Infrastructure for Computing (NAISS), at Chalmers/C3SE and KTH/PDC sites, partially also funded by the Swedish Research Council through the grant No. 2022-06725. I also would like to thank the Excellence Centre META-PIX for fruitful discussions and activities.

Moving slightly to the topic of ”everyday well-being at work”, there is a group of people I would especially like to thank and they are the ”lunch regulars” who frequent the lunch room and my regular lunch buddy Joel. I have always enjoyed the range of subjects discussed during these times, especially when they take a turn towards the philosophical, political, controversial, and obscure. Finally, I would like to thank Ingrid, who both put up with me during my work and continuously supported me.

Bibliography

[1] Z. Ye, P. Zhao, K. Twayana, M. Karlsson, V. Torres-Company, and P. A. Andrekson, *Overcoming the quantum limit of optical amplification in monolithic waveguides,* Science Advances **7**, eabi8150 (2021). doi:10.1126/sciadv.abi8150.

[2] J. W. Choi, B.-U. Sohn, E. Sahin, G. F. R. Chen, P. Xing, D. K. T. Ng, B. J. Eggleton, and D. T. H. Tan, *An optical parametric Bragg amplifier on a CMOS chip,* Nanophotonics **10**, 3507 (2021). doi:10.1515/nanoph-2021-0302.

[3] D. J. Griffiths, *Introduction to Electrodynamics* (Cambridge University Press, 2023).

[4] J. D. Jackson, *Classical electrodynamics* (New York, NY: Wiley, 1999). ISBN 9780471309321.

[5] A. Yariv and P. Yeh, *Optical Waves in Crystals: Propagation and Control of Laser Radiation* (New York, NY: Wiley, 2003). ISBN 0471430811.

[6] J. l. R. d'Alembert, *Recherches Sur La Courbe Que Forme Une Corde* (Deutsche Akademie der Wissenschaften zu Berlin, 1747).

[7] J. l. R. d'Alembert, *Suites Des Recherches Sur La Courbe Que Forme Une Corde* (Deutsche Akademie der Wissenschaften zu Berlin, 1747).

[8] J. l. R. d'Alembert, *Addition Au Memoire Sur La Courbe Que Forme Une Corde* (Deutsche Akademie der Wissenschaften zu Berlin, 1750).

[9] L. Euler, *De la propagation du son,* Mémoires de l'académie des sciences de Berlin , (1759). https://archive.org/details/euler-e305.

[10] J. L. d. Lagrange, *Nouvelles recherches sur la nature et la propagation du son* (Gauthier-Villars, 1760).

[11] J. C. Maxwell, *VIII. A dynamical theory of the electromagnetic field,* Philosophical Transactions of the Royal Society of London **155**, 459 (1865). doi:10.1098/rstl.1865.0008.

[12] H. Hertz, *Electric Waves: Being Researches on the Propagation of Electric Action with Finite Velocity Through Space* (Dover Publications, 1893). ISBN 9780486600574.

[13] O. Heaviside, *Electromagnetic Theory* ("The Electrician" printing and publishing Company, 1893).

[14] O. Heaviside, *Electromagnetic Theory. Oliver Heaviside. Complete and Unabridged Edition of Volume I … II and III with a Critical and Historical Introduction by Ernst Weber* (Dover publications, 1950).

[15] J. Liu, *Photonic Devices* (Cambridge University Press, 2009). ISBN 9781139441148.

[16] B. E. A. Saleh and M. C. Teich, *Nonlinear Optics*. In , *Fundamentals of Photonics* (John Wiley & Sons, Ltd, 1991). doi:10.1002/0471213748.ch19.

[17] P. Markos and C. M. Soukoulis, *Wave propagation: from electrons to photonic crystals and left-handed materials* (Princeton University Press, 2008).

[18] C. Kittel, *Introduction to Solid State Physics* (Wiley, 2004). ISBN 9780471415268.

[19] N. Ashcroft and N. Mermin, *Solid State Physics* (Cengage, 2021). ISBN 9780357670811.

[20] F. Bloch, *Über die Quantenmechanik der Elektronen in Kristallengittern,* Zeitschrift für Physik **52**, 555 (1929). doi:10.1007/BF01339455.

[21] J. M. Ziman, *Principles of the Theory of Solids* (Cambridge University Press, 1972).

[22] J. V. Moloney and A. C. Newell, *Nonlinear Optics* (Westview Press, 2004). ISBN 0-8133-4118-3.

[23] R. W. Boyd, ed., *Nonlinear Optics (Third Edition)* (Burlington: Academic Press, 2008). ISBN 978-0-12-369470-6. doi:10.1016/B978-0-12-369470-6.00020-4.

[24] N. Bloembergen, *Nonlinear Optics* (World Scientific, 1996). doi:10.1142/3046.

[25] ISPNLO, *International School on Parametric Nonlinear Optics,* https://www.youtube.com/@ispnlo9041/, 2015.

[26] K. Dolgaleva and R. W. Boyd, *Local-field effects in nanostructured photonic materials,* Adv. Opt. Photon. **4**, 1 (2012). doi:10.1364/AOP.4.000001.

[27] P. N. Butcher and D. Cotter, *The Elements of Nonlinear Optics* (Cambridge University Press, 1990).

[28] A. Einstein, *Die Grundlage der allgemeinen Relativitätstheorie,* Annalen der Physik **354**, 769 (1916). doi:10.1002/andp.19163540702.

[29] P. Dirac, *The Principles of Quantum Mechanics* (Clarendon Press, 1981). ISBN 9780198520115.

[30] G. B. Arfken, H. J. Weber, and F. E. Harris, *Mathematical Methods for Physicists* (Boston: Academic Press, 2013). ISBN 978-0-12-384654-9. doi:10.1016/B978-0-12-384654-9.00011-6.

[31] P. A. Franken, A. E. Hill, C. W. Peters, and G. Weinreich, *Generation of Optical Harmonics,* Phys. Rev. Lett. **7**, 118 (1961). doi:10.1103/PhysRevLett.7.118.

[32] W. Kaiser and C. G. B. Garrett, *Two-Photon Excitation in CaF_2: Eu^{2+},* Phys. Rev. Lett. **7**, 229 (1961). doi:10.1103/PhysRevLett.7.229.

[33] M. Göppert-Mayer, *Über Elementarakte mit zwei Quantensprüngen,* Annalen der Physik **401**, 273 (1931). doi:10.1002/andp.19314010303.

[34] R. W. Terhune, P. D. Maker, and C. M. Savage, *Optical Harmonic Generation in Calcite,* Phys. Rev. Lett. **8**, 404 (1962). doi:10.1103/PhysRevLett.8.404.

[35] J. A. Armstrong, N. Bloembergen, J. Ducuing, and P. S. Pershan, *Interactions between Light Waves in a Nonlinear Dielectric,* Phys. Rev. **127**, 1918 (1962). doi:10.1103/PhysRev.127.1918.

[36] J. E. Midwinter and J. Warner, *The effects of phase matching method and of uniaxial crystal symmetry on the polar distribution of second-order non-linear optical polarization,* British Journal of Applied Physics **16**, 1135 (1965). doi:10.1088/0508-3443/16/8/312.

[37] M. V. Hobden, *Phase-Matched Second-Harmonic Generation in Biaxial Crystals,* Journal of Applied Physics **38**, 4365 (1967). doi:10.1063/1.1709130.

[38] D. S. Hum and M. M. Fejer, *Quasi-phasematching,* Comptes Rendus Physique **8**, 180 (2007). doi:10.1016/j.crhy.2006.10.022.

[39] S. ning Zhu, Y. yuan Zhu, and N. ben Ming, *Quasi-Phase-Matched Third-Harmonic Generation in a Quasi-Periodic Optical Superlattice,* Science **278**, 843 (1997). doi:10.1126/science.278.5339.843.

[40] R. L. Byer, *Quasi-Phasematched Nonlinear Interactions and Devices,* Journal of Nonlinear Optical Physics & Materials **06**, 549 (1997). doi:10.1142/S021886359700040X.

[41] J. S. Pelc, C. R. Phillips, D. Chang, C. Langrock, and M. M. Fejer, *Efficiency pedestal in quasi-phase-matching devices with random duty-cycle errors,* Opt. Lett. **36**, 864 (2011). doi:10.1364/OL.36.000864.

[42] M. Fejer, G. Magel, D. Jundt, and R. Byer, *Quasi-phase-matched second harmonic generation: tuning and tolerances,* IEEE Journal of Quantum Electronics **28**, 2631 (1992). doi:10.1109/3.161322.

[43] G. Agrawal, *Nonlinear Fiber Optics* (Elsevier Science, 2019). ISBN 9780128170434.

[44] S. E. Harris, M. K. Oshman, and R. L. Byer, *Observation of Tunable Optical Parametric Fluorescence,* Phys. Rev. Lett. **18**, 732 (1967). doi:10.1103/PhysRevLett.18.732.

[45] W. H. Louisell, A. Yariv, and A. E. Siegman, *Quantum Fluctuations and Noise in Parametric Processes. I.,* Phys. Rev. **124**, 1646 (1961). doi:10.1103/PhysRev.124.1646.

[46] C. M. Caves, *Quantum limits on noise in linear amplifiers,* Phys. Rev. D **26**, 1817 (1982). doi:10.1103/PhysRevD.26.1817.

[47] P. Zhao, V. Shekhawat, M. Girardi, Z. He, V. Torres-Company, and P. A. Andrekson, *Ultra-broadband optical amplification using nonlinear integrated waveguides,* Nature **640**, 918 (2025). doi:10.1038/s41586-025-08824-3.

[48] R. Mears, L. Reekie, I. Jauncey, and D. Payne, *Low-noise erbium-doped fibre amplifier operating at 1.54μm,* Electronics Letters **23**, 1026 (1987). doi:10.1049/el:19870719.

[49] P. A. Andrekson and M. Karlsson, *Fiber-based phase-sensitive optical amplifiers and their applications,* Adv. Opt. Photon. **12**, 367 (2020). doi:10.1364/AOP.382548.

[50] R. Slavík, F. Parmigiani, J. Kakande, C. Lundström, M. Sjödin, P. A. Andrekson, R. Weerasuriya, S. Sygletos, A. D. Ellis, L. Grüner-Nielsen, D. Jakobsen, S. Herstrøm, R. Phelan, J. O'Gorman, A. Bogris, D. Syvridis, S. Dasgupta, P. Petropoulos, and D. J. Richardson, *All-optical phase and amplitude regenerator for next-generation telecommunications systems,* Nature Photonics **4**, 690 (2010). doi:10.1038/nphoton.2010.203.

[51] K. J. Lee, F. Parmigiani, S. Liu, J. Kakande, P. Petropoulos, K. Gallo, and D. Richardson, *Phase sensitive amplification based on quadratic cascading in a periodically poled lithium niobate waveguide,* Opt. Express **17**, 20393 (2009). doi:10.1364/OE.17.020393.

[52] Z. Tong, C. Lundström, P. A. Andrekson, C. J. McKinstrie, M. Karlsson, D. J. Blessing, E. Tipsuwannakul, B. J. Puttnam, H. Toda, and L. Grüner-Nielsen, *Towards ultrasensitive optical links enabled by low-noise phase-sensitive amplifiers,* Nature Photonics **5**, 430 (2011). doi:10.1038/nphoton.2011.79.

[53] R. Kakarla, J. Schröder, and P. A. Andrekson, *One photon-per-bit receiver using near-noiseless phase-sensitive amplification,* Light: Science & Applications **9**, 153 (2020). doi:10.1038/s41377-020-00389-2.

[54] K. Vijayan, A. Mirani, J. Schröder, M. Karlsson, and P. Andrekson, *Capacity of phase-sensitively preamplified optical links at low signal-to-noise ratio,* in *2022 European Conference on Optical Communication (ECOC)*, 1, 2022.

[55] T. Kazama, T. Umeki, S. Shimizu, T. Kashiwazaki, K. Enbutsu, R. Kasahara, Y. Miyamoto, and K. Watanabe, *Over-30-dB gain and 1-dB noise figure phase-sensitive amplification using a pump-combiner-integrated fiber I/O PPLN module,* Opt. Express **29**, 28824 (2021). doi:10.1364/OE.434601.

[56] R. Larsson, R. U. Weerasuriya, and P. A. Andrekson, *Ultralow-noise preamplified optical receiver using conventional single-wavelength transmission,* Optica **11**, 1497 (2024). doi:10.1364/OPTICA.539544.

[57] V. Shekhawat, P. Zhao, N. Lindvall, M. Girardi, P. A. Andrekson, and V. Torres-Company, *Rib waveguides for Kerr nonlinear optics,* Opt. Express **33**, 50811 (2025). doi:10.1364/OE.574369.

[58] Z. Tong and S. Radic, *Low-noise optical amplification and signal processing in parametric devices,* Adv. Opt. Photon. **5**, 318 (2013). doi:10.1364/AOP.5.000318.

[59] M. Karlsson, *Transmission Systems With Low Noise Phase-Sensitive Parametric Amplifiers,* Journal of Lightwave Technology **34**, 1411 (2016). doi:10.1109/JLT.2015.2507866.

[60] M. E. Marhic, P. A. Andrekson, P. Petropoulos, S. Radic, C. Peucheret, and M. Jazayerifar, *Fiber optical parametric amplifiers in optical communication systems,* Laser & Photonics Reviews **9**, 50 (2015). doi:https://doi.org/10.1002/lpor.201400087.

[61] S. Obayya, *Computational Photonics* (Wiley, 2011). ISBN 978-1-119-95750-8.

[62] E. Schrödinger, *An Undulatory Theory of the Mechanics of Atoms and Molecules,* Phys. Rev. **28**, 1049 (1926). doi:10.1103/PhysRev.28.1049.

[63] R. Y. Chiao, E. Garmire, and C. H. Townes, *Self-Trapping of Optical Beams,* Phys. Rev. Lett. **13**, 479 (1964). doi:10.1103/PhysRevLett.13.479.

[64] L. D. Landau and V. L. Ginzburg, *On the theory of superconductivity,* JETP **20**, (1950).

[65] L. D. Landau, *73 - On the theory of superconductivity,* in *Collected Papers of L.D. Landau*, edited by D. TER HAAR (Pergamon, 1965), p. 546. doi:10.1016/B978-0-08-010586-4.50078-X.

[66] M. E. Marhic, *Fiber Optical Parametric Amplifiers, Oscillators and Related Devices* (Cambridge University Press, 2007).

[67] R. Y. Chiao, E. Garmire, and C. H. Townes, *Self-Trapping of Optical Beams,* Phys. Rev. Lett. **13**, 479 (1964). doi:10.1103/PhysRevLett.13.479.

[68] A. Hasegawa and F. Tappert, *Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers. II. Normal dispersion,* Applied Physics Letters **23**, 171 (1973). doi:10.1063/1.1654847.

[69] F. D. Tappert and C. N. Judice, *Recurrence of Nonlinear Ion Acoustic Waves,* Phys. Rev. Lett. **29**, 1308 (1972). doi:10.1103/PhysRevLett.29.1308.

[70] J. Fleck, J. R. Morris, and M. D. Feit, *Time-dependent propagation of high energy laser beams through the atmosphere,* Applied Physics **10**, 129 (1976). doi:10.1007/BF00896333.

[71] R. Courant, *Variational methods for the solution of problems of equilibrium and vibrations,* Bull. Am. Math. **49**, 1 (1943). doi:10.1090/S0002-9904-1943-07818-4.

[72] J.-M. Jin, *The Finite Element Method in Electromagnetics* (Wiley, 2014). ISBN 978-1-118-57136-1.

[73] N. Ottosen and H. Petersson, *Introduction to the Finite Element Method* (Prentice-Hall, 1992). ISBN 0-13-473877-2.

[74] *Comsol Multiphysics Reference Manual, pp. 1422-1440. COMSOL Multiphysics® v. 6.2. COMSOL AB, Stockholm, Sweden. 2024",.*

[75] *Comsol Multiphysics Reference Manual, pp. 1268-1342. COMSOL Multiphysics® v. 6.2. COMSOL AB, Stockholm, Sweden. 2024",.*

[76] https://www.comsol.com/multiphysics/finite-element-method.

[77] A. J. Svärdsby and P. Tassin, *Adaptive meshing strategies for nanophotonics using a posteriori error estimation,* Opt. Express **32**, 24592 (2024). doi:10.1364/OE.523907.

[78] *COMSOL Multiphysics,* Stockholm, Sweden, 2022. https://www.comsol.com/.

[79] B. Galerkin, *"On electrical circuits for the approximate solution of the Laplace equation",* Vestnik Inzh , 897 (1915).

[80] G. Floquet, *Sur les équations différentielles linéaires à coefficients périodiques,* Annales scientifiques de l'École Normale Supérieure **2e série, 12**, 47 (1883). doi:10.24033/asens.220.

[81] S. G. Johnson, *Notes on Perfectly Matched Layers (PMLs),* 2021. https://arxiv.org/abs/2108.05348.

[82] J.-P. Berenger, *A perfectly matched layer for the absorption of electromagnetic waves,* Journal of Computational Physics **114**, 185 (1994). doi:10.1006/jcph.1994.1159.

[83] G. Robin, *Sur la distribution de l'électricité à la surface des conducteurs fermés des conducteurs ouverts,* Annales scientifiques de l'École Normale Supérieure **3**, 3 (1886). http://eudml.org/doc/80956.

[84] K. Gustafson and T. Abe, *The third boundary condition - was it Robin's?,* The matematical Intelligencer , 63 (1998). doi:10.1007/BF03024402.

[85] J. D. Joannopoulos, S. G. Johnson, J. N. Winn, and R. D. Meade, *Photonic Crystals: Molding the Flow of Light - Second Edition* (Princeton University Press, 2008). ISBN 9780691124568. doi:10.2307/j.ctvcm4gz9.

[86] K. Sakoda, *Optical Properties of Photonic Crystals* (Springer Berlin, Heidelberg, 2005). ISBN 978-3-642-42408-3. doi:10.1007/b138376.

[87] M. Davanço, Y. Urzhumov, and G. Shvets, *The complex Bloch bands of a 2D plasmonic crystal displaying isotropic negative refraction,* Opt. Express **15**, 9681 (2007). doi:10.1364/OE.15.009681.

[88] C. Fietz, Y. Urzhumov, and G. Shvets, *Complex k band diagrams of 3D metamaterial/photonic crystals.,* Opt. Express **19**, 19027 (2011). doi:10.1364/OE.19.019027.

[89] Z. Zhu and T. G. Brown, *Full-vectorial finite-difference analysis of microstructured optical fibers,* Opt. Express **10**, 853 (2002). doi:10.1364/OE.10.000853.

[90] C. Vassallo, *Optical Waveguide Concepts* (Elsevier, 1991). ISBN 9780444886842.

[91] D. Marcuse, *The coupling of degenerate modes in two parallel dielectric waveguides,* The Bell System Technical Journal **50**, 1791 (1971). doi:10.1002/j.1538-7305.1971.tb02582.x.

[92] R. Vanclooster and P. Phariseau, *The coupling of two parallel dielectric fibers: I. Basic equations,* Physica **47**, 485 (1970). doi:https://doi.org/10.1016/0031-8914(70)90125-4.

[93] M. Digonnet and H. Shaw, *Analysis of a Tunable Single Mode Optical Fiber Coupler,* IEEE Transactions on Microwave Theory and Techniques **30**, 592 (1982). doi:10.1109/TMTT.1982.1131101.

[94] A. W. Snyder and J. D. Love, *Optical Waveguide Theory* (Springer, 1983). ISBN 9780412099502. doi:10.1007/978-1-4613-2813-1.

[95] M. Masek, M. Capek, L. Jelinek, and K. Schab, *Modal Tracking Based on Group Theory,* IEEE Transactions on Antennas and Propagation **68**, 927 (2020). doi:10.1109/TAP.2019.2943354.

[96] G. K. Svendsen, M. W. Haakestad, and J. Skaar, *Reciprocity and the scattering matrix of waveguide modes,* Phys. Rev. A **87**, 013838 (2013). doi:10.1103/PhysRevA.87.013838.

[97] K. R. Schab and J. T. Bernhard, *A Group Theory Rule for Predicting Eigenvalue Crossings in Characteristic Mode Analyses,* IEEE Antennas and Wireless Propagation Letters **16**, 944 (2017). doi:10.1109/LAWP.2016.2615041.

[98] K. Kurokawa, *Power Waves and the Scattering Matrix,* IEEE Transactions on Microwave Theory and Techniques **13**, 194 (1965). doi:10.1109/TMTT.1965.1125964.

[99] D. M. Pozar, *Microwave engineering* (Fourth edition. Hoboken, NJ : Wiley, [2012] ©2012, 2012).

[100] F. Alpeggiani, N. Parappurath, E. Verhagen, and L. Kuipers, *Quasinormal-Mode Expansion of the Scattering Matrix,* Phys. Rev. X **7**, 021035 (2017). doi:10.1103/PhysRevX.7.021035.

[101] H. Zhang and O. D. Miller, *Quasinormal Coupled Mode Theory,* 2020. https://arxiv.org/abs/2010.08650.

[102] P. Lalanne, W. Yan, K. Vynck, C. Sauvan, and J.-P. Hugonin, *Light Interaction with Photonic and Plasmonic Resonances,* Laser & Photonics Reviews **12**, 1700113 (2018). doi:https://doi.org/10.1002/lpor.201700113.

[103] M. Benzaouia, J. D. Joannopoulos, S. G. Johnson, and A. Karalis, *Quasi-normal mode theory of the scattering matrix, enforcing fundamental constraints for truncated expansions,* Phys. Rev. Res. **3**, 033228 (2021). doi:10.1103/PhysRevResearch.3.033228.

[104] V. A. Lilja, A. J. Svärdsby, T. Gahlmann, and P. Tassin, *A general framework for knowledge integration in machine learning for electromagnetic scattering using quasinormal modes,* 2025. https://arxiv.org/abs/2509.06130.

[105] C. Liu, X. Zhang, J. Chang, Y. Lyu, J. Zhao, and S. Qiu, *Programmable mechanical metamaterials: basic concepts, types, construction strategies—a review,* Frontiers in Materials **Volume 11 - 2024**, (2024). doi:10.3389/fmats.2024.1361408.

[106] T. Shah, C. Brendel, V. Peano, and F. Marquardt, *Colloquium: Topologically protected transport in engineered mechanical systems,* Rev. Mod. Phys. **96**, 021002 (2024). doi:10.1103/RevModPhys.96.021002.

[107] H. Huang, J. Chen, and S. Huo, *Recent advances in topological elastic metamaterials,* Journal of Physics: Condensed Matter **33**, 503002 (2021). doi:10.1088/1361-648X/ac27d8.

[108] G. Aydın and S. E. San, *Breaking the limits of acoustic science: A review of acoustic metamaterials,* Materials Science and Engineering: B **305**, 117384 (2024). doi:https://doi.org/10.1016/j.mseb.2024.117384.

[109] G. Ma and P. Sheng, *Acoustic metamaterials: From local resonances to broad horizons,* Science Advances **2**, e1501595 (2016). doi:10.1126/sciadv.1501595.

[110] J. F. Ramsay, *Microwave Antenna and Waveguide Techniques before 1900,* Proceedings of the IRE **46**, 405 (1958). doi:10.1109/JRPROC.1958.286869.

[111] W. E. Kock, *Metallic delay lenses,* The Bell System Technical Journal **27**, 58 (1948). doi:10.1002/j.1538-7305.1948.tb01331.x.

[112] J. B. Pendry, *Negative Refraction Makes a Perfect Lens,* Phys. Rev. Lett. **85**, 3966 (2000). doi:10.1103/PhysRevLett.85.3966.

[113] D. R. Smith, W. J. Padilla, D. C. Vier, S. C. Nemat-Nasser, and S. Schultz, *Composite Medium with Simultaneously Negative Permeability and Permittivity,* Phys. Rev. Lett. **84**, 4184 (2000). doi:10.1103/PhysRevLett.84.4184.

[114] D. Schurig, J. J. Mock, B. J. Justice, S. A. Cummer, J. B. Pendry, A. F. Starr, and D. R. Smith, *Metamaterial Electromagnetic Cloak at Microwave Frequencies,* Science **314**, 977 (2006). doi:10.1126/science.1133628.

[115] R. A. Shelby, D. R. Smith, and S. Schultz, *Experimental Verification of a Negative Index of Refraction,* Science **292**, 77 (2001). doi:10.1126/science.1058847.

[116] P. Vukusic, J. Sambles, and C. Lawrence, *Colour mixing in wing scales of a butterfly,* Nature **404**, 457 (2000). doi:10.1038/35006561.

[117] D. Zhang, W. Zhang, J. Gu, T. Fan, Q. Liu, H. Su, and S. Zhu, *Inspiration from butterfly and moth wing scales: Characterization, modeling, and fabrication,* Progress in Materials Science **68**, 67 (2015). doi:https://doi.org/10.1016/j.pmatsci.2014.10.003.

[118] A. Pitilakis, O. Tsilipakos, F. Liu, K. M. Kossifos, A. C. Tasolamprou, D.-H. Kwon, M. S. Mirmoosa, D. Manessis, N. V. Kantartzis, C. Liaskos, M. A. Antoniades, J. Georgiou, C. M. Soukoulis, M. Kafesaki, and S. A. Tretyakov, *A Multi-Functional Reconfigurable Metasurface: Electromagnetic Design Accounting for Fabrication Aspects,* IEEE Transactions on Antennas and Propagation **69**, 1440 (2021). doi:10.1109/TAP.2020.3016479.

[119] O. Tsilipakos, A. C. Tasolamprou, A. Pitilakis, F. Liu, X. Wang, M. S. Mirmoosa, D. C. Tzarouchis, S. Abadal, H. Taghvaee, C. Liaskos, A. Tsioliaridou, J. Georgiou, A. Cabellos-Aparicio, E. Alarcón, S. Ioannidis, A. Pitsillides, I. F. Akyildiz, N. V. Kantartzis, E. N. Economou, C. M. Soukoulis, M. Kafesaki, and S. Tretyakov, *Toward Intelligent Metasurfaces: The Progress from Globally Tunable Metasurfaces to Software-Defined Metasurfaces with an Embedded Network of Controllers,* Advanced Optical Materials **8**, 2000783 (2020). doi:https://doi.org/10.1002/adom.202000783.

[120] Y. Yang, Y. Jeon, Z. Dong, J. K. W. Yang, M. Haddadi Moghaddam, D.-S. Kim, D. K. Oh, J. Lee, M. Hentschel, H. Giessen, D. Kang, G. Kim, T. Tanaka, Y. Zhao, J. Bürger, S. A. Maier, H. Ren, W. Jung, M. Choi, G. Bae, H. Chen, S. Jeon, J. Kim, E. Lee, H. Kang, Y. Park, D. Du Nguyen, I. Kim, P. Cencillo-Abad, D. Chanda, X. Jing, N. Liu, I. V. Martynenko, T. Liedl, Y. Kwak, J.-M. Nam, S.-M. Park, T. W. Odom, H.-E. Lee, R. M. Kim, K. T. Nam, H. Kwon, H.-H. Jeong, P. Fischer, J. Yoon, S.-H. Kim, S. Shim, D. Lee, L. A. Pérez, X. Qi, A. Mihi, H. Keum, M. Shim, S. Kim, H. Jang, Y. S. Jung, C. Rossner, T. A. König, A. Fery, Z. Li, K. Aydin, C. A. Mirkin, J. Seong, N. Jeon, Z. Xu, T. Gu, J. Hu, H. Kwon, H. Jung,

H. Alijani, I. Aharonovich, J. Kim, and J. Rho, *Nanofabrication for Nanophotonics,* ACS Nano **19**, 12491 (2025). doi:10.1021/acsnano.4c10964.

[121] S. He, Y. Tian, H. Zhou, M. Zhu, C. Li, B. Fang, Z. Hong, and X. Jing, *Review for Micro-Nano Processing Technology of Microstructures and Metadevices,* Advanced Functional Materials **35**, 2420369 (2025). doi:https://doi.org/10.1002/adfm.202420369.

[122] L. Martin-Monier, S. Pajovic, M. G. Abebe, J. Chen, S. Vaidya, S. Min, S. Choi, S. E. Kooi, B. Maes, J. Hu, M. Soljačić, and C. Roques-Carmes, *Large-scale self-assembled nanophotonic scintillators for X-ray imaging,* Nature Communications **16**, 5750 (2025). doi:10.1038/s41467-025-60953-5.

[123] G. von Freymann, A. Ledermann, M. Thiel, I. Staude, S. Essig, K. Busch, and M. Wegener, *Three-Dimensional Nanostructures for Photonics,* Advanced Functional Materials **20**, 1038 (2010). doi:https://doi.org/10.1002/adfm.200901838.

[124] H. Kollmann, X. Piao, M. Esmann, S. F. Becker, D. Hou, C. Huynh, L.-O. Kautschor, G. Bösker, H. Vieker, A. Beyer, A. Gölzhäuser, N. Park, R. Vogelgesang, M. Silies, and C. Lienau, *Toward Plasmonics with Nanometer Precision: Nonlinear Optics of Helium-Ion Milled Gold Nanoantennas,* Nano Letters **14**, 4778 (2014). doi:10.1021/nl5019589.

[125] G. E. Moore, *Cramming more components onto integrated circuits,* Electronics **38**, 114 (1965).

[126] M. E. Warren, R. E. Smith, G. A. Vawter, and J. R. Wendt, *High-efficiency subwavelength diffractive optical element in GaAs for 975 nm,* Opt. Lett. **20**, 1441 (1995). doi:10.1364/OL.20.001441.

[127] F. T. Chen and H. G. Craighead, *Diffractive phase elements based on two-dimensional artificial dielectrics,* Opt. Lett. **20**, 121 (1995). doi:10.1364/OL.20.000121.

[128] P. Lalanne, S. Astilean, P. Chavel, E. Cambril, and H. Launois, *Design and fabrication of blazed binary diffractive elements with sampling periods smaller than the structural cutoff,* J. Opt. Soc. Am. A **16**, 1143 (1999). doi:10.1364/JOSAA.16.001143.

[129] L. Martín-Moreno, F. J. García-Vidal, H. J. Lezec, K. M. Pellerin, T. Thio, J. B. Pendry, and T. W. Ebbesen, *Theory of Extraordinary Optical Transmission through Subwavelength Hole Arrays,* Phys. Rev. Lett. **86**, 1114 (2001). doi:10.1103/PhysRevLett.86.1114.

[130] A. V. Kildishev, A. Boltasseva, and V. M. Shalaev, *Planar Photonics with Metasurfaces,* Science **339**, 1232009 (2013). doi:10.1126/science.1232009.

[131] N. Yu, P. Genevet, M. A. Kats, F. Aieta, J.-P. Tetienne, F. Capasso, and Z. Gaburro, *Light Propagation with Phase Discontinuities: Generalized Laws of Reflection and Refraction,* Science **334**, 333 (2011). doi:10.1126/science.1210713.

[132] J. Lu and J. Vučković, *Nanophotonic computational design,* Opt. Express **21**, 13351 (2013). doi:10.1364/OE.21.013351.

[133] A. Y. Piggott, J. Lu, K. G. Lagoudakis, J. Petykiewicz, T. M. Babinec, and J. Vučković, *Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer,* Nature Photonics **9**, 374 (2015). doi:10.1038/nphoton.2015.69.

[134] L. Su, A. Y. Piggott, N. V. Sapra, J. Petykiewicz, and J. Vučković, *Inverse Design and Demonstration of a Compact on-Chip Narrowband Three-Channel Wavelength Demultiplexer,* ACS Photonics **5**, 301 (2018). doi:10.1021/acsphotonics.7b00987.

[135] D. Sell, J. Yang, S. Doshay, R. Yang, and J. A. Fan, *Large-Angle, Multifunctional Metagratings Based on Freeform Multimode Geometries,* Nano Letters **17**, 3752 (2017). doi:10.1021/acs.nanolett.7b01082.

[136] W.-K. Lee, S. Yu, C. J. Engel, T. Reese, D. Rhee, W. Chen, and T. W. Odom, *Concurrent design of quasi-random photonic nanostructures,* Proceedings of the National Academy of Sciences **114**, 8734 (2017). doi:10.1073/pnas.1704711114.

[137] M. Meem, S. Banerji, A. Majumder, F. G. Vasquez, B. Sensale-Rodriguez, and R. Menon, *Broadband lightweight flat lenses for long-wave infrared imaging,* Proceedings of the National Academy of Sciences **116**, 21375 (2019). doi:10.1073/pnas.1908447116.

[138] A. C. Tasolamprou, E. Skoulas, G. Perrakis, M. Vlahou, Z. Viskadourakis, E. N. Economou, M. Kafesaki, G. Kenanakis, and E. Stratakis, *Highly ordered laser imprinted plasmonic metasurfaces for polarization sensitive perfect absorption,* Scientific Reports **12**, 19769 (2022). doi:10.1038/s41598-022-21647-w.

[139] M. Atwany, S. Pardo, S. Serunjogi, and M. Rasras, *A review of emerging trends in photonic deep learning accelerators,* Frontiers in Physics **12**, (2024). doi:10.3389/fphy.2024.1369099.

[140] H. Zhang, L. Kang, S. D. Campbell, J. T. Young, and D. H. Werner, *Data driven approaches in nanophotonics: a review of AI-enabled metadevices,* Nanoscale , (2025). doi:10.1039/d5nr02043c.

[141] T. Fu, J. Zhang, R. Sun, Y. Huang, W. Xu, S. Yang, Z. Zhu, and H. Chen, *Optical neural networks: progress and challenges,* Light: Science & Applications **13**, (2024). doi:10.1038/s41377-024-01590-3.

[142] N. Wu, Y. Sun, J. Hu, C. Yang, Z. Bai, F. Wang, X. Cui, S. He, Y. Li, C. Zhang, K. Xu, J. Guan, S. Xiao, and Q. Song, *Intelligent nanophotonics: when machine learning sheds light,* eLight **5**, 5 (2025). doi:10.1186/s43593-025-00085-x.

[143] I. Malkiel, M. Mrejen, L. Wolf, and H. Suchowski, *Machine learning for nanophotonics,* MRS Bulletin **45**, 221 (2020). doi:10.1557/mrs.2020.66.

[144] P. Genevet, F. Capasso, F. Aieta, M. Khorasaninejad, and R. Devlin, *Recent advances in planar optics: from plasmonic to dielectric metasurfaces,* Optica **4**, 139 (2017). doi:10.1364/OPTICA.4.000139.

[145] J. Engelberg and U. Levy, *The advantages of metalenses over diffractive lenses,* Nature Communications **11**, (2020). doi:10.1038/s41467-020-15972-9.

[146] A. Majumder, B. Shen, R. Polson, and R. Menon, *Ultra-compact polarization rotation in integrated silicon photonics using digital metamaterials,* Opt. Express **25**, 19721 (2017). doi:10.1364/OE.25.019721.

[147] S. A. Schulz, R. F. Oulton, M. Kenney, A. Alù, I. Staude, A. Bashiri, Z. Fedorova, R. Kolkowski, A. F. Koenderink, X. Xiao, J. Yang, W. J. Peveler, A. W. Clark, G. Perrakis, A. C. Tasolamprou, M. Kafesaki, A. Zaleska, W. Dickson, D. Richards, A. Zayats, H. Ren, Y. Kivshar, S. Maier, X. Chen, M. A. Ansari, Y. Gan, A. Alexeev, T. F. Krauss, A. Di Falco, S. D. Gennaro, T. Santiago-Cruz, I. Brener, M. V. Chekhova, R.-M. Ma, V. V. Vogler-Neuling, H. C. Weigand, ×.-L. Talts, I. Occhiodori, R. Grange, M. Rahmani, L. Xu, S. M. Kamali, E. Arababi, A. Faraon, A. C. Harwood, S. Vezzoli, R. Sapienza, P. Lalanne, A. Dmitriev, C. Rockstuhl, A. Sprafke, K. Vynck, J. Upham, M. Z. Alam, I. De Leon,

R. W. Boyd, W. J. Padilla, J. M. Malof, A. Jana, Z. Yang, R. Colom, Q. Song, P. Genevet, K. Achouri, A. B. Evlyukhin, U. Lemmer, and I. Fernandez-Corbaton, *Roadmap on photonic metasurfaces,* Applied Physics Letters **124**, 260701 (2024). doi:10.1063/5.0204694.

[148] G. Rosolen, L. J. Wong, N. Rivera, B. Maes, M. Soljačić, and I. Kaminer, *Metasurface-based multi-harmonic free-electron light source,* Light: Science & Applications **7**, 64 (2018). doi:10.1038/s41377-018-0065-2.

[149] T. van Loon, M. Liang, T. Delplace, B. Maes, S. Murai, P. Zijlstra, and J. G. Rivas, *Refractive index sensing using quasi-bound states in the continuum in silicon metasurfaces,* Opt. Express **32**, 14289 (2024). doi:10.1364/OE.514787.

[150] M. G. Abebe, A. De Corte, and B. Maes, *Thermal Plasmonics and Metamaterials for a Low-Carbon Society* (CRC Press, 2024). ISBN 9781003409090. doi:https://doi.org/10.1201/9781003409090.

[151] A. Krasnok, M. Tymchenko, and A. Alù, *Nonlinear metasurfaces: a paradigm shift in nonlinear optics,* Materials Today **21**, 8 (2018). doi:10.1016/j.mattod.2017.06.007.

[152] K. P. Kalinin, J. Gladrow, J. Chu, J. H. Clegg, D. Cletheroe, D. J. Kelly, B. Rahmani, G. Brennan, B. Canakci, F. Falck, M. Hansen, J. Kleewein, H. Kremer, G. O'Shea, L. Pickup, S. Rajmohan, A. Rowstron, V. Ruhle, L. Braine, S. Khedekar, N. G. Berloff, C. Gkantsidis, F. Parmigiani, and H. Ballani, *Analog optical computer for AI inference and combinatorial optimization,* Nature **645**, 354 (2025). doi:10.1038/s41586-025-09430-z.

[153] L. Abou-Hamdan, E. Marinov, P. Wiecha, P. del Hougne, T. Wang, and P. Genevet, *Programmable metasurfaces for future photonic artificial intelligence,* Nature Reviews Physics **7**, 331 (2025). doi:10.1038/s42254-025-00831-7.

[154] V. Nikkhah, A. Pirmoradi, F. Ashtiani, B. Edwards, F. Aflatouni, and N. Engheta, *Inverse-designed low-index-contrast structures on a silicon photonics platform for vector–matrix multiplication,* Nature Photonics **18**, 501 (2024). doi:10.1038/s41566-024-01394-2.

[155] F. Zangeneh-Nejad, D. L. Sounas, A. Alù, and R. Fleury, *Analogue computing with metamaterials,* Nature Reviews Materials **6**, 207 (2021). doi:10.1038/s41578-020-00243-2.

[156] S. Raschka, Y. H. Liu, and V. Mirjalili, *Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python* (Packt Publishing Ltd, 2022). ISBN 978-1-80181-931-2.

[157] G. James, D. Witten, T. Hastie, R. Tibshirani, and J. Taylor, *An Introduction to Statistical Learning with Applications in Python* (Cham: Springer, 2023). ISBN 978-3-031-38746-3. doi:10.1007/978-3-031-38747-0.

[158] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Z. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, *PyTorch: An Imperative Style, High-Performance Deep Learning Library.*, in *NeurIPS*, edited by H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. B. Fox, and R. Garnett, 8024, 2019.

[159] *Pytorch*, 2025. https://www.pytorch.org/.

[160] *pytorch tutorial*, 2025. https://docs.pytorch.org/tutorials/.

[161] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia,

R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015. https://www.tensorflow.org/.

[162] *Tensorflow tutorial*, 2025. https://www.tensorflow.org/tutorials.

[163] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *Scikit-learn: Machine Learning in Python*, Journal of Machine Learning Research **12**, 2825 (2011).

[164] F. Chollet *et al.*, *Keras*, https://github.com/fchollet/keras, 2015.

[165] P. R. Wiecha, *Deep learning for nano-photonic materials – The solution to everything!?*, Current Opinion in Solid State and Materials Science **28**, 101129 (2024). doi:https://doi.org/10.1016/j.cossms.2023.101129.

[166] Y. Deng, K. Fan, B. Jin, J. Malof, and W. J. Padilla, *Physics-informed learning in artificial electromagnetic materials*, Applied Physics Reviews **12**, 011331 (2025). doi:10.1063/5.0232675.

[167] G. You, C. Qian, S. Tan, E. Li, and H. Chen, *Driving deep-learning-based metasurface design with Kramers-Kronig relations*, Phys. Rev. Appl. **22**, L041002 (2024). doi:10.1103/PhysRevApplied.22.L041002.

[168] Y. Xu, J.-Q. Yang, K. Fan, S. Wang, J. Wu, C. Zhang, D.-C. Zhan, W. J. Padilla, B. Jin, J. Chen, and P. Wu, *Physics-Informed Inverse Design of Programmable Metasurfaces*, Advanced Science **11**, 2406878 (2024). doi:https://doi.org/10.1002/advs.202406878.

[169] S. W. Kim, I. Kim, J. Lee, and S. Lee, *Knowledge Integration into deep learning in dynamical systems: an overview and taxonomy*, Journal of Mechanical Science and Technology **35**, 1331 (2021). doi:10.1007/s12206-021-0342-5.

[170] L. von Rueden, S. Mayer, K. Beckh, B. Georgiev, S. Giesselbach, R. Heese, B. Kirsch, J. Pfrommer, A. Pick, R. Ramamurthy, M. Walczak, J. Garcke, C. Bauckhage, and J. Schuecker, *Informed Machine Learning – A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems*, IEEE Transactions on Knowledge and Data Engineering **35**, 614 (2023). doi:10.1109/TKDE.2021.3079836.

[171] S. Bulusu, M. Favoni, A. Ipp, D. I. Müller, and D. Schuh, *Generalization capabilities of translationally equivariant neural networks*, Phys. Rev. D **104**, 074504 (2021). doi:10.1103/PhysRevD.104.074504.

[172] V. Biscione and J. Bowers, *Learning Translation Invariance in CNNs*, 2020. https://arxiv.org/abs/2011.11757.

[173] A. Azulay and Y. Weiss, *Why do deep convolutional networks generalize so poorly to small image transformations?*, 2019. https://arxiv.org/abs/1805.12177.

[174] R. Zhang, *Making Convolutional Networks Shift-Invariant Again*, in *Proceedings of the 36th International Conference on Machine Learning*, edited by K. Chaudhuri and R. Salakhutdinov, vol. 97 of *Proceedings of Machine Learning Research*, 7324, PMLR, 2019.

[175] Z. Wang, P. Wang, K. Liu, P. Wang, Y. Fu, C.-T. Lu, C. C. Aggarwal, J. Pei, and Y. Zhou, *A Comprehensive Survey on Data Augmentation*, 2025. https://arxiv.org/abs/2405.09591.

[176] T. Gahlmann and P. Tassin, *Evaluation of machine learning techniques for conditional generative adversarial networks in inverse design*, 2025. https://arxiv.org/abs/2502.11934.

[177] *TorchVision: PyTorch's Computer Vision library*, https://github.com/pytorch/vision, 2016.

[178] M. Weiler, P. Forré, E. Verlinde, and M. Welling, *Equivariant and Coordinate Independent Convolutional Networks* (, 2023). https://maurice-weiler.gitlab.io/cnn_book/EquivariantAndCoordinateIndependentCNNs.pdf.

[179] T. Cohen and M. Welling, *Group Equivariant Convolutional Networks*, in *Proceedings of The 33rd International Conference on Machine Learning*, edited by M. F. Balcan and K. Q. Weinberger, vol. 48 of *Proceedings of Machine Learning Research*, (New York, New York, USA), 2990, PMLR, 2016. https://proceedings.mlr.press/v48/cohenc16.html.

[180] M. Weiler and G. Cesa, *General E(2)-Equivariant Steerable CNNs*, CoRR **abs/1911.08251**, (2019). http://arxiv.org/abs/1911.08251.

[181] G. Cesa, L. Lang, and M. Weiler, *A Program to Build E(N)-Equivariant Steerable CNNs*, in *International Conference on Learning Representations*, 2022. https://openreview.net/forum?id=WE4qe9xlnQw.

[182] *"E(n)-equivariant Steerable CNNs (escnn)"*, 2025. https://github.com/QUVA-Lab/escnn.

[183] F. Casenave, X. Roynard, B. Staber, W. Piat, M. A. Bucci, N. Akkari, A. Kabalan, X. M. V. Nguyen, L. Saverio, R. C. Perez, A. Kalaydjian, S. Fouché, T. Gonon, G. Najjar, E. Menier, M. Nastorg, G. Catalani, and C. Rey, *Physics-Learning AI Datamodel (PLAID) datasets: a collection of physics simulations for machine learning*, 2025. https://arxiv.org/abs/2505.02974.

[184] J. Jiang, R. Lupoiu, E. W. Wang, D. Sell, J. P. Hugonin, P. Lalanne, and J. A. Fan, *MetaNet: a new paradigm for data sharing in photonics research*, Opt. Express **28**, 13670 (2020). doi:10.1364/OE.388378.

[185] T. Ma, M. Tobah, H. Wang, and L. J. Guo, *Benchmarking deep learning-based models on nanophotonic inverse design problems*, Opto-Electronic Science **1**, 210012 (2022). doi:10.29026/oes.2022.210012.

[186] T. Ma, M. Tobah, H. Wang, and L. J. Guo, *Benchmarking deep learning-based models on nanophotonic inverse design problems (Data)*, https://github.com/taigaoma1997/benchmark_in_de, 2022.

[187] R. Mojumder, Z. Guo, N. Li, and M. Swaminathan, *rayid-mojumder/Physics-inspired-ML-dataset-for- Photonic-Waveguide-Design: v1.0.0*, 2025. doi:10.5281/zenodo.15319336.

[188] Y. Deng, J. Dong, S. Ren, O. Khatib, M. Soltani, V. Tarokh, W. Padilla, and J. Malof, *Benchmarking Data-driven Surrogate Simulators for Artificial Electromagnetic Materials*, in *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, edited by J. Vanschoren and S. Yeung, vol. 1, 2021. https://datasets-benchmarks-proceedings.neurips.cc/paper_files/paper/2021/file/8c19f571e251e61cb8dd3612f26d5ecf-Paper-round2.pdf.

[189] T. Zhang, C. Y. Kee, Y. S. Ang, E. Li, and L. K. Ang, *Symmetry Enhanced Network Architecture Search for Complex Metasurface Design*, IEEE Access **10**, 73533 (2022). doi:10.1109/ACCESS.2022.3190419.

[190] J. Chen, W. Zhan, H. Wang, Z. Jia, J. Gan, J. Zhang, J. Qi, T. Chen, L. Huang, M. Chen, L. Li, W. Wang, and D. Zhou, *MetamatBench: Integrating Heterogeneous Data, Computational Tools, and Visual Interface for Metamaterial Discovery,* 2025. https://arxiv.org/abs/2505.20299.

[191] S. An, C. Fowler, B. Zheng, M. Y. Shalaginov, H. Tang, H. Li, L. Zhou, J. Ding, A. M. Agarwal, C. Rivero-Baleine, K. A. Richardson, T. Gu, J. Hu, and H. Zhang, *A Deep Learning Approach for Objective-Driven All-Dielectric Metasurface Design,* ACS Photonics **6**, 3196 (2019). doi:10.1021/acsphotonics.9b00966.

[192] S. An, *A Deep Learning Approach for Objective-Driven All-Dielectric Metasurface Design - dataset,* https://github.com/SensongAn/Meta-atoms-data-sharing, 2021.

[193] J. Hadamard, *Sur les Problemes Aux Derivees Partielles et Leur Signification Physique,* Princeton university bulletin , (1902).

[194] S. Molesky, Z. Lin, A. Piggott, W. Jin, J. Vucković, and A. Rodriguez, *Inverse design in nanophotonics,* Nature Photonics **12**, 659 (2018). doi:10.1038/s41566-018-0246-9.

[195] C. Qian, I. Kaminer, and H. Chen, *A guidance to intelligent metamaterials and metamaterials intelligence,* Nature Communications **16**, (2025). doi:10.1038/s41467-025-56122-3.

[196] O. Khatib, S. Ren, J. Malof, and W. J. Padilla, *Deep Learning the Electromagnetic Properties of Metamaterials—A Comprehensive Review,* Advanced Functional Materials **31**, 2101748 (2021). doi:https://doi.org/10.1002/adfm.202101748.

[197] C. Ma, Z. Wang, H. Zhang, F. Yang, J. Chen, Q. Ren, Y. Ma, and N. Wang, *Inverse design of electromagnetic metamaterials: from iterative to deep learning-based methods,* Journal of Micromechanics and Microengineering **34**, 053001 (2024). doi:10.1088/1361-6439/ad3a72.

[198] G. Cerniauskas, H. Sadia, and P. Alam, *Machine intelligence in metamaterials design: a review,* Oxford Open Materials Science **4**, itae001 (2024). doi:10.1093/oxfmat/itae001.

[199] A. Khaireh-Walieh, D. Langevin, P. Bennet, O. Teytaud, A. Moreau, and P. R. Wiecha, *A newcomer's guide to deep learning for inverse design in nano-photonics,* Nanophotonics **12**, 4387 (2023) [cited 2025-09-10]. doi:doi:10.1515/nanoph-2023-0527.

[200] C. Shirpurkar, J. Zang, K. Y. Yang, D. Carlson, S. P. Yu, E. Lucas, S. V. Pericherla, J. Yang, M. Guidry, D. Lukin, G. H. Ahn, J. Lu, L. Trask, F. Aflatouni, J. Vučković, S. B. Papp, and P. J. Delfyett, *Photonic crystal resonators for inverse-designed multi-dimensional optical interconnects,* Opt. Lett. **47**, 3063 (2022). doi:10.1364/OL.461272.

[201] K. Y. Yang, C. Shirpurkar, A. D. White, J. Zang, L. Chang, F. Ashtiani, M. A. Guidry, D. M. Lukin, S. V. Pericherla, J. Yang, H. Kwon, J. Lu, G. H. Ahn, K. Van Gasse, Y. Jin, S.-P. Yu, T. C. Briles, J. R. Stone, D. R. Carlson, H. Song, K. Zou, H. Zhou, K. Pang, H. Hao, L. Trask, M. Li, A. Netherton, L. Rechtman, J. S. Stone, J. L. Skarda, L. Su, D. Vercruysse, J.-P. W. MacLean, S. Aghaeimeibodi, M.-J. Li, D. A. B. Miller, D. M. Marom, A. E. Willner, J. E. Bowers, S. B. Papp, P. J. Delfyett, F. Aflatouni, and J. Vučković, *Multi-dimensional data transmission using inverse-designed silicon photonics and microcombs,* Nature Communications **13**, 7862 (2022). doi:10.1038/s41467-022-35446-4.

[202] O. Sigmund and K. Maute, *Topology optimization approaches,* Structural and Multidisciplinary Optimization **48**, 1031 (2013). doi:10.1007/s00158-013-0978-6.

[203] J. Jensen and O. Sigmund, *Topology optimization for nano-photonics,* Laser & Photonics Reviews **5**, 308 (2011). doi:https://doi.org/10.1002/lpor.201000014.

[204] R. E. Christiansen and O. Sigmund, *Inverse design in photonics by topology optimization: tutorial,* J. Opt. Soc. Am. B **38**, 496 (2021). doi:10.1364/JOSAB.406048.

[205] A. M. Hammond, A. Oskooi, S. G. Johnson, and S. E. Ralph, *Photonic topology optimization with semiconductor-foundry design-rule constraints,* Opt. Express **29**, 23916 (2021). doi:10.1364/OE.431188.

[206] M. Gendreau and J.-Y. Potvin, *Handbook of Metaheuristics* (Springer Cham, 2018). ISBN 978-3-319-91085-7. doi:10.1007/978-3-319-91086-4.

[207] K. Sörensen, M. Sevaux, and F. Glover, *A History of Metaheuristics*. In R. Martí, P. M. Pardalos, and a. G. Resende, eds., *Handbook of Heuristics* (Cham: Springer Nature Switzerland, 2025). doi:10.1007/978-3-032-00385-0_4.

[208] L. Shen, Z. Ye, and S. He, *Design of two-dimensional photonic crystals with large absolute band gaps using a genetic algorithm,* Phys. Rev. B **68**, 035109 (2003). doi:10.1103/PhysRevB.68.035109.

[209] F. A. A. Nugroho, P. Bai, I. Darmadi, G. W. Castellanos, J. Fritzsche, C. Langhammer, J. Gómez Rivas, and A. Baldi, *Inverse designed plasmonic metasurface with parts per billion optical hydrogen detection,* Nature Communications **13**, 5737 (2022). doi:10.1038/s41467-022-33466-8.

[210] E. Lucas, S.-P. Yu, T. C. Briles, D. R. Carlson, and S. B. Papp, *Tailoring microcombs with inverse-designed, meta-dispersion microresonators,* Nature Photonics **17**, 943 (2023). doi:10.1038/s41566-023-01252-7.

[211] D. Liu, Y. Tan, E. Khoram, and Z. Yu, *Training Deep Neural Networks for the Inverse Design of Nanophotonic Structures,* ACS Photonics **5**, 1365 (2018). doi:10.1021/acsphotonics.7b01377.

[212] D. P. Kingma and M. Welling, *Auto-Encoding Variational Bayes,* 2022. https://arxiv.org/abs/1312.6114.

[213] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative adversarial nets,* in *Advances in neural information processing systems*, 2672, 2014.

[214] M. Mirza and S. Osindero, *Conditional Generative Adversarial Nets,* 2014. https://arxiv.org/abs/1411.1784.

[215] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, *Score-Based Generative Modeling through Stochastic Differential Equations,* CoRR **abs/2011.13456**, (2020). https://arxiv.org/abs/2011.13456.

[216] Y. Kossale, M. Airaj, and A. Darouichi, *Mode Collapse in Generative Adversarial Networks: An Overview,* in *2022 8th International Conference on Optimization and Applications (ICOA)*, 1, 2022. doi:10.1109/ICOA55659.2022.9934291.

[217] D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, J. Müller, J. Penna, and R. Rombach, *SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis,* 2023. https://arxiv.org/abs/2307.01952.

[218] T. Brooks, B. Peebles, C. Holmes, W. DePue, Y. Guo, L. Jing, D. Schnurr, J. Taylor, T. Luhman, E. Luhman, C. Ng, R. Wang, and A. Ramesh, *Video generation models as world simulators,* 2024. https://openai.com/research/video-generation-models-as-world-simulators.

[219] D. H. Ballard, *Modular learning in neural networks,* in *Proceedings of the Sixth National Conference on Artificial Intelligence - Volume 1*, AAAI'87, 279–284, AAAI Press, 1987.

[220] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning internal representations by error propagation*. In , *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations* (Cambridge, MA, USA: MIT Press, 1986).

[221] Y. Lecun and F. Soulie Fogelman, *Modeles connexionnistes de l'apprentissage,* Intellectica, special issue apprentissage et machine **2**, (1987). doi:10.3406/intel.1987.1804.

[222] P. Gallinari, Y. Lecun, S. Thiria, and F. Soulie Fogelman, *Mémoires associatives distribuées: une comparaison (distributed associative memories: a comparison)*, 1987.

[223] H. Bourlard and Y. Kamp, *Auto-association by multilayer perceptrons and singular value decomposition,* Biological Cybernetics **59**, 291 (1988). doi:10.1007/BF00332918.

[224] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016).

[225] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, *Extracting and composing robust features with denoising autoencoders,* in *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, (New York, NY, USA), 1096–1103, Association for Computing Machinery, 2008. doi:10.1145/1390156.1390294.

[226] G. E. Hinton and R. R. Salakhutdinov, *Reducing the Dimensionality of Data with Neural Networks,* Science **313**, 504 (2006). doi:10.1126/science.1127647.

[227] K. Pearson, *LIII. On lines and planes of closest fit to systems of points in space,* 1901. doi:10.1080/14786440109462720.

[228] S. Kullback and R. A. Leibler, *On Information and Sufficiency,* The Annals of Mathematical Statistics **22**, 79 (1951). doi:10.1214/aoms/1177729694.

[229] K. P. Murphy, *Probabilistic Machine Learning: Advanced Topics* (MIT Press, 2023). http://probml.github.io/book2.

[230] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, *beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework,* in *International Conference on Learning Representations*, 2017. https://openreview.net/forum?id=Sy2fzU9gl.

[231] C. P. Burgess, I. Higgins, A. Pal, L. Matthey, N. Watters, G. Desjardins, and A. Lerchner, *Understanding disentangling in β-VAE,* 2018. https://arxiv.org/abs/1804.03599.

[232] D. J. Rezende, S. Mohamed, and D. Wierstra, *Stochastic Backpropagation and Approximate Inference in Deep Generative Models,* in *Proceedings of the 31st International Conference on Machine Learning*, edited by E. P. Xing and T. Jebara, vol. 32 of *Proceedings of Machine Learning Research*, (Bejing, China), 1278, PMLR, 2014. https://proceedings.mlr.press/v32/rezende14.html.

[233] W. Ma, F. Cheng, Y. Xu, Q. Wen, and Y. Liu, *Probabilistic Representation and Inverse Design of Metamaterials Based on a Deep Generative Model with Semi-Supervised Learning Strategy,* Advanced Materials **31**, 1901111 (2019). doi:https://doi.org/10.1002/adma.201901111.

[234] K. Sohn, H. Lee, and X. Yan, *Learning Structured Output Representation using Deep Conditional Generative Models*, in *Neural Information Processing Systems*, 2015. https://api.semanticscholar.org/CorpusID:13936837.

[235] D. Lee, Y.-C. Chan, W. W. Chen, L. Wang, A. van Beek, and W. Chen, *t-METASET: Task-Aware Acquisition of Metamaterial Datasets Through Diversity-Based Active Learning*, Journal of Mechanical Design **145**, 031704 (2022). doi:10.1115/1.4055925.

[236] X. Ge, R. T. Goodwin, H. Yu, P. Romero, O. Abdelrahman, A. Sudhalkar, J. Kusuma, R. Cialdella, N. Garg, and L. R. Varshney, *Accelerated Design and Deployment of Low-Carbon Concrete for Data Centers*, 2022. https://arxiv.org/abs/2204.05397.

[237] S. Barmada, P. D. Barba, N. Fontana, M. E. Mognaschi, and M. Tucci, *Electromagnetic Field Reconstruction and Source Identification Using Conditional Variational Autoencoder and CNN*, IEEE Journal on Multiscale and Multiphysics Computational Techniques **8**, 322 (2023). doi:10.1109/JMMCT.2023.3304709.

[238] C. Xiao, N. Wu, Z. Hong, X. Wu, H. Chen, and C. Qian, *Intelligent Adaptive Broadband Metasurface with Multilayer Generative Network*, Advanced Theory and Simulations , e01063 (2025). doi:https://doi.org/10.1002/adts.202501063.

[239] G. Li, L. Tang, V. Sorokin, and S. Wang, *CVAE-based inverse design of two-dimensional honeycomb pentamode metastructure for acoustic cloaking*, Thin-Walled Structures **206**, 112623 (2025). doi:https://doi.org/10.1016/j.tws.2024.112623.

[240] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, *et al.*, *Improving language understanding by generative pre-training*, 2018.

[241] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, *Attention is all you need*, in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, (Red Hook, NY, USA), 6000–6010, Curran Associates Inc., 2017.

[242] S. Hochreiter and J. Schmidhuber, *Long Short-Term Memory*, Neural Computation **9**, 1735 (1997). doi:10.1162/neco.1997.9.8.1735.

[243] D. Lu, Y. Deng, J. M. Malof, and W. J. Padilla, *Learning Electromagnetic Metamaterial Physics With ChatGPT*, IEEE Access **13**, 51513 (2025). doi:10.1109/ACCESS.2025.3552418.

[244] D. Lu, J. M. Malof, and W. J. Padilla, *An Agentic Framework for Autonomous Metamaterial Modeling and Inverse Design*, 2025. https://arxiv.org/abs/2506.06935.

[245] R. Lupoiu, Y. Shao, T. Dai, C. Mao, K. Edée, and J. A. Fan, *A multi-agentic framework for real-time, autonomous freeform metasurface design*, Science Advances **11**, eadx8006 (2025). doi:10.1126/sciadv.adx8006.

[246] M. Hasenjäger and H. Ritter, *Active Learning in Neural Networks*. In L. C. Jain and J. Kacprzyk, eds., *New Learning Paradigms in Soft Computing* (Heidelberg: Physica-Verlag HD, 2002). doi:10.1007/978-3-7908-1803-1_5.

[247] P. Belcak, G. Heinrich, S. Diao, Y. Fu, X. Dong, S. Muralidharan, Y. C. Lin, and P. Molchanov, *Small Language Models are the Future of Agentic AI*, 2025. doi:10.48550/arXiv.2506.02153.

[248] S. E. Miller, *Integrated Optics: An Introduction*, Bell System Technical Journal **48**, 2059 (1969). doi:10.1002/j.1538-7305.1969.tb01165.x.