Investigating the use of Semiconductor Nanowires for Neural Networks

Alfredo Serafini

Thesis presented for the degree of Master in Materials Science

LUND UNIVERSITY
Faculty of Science

Supervised by Anders Mikkelsen and David Winge

Department of Physics
Division of Synchrotron Radiation Research

December 2021

A Perla,

Abstract

In this work, we intend to employ nanowires for the realisation of an artificial neuron which can be used to design a neural network that will guide the new generation of non-von Neumann architectures such as Neuromorphic Computing. Possible applications are strongly related to Neuromorphic architectures and artificial intelligence, such as high-performance computers, robotics hardware and autonomous drones.

Here, we aim to design a device that can perform as a node in an optically communicating neural network. We employ nanowires to form a single artificial neuron by coupling a receiver and a transmitter of optical signals, together with a transistor as an in-between control element. Based on this configuration, we find optimal optical properties for our desired neural network communication. The node-to-node communication relies on the strength of the connection known as weights. The connection strength can be tailored by the position and rotation of the individual nanowires making up the nodes and layers; however, this might set some correlations and limitations on the weights.

We frame our main research questions as a set of hypotheses that we test by numerical experiments using the finite-difference time-domain (FDTD) method. The FDTD method is one of many numerical methods that exist to tackle this type of electromagnetic simulation. Among the valid alternatives are the finite element method and the method of the moment. However, we decided on the FDTD method because it is the one that best suits our needs to perform electromagnetic simulation on nanowires and to study a node-to-node communication.

We split our study and perform separate simulations on the receiver and transmitter. During the simulation of the receiver, we sweep the incoming light angle to study the absorption difference between two distinct depletion regions. The difference in absorption between the two regions yields a potential difference that allows switching on and off the transistor that controls the current through the transmitter. The emitted light from the transmitter is assumed proportional to the current. To study the transmitter, a dipole source was placed in the middle of a nanowire to study and evaluate the ability of the emitter to emit light in a specific direction, namely directivity.
The results from the receiver and transmitter studies are combined into a weight function after connecting the two simulated nanowires that form an artificial neuron. The weight function is our ultimate result from our experiment, which can be employed to design a custom neural network based on already known weights.

Acknowledgements

My most profound thanks go to Professor Anders Mikkelsen and Professor Timm Rainer for introducing and initialising this opportunity. A particular thanks to Doctor David Winge for his constant and kind support along with all our meeting sessions.
My most heartfelt thanks to my family for being so supportive during these years.
Cogito cum grano salis, ergo sum!

Contents

Lists of Abbreviations

ANN Artificial Neural Network.

BC Boundary Condition.

CDA Central Difference Approximation.
CMT Comformal Mesh Technique.

DNN Deep Neural Network.

EM Electromagnetic.

FDTD Finite Difference Time Domain.

GaInP Gallium Indium Phosphide.

HPC High-Performance Computing.

IBP Initial Boundary Problem.
InP Indium Phosphide.
IVP Initial Value Problem.

MLP Multilayer Perceptron.
MOSFET Metal–oxide–semiconductor field-effect transistor.

NN Neural Network.
NW Nanowire.

w.l.o.g without lost of generality.
w.r.t with respect to.
WF Weight Function.
WM Weight Matrix.

1 Introduction

1.1 Motivation

The last century was packed with brilliant minds that boosted our society's knowledge and information through the advent of computers, the internet, and television that constantly influence our life. As an example, one interesting object that seems to propel research efforts in the following decades is an *artificial neuron* (AN) that was idealised and reported for the first time in 1944 by *Pitts and McCulloch* [1]. *Pitts and McCulloch* were the pioneers of the field for what later became known as *Neural Networks*. Later in 1951, *Minsky and Edmonds* inspired by *Pitts and McCulloch* built the first neural network machine at MIT Media Laboratory, named SNARC[1], which was one of the first attempts in the field of Artificial Intelligence [2][3]. These technical advances paved the fundamental building block of a new generation of *non-von Neumann* architectures, e.g. Neuromorphic Computing [4]. The technique has a wide variety of applications, such as High-Performance Computing [5], Robotics hardware and Autonomous Drones [6].

In this work, we put forward and test a device design with the aim that it can be used as a node in neural networks (NNs). Furthermore, a set of nodes should communicate with each other if they are coupled together, and we intend to describe the strength of such a coupling. The neural node component that serves as the fundamental building block of a general neural network (NN) is shown schematically in Fig.1. A single neuron may compute simple logical gates such as `AND`, `OR` and `NOT` similarly to addition and multiplication operations between layers [7]. Therefore, node communications are mainly basic mathematical operations such as vector or matrix operations as $a = \boldsymbol{\omega}^T \mathbf{x} + b$ or $\boldsymbol{a} = \Theta \cdot \mathbf{x} + \mathbf{b}$ in matrix notation. In addition, an activation function, $\phi(a)$, is necessary to map out the output as shown in Fig.1.

In our envisioned realisation of an *artificial neural network* (ANN), the node-to-node weighted connectivity is attained by emitting and receiving overlapping light signals inside a shared 2D waveguide, under the right circumstances. The neural network that we are going to simulate is a *broadcasting network*, in contrast to having individual guiding channels, like fibres, between all the neurons. The strength of the communication or coupling is described by a set of parameters called weights, $\omega_{i,j}$, as exemplified in Fig.1, which can be seen as a weight matrix.

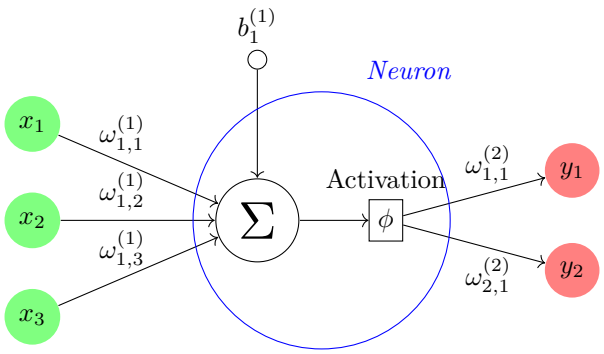


Figure 1. An abstraction of a minimal Neural Network is composed, from the leftmost, of an input layer x_i, at least one neuron, and an output layer y_i, where each neuron has an activation function $\phi(a)$.

The node design consists of a Receiver nanowire (NW) based on photovoltaic nanotechnology [8] such as two distinct photo-diodes labelled by +/− as depicted on the left side of Fig. 2, and a Transmitter or emitter NW containing a nanowire-LED. A key element of the NN communication proposed here is the illumination angle from the leftmost end which influences the amount of absorption between these two different photo-diodes. The absorbed carriers[2] are gathered at the middle of the Receiver NW via a *gold contact pad* and form the gate voltage, V_{gate} or net charge. The gate voltage controls the current through the LED via a device, namely a Metal–oxide–semiconductor field-effect transistor (MOSFET). The MOSFET behaves as a physical activation function that maps the built-up potential on the V_{gate} contact to a current through the nanowire-LED. The LED produces the radiation pattern, and it is set in the middle of the emitter NW as far as possible from the gold contact pads. The emitted power is

[1] Stochastic neural analogue reinforcement calculator (SNARC).

[2] One of the carriers accumulated on the V_{gate} contact, e.g. holes or electrons may rule over the other one, leading to a net charge that switch on/off the transistor and control the current flow from the circuit to the LED in the Transmitter as shown in Fig. 2.

transmitted towards both ends of the NW. However, it is possible to limit the transmission to one side of a Transmitter NW by placing a *Bragg reflector* at one side that acts as a mirror [9]. What makes these nanoscale devices so desirable for both research and application fields is the limited footprint, an improving conversion efficiency [8] and last, the capability to absorb a broad variety of wavelengths, λ, although its nanoscale size[3]. So, we are interested in leveraging the superior speed, and energy efficiency of light as a carrier of information [10].

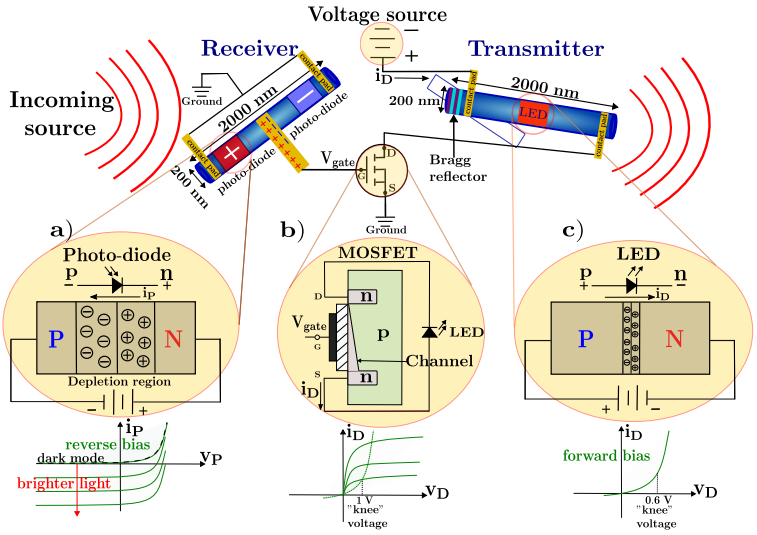


Figure 2. A single physical artificial neuron composed, from the leftmost, by a Receiver NW with two distinct photo-diodes +/−, a transistor, and a Transmitter NW with an LED in the middle. (a) Photo-diode p-n junction labelled with + or −, detects the incoming source **(b) The activation function** seen in Fig. 1 is replaced with a transistor current-voltage characteristic [11]. The activation function maps the built-up potential or net charge on the V_{gate} contact to a current source through **(c) the Transmitter LED.**

Fig. 3 represents the fundamental building blocks of a simple NN, e.g. an input layer, a hidden layer and an output layer, accordingly. The input layer, performed by a Transmitter, generates the incoming light source, while the hidden layer is a physical artificial neuron obtained via an electrical coupling of a Receiver and a Transmitter through a transistor and *gold contact pads* as also shown in Fig. 2 and last, a Receiver as an output layer.

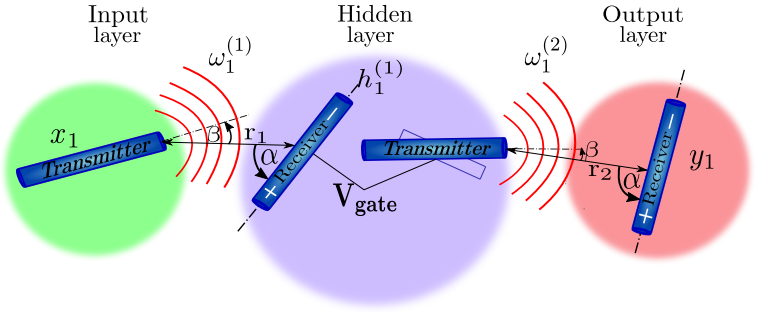
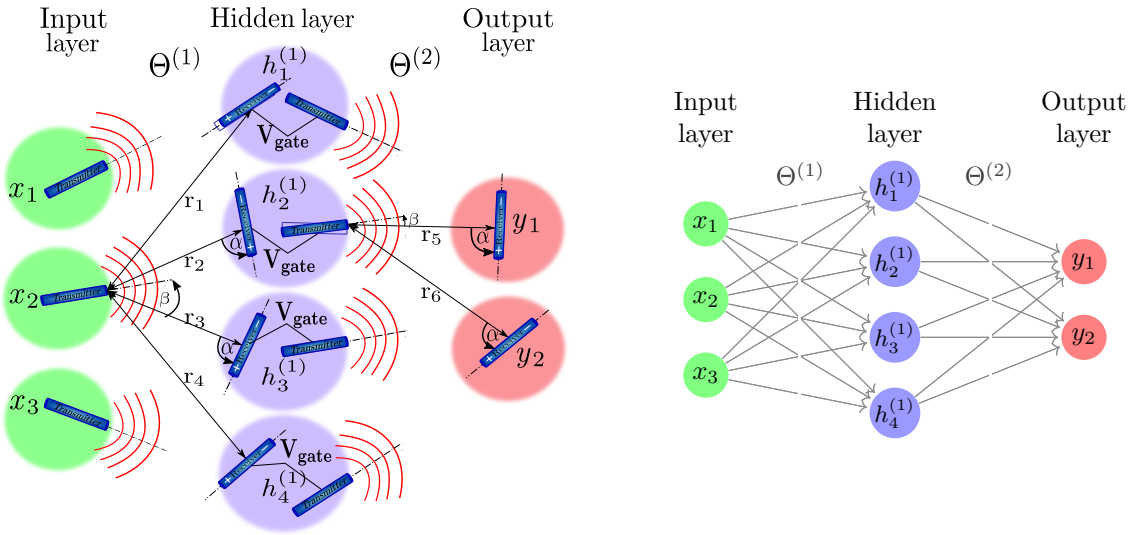


Figure 3. Sketch of a basic physical Neural Network framework formed by an input layer x_1, a single hidden neuron $h_1^{(1)}$ and an output layer y_1, respectively. Transmitter and Receiver NWs rotate anticlockwise w.r.t to the baseline r_1 or r_2, which is a virtual line connecting a Transmitter end-facet transmission point to the centre of a Receiver NW. For $\alpha = \beta = 0°$, both Receiver and Transmitter NWs are aligned with the baselines, r_1, r_2.

Figure 4(a) depicts the interrelation between nanowires through a set of parameters, e.g. (r, α, β) where r is the distance between Transmitter and Receiver, α the Receiver rotation angle and β the Transmitter rotation angle, respectively. Figure 4(b) is a FCNN[4] where an example input node x_2 is Fully-Connected to the next hidden layer similar to the emitter case of Fig. 4(a). The weight values, Θ, represent the strength of the connection between nodes. In short, the below images depict how one can tune the

[3]The ratio of simulated cross-section versus geometrical cross-section shown in Fig. 24, is an appealing factor of NWs.
[4]The Transmitter-Receiver correlation can be seen as a 2D graphic representation of a fully-connected neural network.

strength of optical communication by simply rotating a nanowire or a set of NWs. This is one of the main aspects of this network design strategy. The key point is how the rotation of one nanowire influences not only one weight but an entire row or column of a weight matrix for a network with more than two neurons.



(a) Framework of Neural Network communication.

(b) A Fully-Connected 3-4-2 Neural Network.

Figure 4. (a) Schematic representation of an optical communication between Transmitter and Receiver NWs between layers in an extended NN compare to Fig. 3. **(b) An abstraction of the previous physical NN** such as a fully-connected neural network (FCNN) with 3-4-2 architecture.

In conclusion, the nodes shown in Fig. 4 are made up of nanowires, and the connection strength ω can be tailored by the position and rotation of NWs; this will, however, set some correlations and limitations on the weights. One can view these NWs as a part of a circuit board that has been placed down and contacted, as exemplified in Fig 2. We cannot move them; thus, the weights are fixed. Consequently, the weights are set and cannot be trained *in-loco*. However, the training part of these weights is done externally on *Supercomputers*. Therefore, one needs to implement a NN model that is solid and reliable such as a pre-trained model, e.g. the *AlexNet* [12] to extract these weights. Given these pre-trained weights, one can eventually build a mathematical model that describes the position and rotation of NWs in a NN as illustrated in Fig 4(a).

Exploiting nanowire features under the right circumstances, we want to prove three main *hypotheses* that are important if the types of networks that we propose here are to be realised:

1. Can the absorption of a receiver NW be so specific to suit our neural network requirements, even though, nanowires have tiny cross-section?
2. Can the emission of the transmitter nanowire also be tuned to control the spread of light through the neural network?
3. If a network has more than two neurons, then the transmitter rotations affect not only one weight but rather an entire row or column of a layer-to-layer weight matrix as depicted in Fig 4(a). Can such a network be useful nevertheless?

Hypothesis 1) We assume that we can have more local absorption in one of the two end depletion regions for a certain illumination angle. After selecting an appropriate semiconductor composed of indium and phosphorus, namely, indium phosphide (InP), we can tune the absorption by sweeping the illumination angle, which means that we can adjust the weights and the strength of communication between nodes or layers of a neural network.
Hypothesis 2) We will investigate if it is possible to adjust the Transmitter parameters to direct and control the light cone emission through a neural network.
Hypothesis 3) We will try to develop an algorithm that allows our proposal to mimic a specific neural network to test this hypothesis.

1.2 Aim

This thesis aims to study the limitations of a physical neural network based on broadcasting optical communication. Such a scheme allows all transmitted signals to share a common waveguide, limiting the footprint of such a computational network. The aim in designing this network is that the weights are controlled by geometry, like the relative angle and distance between two components in the network. The shared waveguide approach introduces correlations between different weights. For example, if one component is rotated to tilt the transmitted radiation pattern, this will affect the connections to all components in the following layer. We intend to study if it is possible to perform interesting computations under these limitations.

By performing electromagnetic (EM) simulation within the finite-difference time-domain (FDTD) using a software package called Lumerical, we are aiming to test hypotheses 1) and 2) and by analyzing those results estimate whether hypothesis 3) can be fulfilled.
We will try to investigate if the absorption of the *Receiver* NW can be drastically different between the two ends of the NW by simulating certain illumination angles, wavelengths and NW diameters, for instance. For the *Transmitter*, after selecting a proper power source, we will study different communication wavelengths and NW diameters to observe how the Directivity, which is the ability of NWs to direct light, changes and how it can be controlled. After coupling electrically *Receiver* and *Transmitter* to form a neuron, we aim to describe the *Weight function* used to calculate the weights that we change by rotating nanowires of our neural network, which is our end result.

1.3 Scope

The devices making up the neural network are composed of advanced nanowire devices with electrical properties, like transistors and diodes. In this thesis, we are only interested in their optical properties. For example, we study how they absorb and transmit light as a function of their relative positions and rotations. Thus, we will limit ourselves to single nanowire components without considering the necessary electronics, i.e. the contact pads and wires as shown in Fig. 2. In addition, we will perform 3D simulations in free space to map out the optical characteristics of the nanowire devices, which is a simplification compared to simulating them in a waveguide environment. We believe that this is the correct starting point to study the physics of the devices first, while the waveguide physics could enter at a later stage in order not to obstruct our view.

1.4 Outline

After providing the Motivation above and describing the main research questions, we introduce the Theory chapter which is dedicated to cast our simulation problem onto the form of an initial value problem (IVP). Entering this problem are Maxwell's equations accompanied by examples of their conditions at material interfaces, problem geometry, initial conditions, and boundary conditions (BC). As an aside, we provide a useful definition of the Directivity.

The Method chapter is about solving the initial value problem specified in the Theory. First, the principle of the FDTD method is given in the style of Yee's FDTD technique which is described and implemented with a few algorithm examples with periodic BC, non-reflecting BC and propagation in a medium in 1D, to get a basic idea of the method. Expanding on this Yee's FDTD method, specific Lumerical features are introduced to show how Lumerical treats the boundaries and minimises the numerical error via perfect matching layers (PML). The final part of the Method is dedicated to Sources, Monitors and Limitations to complete the FDTD method as implemented in Lumerical. After explaining the physics in the Theory chapter and how the methods are applied, the outcome of these investigations are presented in the Results and Discussion chapter to consolidate our *hypotheses 1), 2), and 3).* The Conclusion chapter is devoted to the main findings and discusses possible applications and further development of this project.

2 Theory

This chapter explains briefly the theoretical principles that govern the physical phenomena that interests us. We will describe the physical situation of NWs absorbing and emitting light as an *initial value problem* (IVP). IVPs frequently occur in applications, hence, the numerical solution of this kind of problem is a central task in many simulation environments for mechanical, electrical, chemical systems. IVP is a type of problem involving an initial condition, a set of equations and boundary conditions. The solution of an IVP is an unknown function, y, that requires the use of an ordinary differential equation (ODE) with one independent variable or partial differential equation (PDE) in the case of more than one independent variable plus an initial condition. Let $y \in C^1[t_0, t_e]$ be a function that is a solution of the following *initial value problem*:

$$y' = f(t, y(t)) \quad \text{with}\, y(t_0) = y_0 \,\text{as initial condition} \tag{1}$$

Similarly, in higher dimensions, the differential equations are a class of $y_i' = f_i(t, y_1(t), \ldots, y_n(t))$ equations and $y(t)$ is a vector function. At higher orders, the IVP derivatives are considered as independent function, e.g. $y''(t) = g(t, y(t), y'(t))$ [13]. Notably, the boundary conditions are used to fully specify the problem on a certain physical domain.

This Theory chapter is organised according to the key components of the IVP and introduces, first, the set of Maxwell's equations in linear matter, second, how the geometry is specified within the problem space, third, a set of relevant initial conditions including plane waves and dipoles, and fourth, boundary conditions. As an aside, in the end, we add the definition of Directivity with examples of isotropic emission and emission from a Hertzian dipole.

2.1 Maxwell's equations in matter

Here, Maxwell's equations are shown in regions of space where there are charge distribution ρ and current density $\mathbf{J}$. Note, in linear material, the free current density in Maxwell's equations is generally $\mathbf{J} \neq 0$.

$$\begin{aligned} &i) \quad \boldsymbol{\nabla} \cdot \mathbf{D} = \frac{\rho}{\epsilon} \quad \text{Gauss's law for}\, \mathbf{E}, \qquad iii) \quad \boldsymbol{\nabla} \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \text{Faraday's law}, \\ &ii) \quad \boldsymbol{\nabla} \cdot \mathbf{B} = 0 \quad \text{Gauss's law for}\, \mathbf{B}, \qquad iv) \quad \boldsymbol{\nabla} \times \mathbf{B} = \mu\mathbf{J} + \mu\epsilon\frac{\partial \mathbf{D}}{\partial t} \quad \text{Ampère's law}, \end{aligned} \tag{2}$$

$$\text{where} \quad \mathbf{D} = \epsilon\mathbf{E}, \quad \mathbf{H} = \frac{1}{\mu}\mathbf{B} \,\text{and}\, \mathbf{J} = \sigma\mathbf{E}$$

These equations are still valid in a vacuum with no sources; under such circumstances, the charge distribution ρ and the current $\mathbf{J}$ disappear; while, the permeability and permittivity are the ones in free space, e.g. μ_0, ϵ_0, respectively [14][15].

2.1.1 Conditions at material interfaces

The following equations relate the electric and magnetic fields to the left and right of the interface between two linear media when a wave passes through two distinguished materials, e.g. air to semiconductor or fibreglass to air. The details of the following equations are related to the conservation laws such as energy and charge conservation, and specify the *internal boundary conditions* at material interfaces [14]:

$$\begin{aligned} &i) \quad \epsilon_1 E_1^{\perp} = \epsilon_2 E_2^{\perp}, \qquad iii) \quad \mathbf{E}_1^{\parallel} = \mathbf{E}_2^{\parallel} \\ &ii) \quad B_1^{\perp} = B_2^{\perp}, \qquad iv) \quad \frac{1}{\mu_1}\mathbf{B}_1^{\parallel} = \frac{1}{\mu_2}\mathbf{B}_2^{\parallel} \end{aligned} \tag{3}$$

where the fields are split into $\mathbf{E}^{\perp}$, $\mathbf{E}^{\parallel}$, the orthogonal and parallel components, respectively, same for $\mathbf{B}^{\perp}$, $\mathbf{B}^{\parallel}$ with respect to the interface, as depicted in Fig. 5. Note that $\epsilon_1 \neq \epsilon_2$.

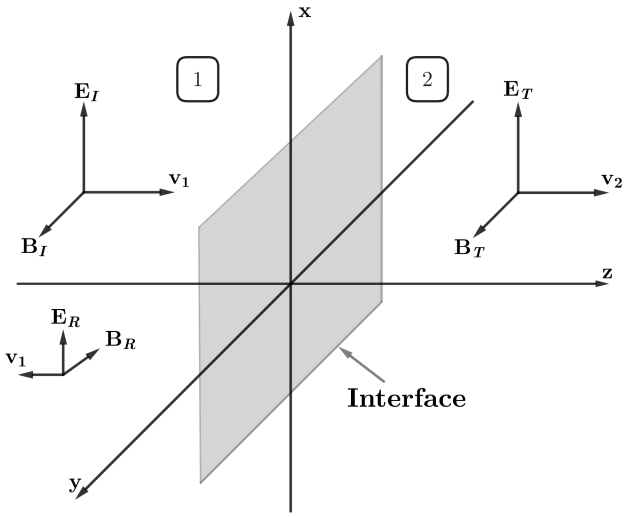


Figure 5. Sketch of a material interface where interior boundary conditions of Eq. (3) should be fulfilled.

2.2 Geometry within the problem space

In a linear material, electromagnetic waves propagate at speed:

$$v = \frac{1}{\sqrt{\epsilon\mu}} = \frac{c}{n} \tag{4}$$

$$n \equiv \sqrt{\frac{\epsilon\mu}{\epsilon_0\mu_0}} \cong \sqrt{\epsilon_r} \quad \mu \text{ close to } \mu_0 \text{ for most of media} \tag{5}$$

where n is the index of refraction of a medium, see Appendix **A.2**. ϵ_r is the dielectric constant[5] that is often greater than 1; hence, incoming light source propagates more slowly in a material. Notably, the dielectric constant, ϵ_r enters into Gauss's law for $\mathbf{E}$ and Ampere's law, and this is how Maxwell's equations know that there is a device in the simulation environment due to the refractive index variation. An important point here is that the refractive index as a function of position, n($\mathbf{r}$, ω), is an relevant function to specify in the system of equations; this is one of the prerequisites of the problem, just as the initial and boundary conditions.

In idealized conditions, the refractive index, n, is sometimes assumed real and the material do not absorb any light. However, real materials absorb part of incoming light source, which effectively can be described by a complex refractive index:

$$\tilde{n} = n + i\kappa, \tag{6}$$

where n is the real part and the imaginary part is the extinction coefficient, κ.

Inserting the $\tilde{n}$ into the wave vector, $\tilde{k} = \left(\frac{\omega}{c}\right)\tilde{n} = \left(\frac{\omega}{c}\right)[n + i\kappa]$ of the plane wave, and rearranging the plane wave equation, we find

$$\begin{aligned}\mathbf{E}(z,t) &= \mathbf{E}_0 \cdot e^{i(\tilde{k}z - \omega t)}, \\ &= \mathbf{E}_0 \cdot e^{-\alpha\frac{z}{2}} e^{i(kz-\omega t)}\end{aligned} \tag{7}$$

where k in the second line of above equation is real and $e^{-\alpha\frac{z}{2}}$ is the exponential damping factor which is related to the absorption coefficient $\alpha = \frac{2\omega\kappa}{c} = \frac{4\pi\kappa}{\lambda_0}$ [16].

An example of values is the refractive index, $n = 3.5896$, $\lambda_0 = 0.5876\,\mu\text{m}$, extinction coefficient, $\kappa = 0.35045$ and $\alpha = 74\,947\,\text{cm}^{-1}$ for InP compound material[6].

2.3 Initial conditions, Sources

A source, such as a plane wave, injects electromagnetic radiation that subsequently impinges a device or material under study. In this study we use sources that inject electromagnetic energy during a short time period and then follow the propagation of the created waves through the system of interest.

[5] Dielectric constant is independent of the amplitude of $\mathbf{E}$; however, it depends on the frequency.
[6] These values are taken from the Refractive index database.

2.3.1 Plane waves

In the case of plane wave propagating in z-direction in free space with no sources, the Maxwell's equations simplify to:

$$\nabla \cdot \mathbf{E} = 0 \implies \frac{\partial E_x(z,t)}{\partial x} = 0 \qquad \text{Gauss's law for } \mathbf{E},$$

$$\nabla \cdot \mathbf{H} = 0 \implies \frac{\partial H_y(z,t)}{\partial y} = 0 \qquad \text{Gauss's law for } \mathbf{H},$$

$$\frac{\partial \mathbf{E}}{\partial t} = \frac{1}{\epsilon_0}\nabla \times \mathbf{H} \implies \frac{\partial E_x}{\partial t} = -\frac{1}{\epsilon_0}\frac{\partial H_y(z,t)}{\partial z} \qquad \text{Ampère's law,} \tag{8}$$

$$\frac{\partial \mathbf{H}}{\partial t} = -\frac{1}{\mu_0}\nabla \times \mathbf{E} \implies \frac{\partial H_y}{\partial t} = -\frac{1}{\mu_0}\frac{\partial E_x(z,t)}{\partial z} \qquad \text{Faraday's law,} \tag{9}$$

where both $\mathbf{E}$ and $\mathbf{H}$ are vectors in 3D that represents three equations each, in general. The plane waves are solutions to these equations and are given as

$$\mathbf{E} = \mathbf{E}_0 \cdot e^{i(kz-\omega t)}, \quad \mathbf{B} = \mathbf{B}_0 \cdot e^{i(kz-\omega t)}, \tag{10}$$

and such waves can be used as initial conditions to the IVP. If the physical fields, i.e. the real components of $\mathbf{E}_0$ and $\mathbf{B}_0$ are considered in place of the complex amplitudes, respectively, along the x- and y-axis, the following linearly polarised wave is a a possible solution:

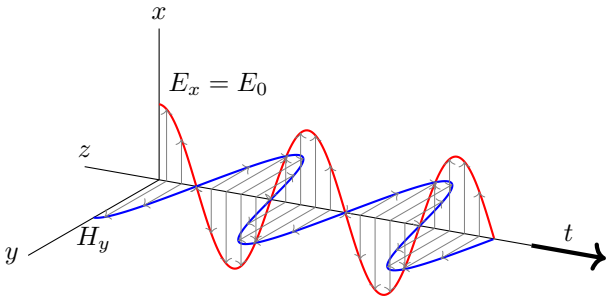
$$\mathbf{E}(z,t) = E_0 \cos(kz - \omega t + \delta)\hat{\mathbf{x}}, \quad \mathbf{B}(z,t) = B_0 \cos(kz - \omega t + \delta)\hat{\mathbf{y}}, \tag{11}$$

Note that $\mathbf{E}$ and $\mathbf{B}$[7] are mutually orthogonal, with real amplitude related by $B_0 = \frac{k}{\omega}E_0 = \frac{1}{c}E_0$. Additional constraints from the Maxwell's equations on $\mathbf{E_0}$ and $\mathbf{B_0}$, since $\nabla \cdot \mathbf{E} = 0$ and $\nabla \cdot \mathbf{H} = 0$, require that

$$(E_0)_z = (B_0)_z = 0$$

which proves, *electromagnetic waves are transverse* [14][16], such that wave propagation is orthogonal to the electric and magnetic fields.

Figure 6. Transverse electromagnetic wave (TEM)



$$\mathbf{B} = \mu_0 \mathbf{H}$$

E_x = electric field amplitude
H_y = magnetic field amplitude
n = index of refraction

$$v = \frac{1}{\sqrt{\epsilon\mu}} = \frac{c}{n}, \quad c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}$$

μ_0 = magnetic permeability in a vacuum,
ε_0 = electric permittivity in a vacuum,
c = speed of light (3×10^8m/s)

Fig. 6 depicts the vector nature of the fields of a wave propagating in three dimensions along the z-axis due to set initial conditions. $\mathbf{E}(z,t)$ field oscillates in the x direction and $\mathbf{H}(z,t)$ in the y direction, while the power source flows along z. The right-hand rule defines the direction of the power flow by the cross-product $\mathbf{E} \times \mathbf{H}$. By summing an ensemble of plane waves, arbitrary shapes in time, such as a Gaussian pulse, $G(t)$, can be built. The Gaussian pulse oscillates w.r.t time, but it is constant in space along its wavefronts, i.e. $\mathbf{E}$ and $\mathbf{H}$.

[7] Every solution to Maxwell's equations, in vacuum, must fulfil the wave equation in Appendix 1, the vice versa is not true.

2.3.2 Energy in Electromagnetic waves, Poynting vector

In relation to the plane waves, we can define another relevant element of electrodynamics: the **Poynting vector**. Here we will provide a brief introduction and discussion. The **energy density** or total energy stored, per unit volume, in electromagnetic fields, **E**, is [14]:

$$u = W_e + W_m = \frac{1}{2}\left(\epsilon E^2 + \frac{1}{\mu}B^2\right) \tag{12}$$

For a **monochromatic** plane wave

$$B^2 = \frac{1}{c^2}E^2 = \mu\epsilon E^2 \tag{13}$$

where is $E = E_0\cos(kz - \omega t + \delta)$, it follows that we have equal electric and magnetic contributions:

$$u = \epsilon E^2 = \epsilon E_0^2\cos^2(kz - wt + \delta) \tag{14}$$

where the cosine argument is the *phase* and k is the wavenumber, and δ is a phase constant.
From the Lorentz force, $\mathbf{F} = q(\mathbf{E} + \mathbf{v}\times\mathbf{B})$, the work done on a charge q is expressed by

$$\mathrm{d}W = \mathbf{F}\cdot\mathrm{d}\mathbf{l} = q(\mathbf{E} + \mathbf{v}\times\mathbf{B})\cdot\mathbf{v}\ \mathrm{d}t = q\mathbf{E}\cdot\mathbf{v}\ \mathrm{d}t, \quad \mathbf{J} = q\mathbf{v}$$

that describes the electromagnetic forces on the charges in a domain V during the time $\mathrm{d}t$. The rate at which work is applied on all the charges in a volume V is

$$\frac{\mathrm{d}W}{\mathrm{d}t} = \int_V (\mathbf{E}\cdot\mathbf{J})\ \mathrm{d}\tau \tag{15}$$

where $(\mathbf{E}\cdot\mathbf{J})$ is the power per unit volume which can be expanded as:

$$(\mathbf{E}\cdot\mathbf{J}) = -\frac{1}{2}\frac{\partial}{\partial t}\left(\epsilon E^2 + \frac{1}{\mu}B^2\right) - \frac{1}{\mu}\boldsymbol{\nabla}\cdot(\mathbf{E}\times\mathbf{B}), \tag{16}$$

obtained by using the Ampère-Maxwell's law to cancel the term **J**, the product rule and the Faraday's law[8]. Inserting the Eq. (16) into Eq. (15) and applying the divergence theorem to the second term, follows the final form of the **Poynting's theorem**[9]:

$$\frac{\mathrm{d}W}{\mathrm{d}t} = -\frac{\mathrm{d}}{\mathrm{d}t}\int_V \frac{1}{2}\left(\epsilon E^2 + \frac{1}{\mu}B^2\right)\mathrm{d}t - \frac{1}{\mu}\oint_S (\mathbf{E}\times\mathbf{B})\cdot\mathrm{d}\mathbf{a}, \tag{17}$$

where S is the surface that bounds V. The first term on the right represents the change in total energy stored in the fields; the second term is the rate at which energy is transported out of V by **E** and **B** through its boundary surface S with $\mathrm{d}\mathbf{a}$ normal to the surface S.

A wave travels and carries this energy flux density **S**, measured as energy per unit time and area, called the **Poynting vector**

$$\boxed{\mathbf{S} \equiv \frac{1}{\mu}\mathbf{E}\times\mathbf{B}}. \tag{18}$$

For a monochromatic plane wave in z direction, $\mathbf{S} = cu\ \vec{\mathbf{z}}$ where cu is the energy per unit time and area, carried by the wave; which can be related to the intensity[10] by making use of Eq. (13):

$$I \equiv \langle S\rangle = \frac{1}{2}\epsilon v E_0^2 \tag{19}$$

All things considered, we are interested in measuring the time-averaged value $\langle S(t)\rangle$ [14]. As the wave travels, it carries this energy along with it, rather than the fluctuating cosine-squared term, Eq. (14). Thereby, the **intensity** is the average power per unit area carried by an electromagnetic wave which is what we work with a monitor when we perform our FDTD simulations.

[8] More detailed derivations on *Griffith*'s 4^{th} edition, Introduction to Electrodynamics [14]

[9] Poynting's theorem says that *the work done on the charges by the electromagnetic force is equal to the decrement in energy left in the fields, minus the energy that flowed out via the surface.*

[10] This holds also in linear media. The Poynting vector itself is always true, in general, regardless the media.

2.3.3 Dipole sources

In this section we will distinguish between far-field and near-field. In the case of the far-field, at large distances such as $r \gg \lambda$, the contribution of a field from many multipoles can be seen as a localised charge distribution or a single charge with potential approximated to $V(r) = \frac{1}{4\pi\epsilon_0}\frac{Q}{r}$ [17]. While, in the near-field, for $r \ll \lambda$, one may consider the contribution of many multipoles such as the dipole, quadrupole and octopole, mainly. The potential $V(\mathbf{r})$ of any localised charge distribution can be expanded in powers of $\frac{1}{r}$ where the law of cosines in Eq. (20) and Taylor expansion are involved, and their coefficients are Legendre polynomials. [14]. Below follows the law of cosine and the **multipole expansion** w.r.t $\frac{1}{r}$ truncated to the 4^{th} order:

$$r^2 = r^2 + (r')^2 - 2rr'\cos\alpha = r^2\left[1 + \left(\frac{r'}{r}\right)^2 - 2\left(\frac{r'}{r}\right)\cos\alpha\right], \tag{20}$$

$$V(\boldsymbol{r}) = \frac{1}{4\pi\epsilon_0}\int \frac{1}{r}\rho(\boldsymbol{r}')\,\mathrm{d}t' = \frac{1}{4\pi\epsilon_0}\left[\frac{1}{r}\int \rho(\boldsymbol{r}')\,\mathrm{d}t' + \frac{1}{r^2}\int r'\cos\alpha\ \rho(\boldsymbol{r}')\,\mathrm{d}t' + \frac{1}{r^3}\int (r')^2\left(\frac{3}{2}\cos^2\alpha - \frac{1}{2}\right)\cos\alpha\ \rho(\boldsymbol{r}')\,\mathrm{d}t' + ...\right] \tag{21}$$

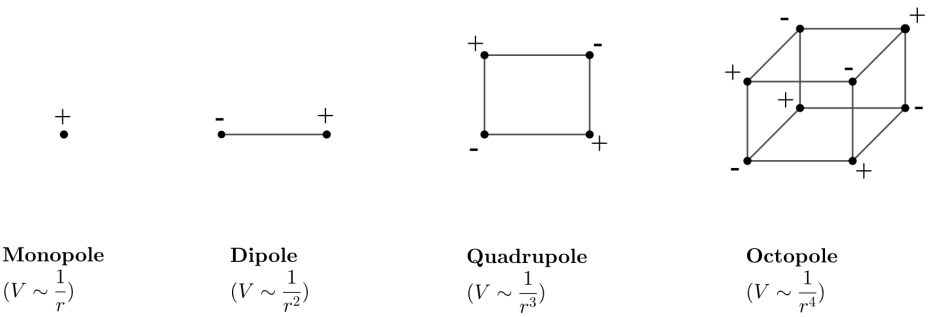


Figure 7. Shows several multipole potentials as function of $\frac{1}{r}$, $\frac{1}{r^2}$, $\frac{1}{r^3}$ and $\frac{1}{r^4}$, respectively.

The above equation, in a far-field environment where the distance $r \gg \lambda$, approaches a monopole potential as follows:

$$V(\boldsymbol{r}) = \frac{1}{4\pi\epsilon_0}\frac{1}{r}\int \rho(\boldsymbol{r}')\,\mathrm{d}t'$$

In the case of dominant time-dependent **dipole potential** where the total charge $Q = \int \rho \mathrm{d}t = 0$ and $\rho(\mathbf{r},t) = \rho(\mathbf{r})e^{i\omega t}$, then $V(\boldsymbol{r},t)$ from Eq.(21) can be written concisely as:

$$V(\mathbf{r},t) = \frac{1}{4\pi\epsilon_0}\frac{\mathbf{p}(t)\cdot\hat{\mathbf{r}}}{r^2} \qquad \text{where} \quad \mathbf{p}(t) = \int \mathbf{r}'\rho(\mathbf{r}',t)\,\mathrm{d}t'$$

where we note that the dipole moment is defined by the oscillating charge density $\rho(\mathbf{r},t)$. The dipole moment $\mathbf{p}(t)$ oscillates in time. Dipole sources can be used as an initial condition in a FDTD simulation.

2.4 Boundary conditions

2.4.1 Free-flow boundary condition

In the context of EM simulations, the perfectly matching layer (PML) was originally proposed by *J.P. Berenger* in 1994 [18]. The PML is a type of boundary that allows **E** and **H** fields to travel through and escape the problem space in order to avoid that radiation reflection may bounce back into the FDTD simulation region. Whereas, when they are applied to the discretised spaces used in the FDTD method, there are often some imperfections due to reflections[11] and instabilities related to the discretisation of the PML equations, see [19]. Nevertheless, PML is widely considered the state-of-art for the termination of

[11] Reflection yields disturbance known as aleatoric noise that is relatively common in signal processing.

FDTD grids, especially when one wants a free-flow boundary condition. In other words, PML is necessary to keep outgoing **E** and **H** fields from being reflected back into the simulation environment [20] as shown in Fig. 8. Though there are several alternatives to this type of boundary conditions as suggested by *J.P. Berenger* [18], we will focus mainly on PML because it is the one implemented in Lumerical.

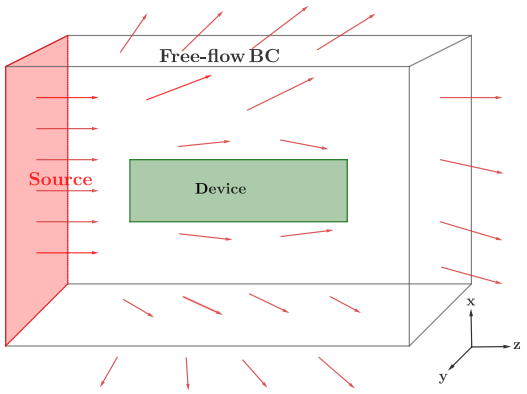


Figure 8. Visualisation of E and H fields propagating outward through a PML boundary condition, avoiding reflection of radiation back into the simulation region.

2.4.2 Periodic boundary condition

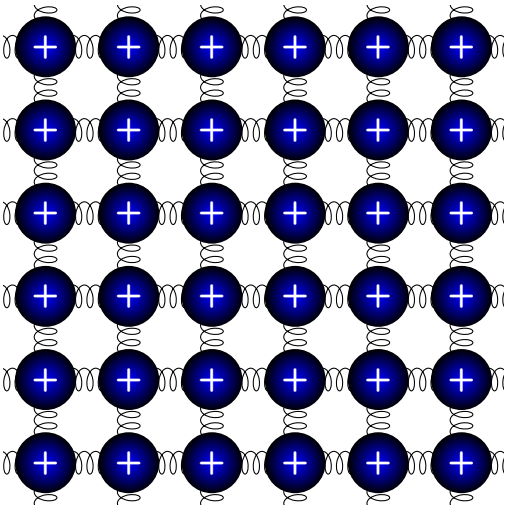
To be able to understand periodic boundary conditions, we describe here the physics of wave oscillations that travel in space and time. First, we set the boundary conditions

$$u(x) = u(x - L), \tag{22}$$

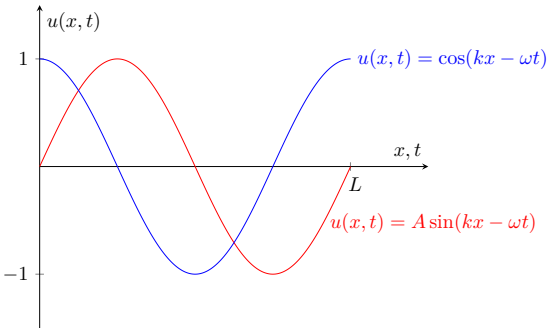
where L is a period length. Then a traveling wave example function as shown in Fig. 9 can be

$$u(x,t) = f(kx - \omega t) = A\sin(kx - \omega t), \tag{23}$$

To this end, by applying the BC from Eq. (22) to the above Eq. (23) gives discrete allowed values for k.



a) Lattice vibration or beaded string in equilibrium.



b) Periodic boundary conditions.

Figure 9. Sketch of two examples of periodic boundary conditions. (a) Shows a periodic BC in a lattice structure. While (b) two examples of periodic BC such as a string.

2.5 Definition of directivity

Directivity is the ability of an antenna to focus radiation along the direction of maximum intensity. This property can be used to benchmark the antennas in our study. Before looking at the directivity, we start by looking at the *directive gain* which is the emission pattern of the antenna as a function of direction, defined as

$$G(\theta)_{2D} = \frac{2\pi U(\theta)}{P_{\text{rad}}} \quad \text{or} \quad G(\theta,\phi)_{3D} = \frac{4\pi U(\theta,\phi)}{P_{\text{rad}}}. \tag{24}$$

in both 2D and 3D, where $U(\theta)$ and $U(\theta,\phi)$ are the angular distributions of radiation intensity per unit solid angle and P_{rad} is the total power. Note, for a homogeneous pattern, i.e. a constant function $U(\theta)$ or $U(\theta,\phi)$, the directive gain in Eq. (30), is also a constant equal to 1.

Formulation of Directivity

The *directivity* is defined as the maximum value of the directive gain, here in 2D:

$$D_{2D} = \frac{2\pi U_{\max}}{P_{\text{rad}}}. \tag{25}$$

where directivity is not a function of θ but a scalar.
Below we present formulas for calculating the directivity in 2D when $U(\theta)$ is known, both when it is known as a continuous function or a discrete set of data points [15]:

$$\text{Continuous 2D: } D = \frac{U(\theta)|_{\max}}{\frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\theta\, U(\theta)} \tag{26}$$

$$\text{Discrete 2D: } \quad D = \max_i \left(\frac{U_i(\theta)}{\frac{1}{2\pi}\sum_{i=0}^{n-1} U_i(\theta)\cdot\Delta\theta} \right), \quad \Delta\theta = \theta_{\mathrm{i}} - \theta_{\mathrm{i-1}} \tag{27}$$

using the transformation $\int \mathrm{d}\theta \to \sum_i \Delta\theta$ with the appropriate variable substitutions. Here we can use

$$\Delta\theta = \frac{\theta_{\max} - \theta_{\min}}{N} = \frac{2\pi - 0}{N}. \tag{28}$$

to eliminate 2π from our expressions, resulting in

$$D = \max_i \left(\frac{U_i(\theta) N}{\sum_{i=0}^{n-1} U_i(\theta)} \right). \tag{29}$$

To exemplify the use of the directivity concept, we provide two examples: isotropic emission with D=1 and the Hertzian dipole emission [15] in 2D where D=2 . These values can be calculated inserting the known function for $U(\theta)$ in each case. First, for an isotropic emitter we have $U(\theta) = 1$, which gives

$$D_{\text{isotropic}} = \frac{U(\theta)|_{\max}}{\frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\theta\, U(\theta)} = \frac{1}{\frac{1}{2\pi} 2\pi} = 1. \tag{30}$$

In contrast, the Hertzian dipole with $U(\theta) = \sin^2(\theta)$ and given $\int_0^{\frac{\pi}{2}} \sin^2\theta\, \mathrm{d}\theta \equiv \int_0^{\frac{\pi}{2}} \cos^2\theta\, \mathrm{d}\theta$[12] follows

$$\begin{aligned} D_{\text{Hertzian}} &= \frac{U(\theta)|_{\max}}{\frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\theta\, U(\theta)} = \\ &= \frac{1}{\frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\theta \cdot \frac{\sin\theta^2 + \cos\theta^2}{2}} = 2 \end{aligned} \tag{31}$$

These two examples can aid in interpreting the results from our modeling.

[12] Alternative option $\longrightarrow$ Double angle formula $U(\theta) = \sin^2\theta \equiv \frac{1}{2}(1 - \cos 2\theta)$ follows same result.

3 Methods

After introducing the fundamental building blocks of the Theory, in this Method chapter, we will describe the FDTD method and the use of Yee's staggered lattice [21] to solve the *initial value problem* (IVP). The chapter is outlined in the following way:

1. A few words of merit is dedicated to the FDTD method on the notes of a specific variation introduced by *Dr Kane S. Yee*.
2. We explain the Discretisation in the simulation space and the Courant stability condition.
3. Algorithm examples are suggested to overview FDTD method in a simple 1D scenario.
4. Geometry and Initial conditions with sources are given.
5. Boundary conditions are illustrated to exhibit how Lumerical treats boundaries to optimise accuracy and computational time.
6. We discuss both general and specific limitations of the FDTD method.

3.1 The FDTD technique

The finite-difference time-domain method (FDTD) is well-known in the context of Numerical Methods and in Computational Fluid Dynamics problems to compute approximate solutions to a related system of differential equations under certain boundary value conditions. FDTD is widely used in the fields of nanotechnology as nanophotonics and spintronics [20] due to its relative easy implementation of first derivatives in Maxwell's equations, where **E** and **H** are intermixed. Another evident reason for its successful use is its ability to cover efficiently a wide broadband per simulation run, while handling various media such as vacuum, dielectrics and metals [20][21].

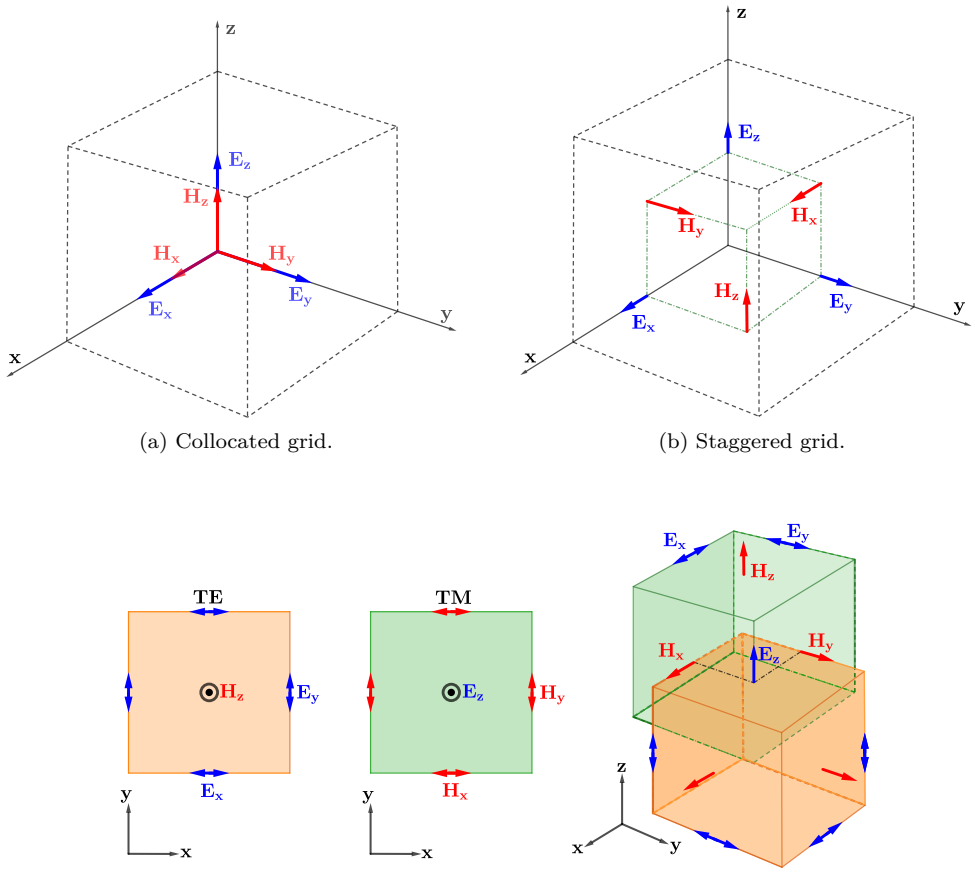
Nevertheless, FDTD was introduced in the field of electromagnetism (EM) in 1966 by *Dr. Kane S. Yee* with a paper on the use of a finite difference staggered grids algorithm in the solution of Maxwell's equations [21]. Yee's staggered lattice leads to an accurate and robust algorithm, compared to the collocated grid, even though it's an extra step of complication.
The basic technique that underlies FDTD is the finite difference method. Using the *central difference approximation* (CDA) to both spatial and temporal partial derivatives, the Ampere's and Faraday's law are discretised. Due its better accuracy, CDA[13] is the one employed by the FDTD method to study EM wave simulation.

3.1.1 Discretisation of the simulation space

Calculating the *curl* of the two fields in 1D, 2D, 3D can be computationally challenging. The FDTD method is a grid-based finite-difference algorithm where the simulation environment is discrete in space and time via collocated grid as shown in Fig. 10(a). However, with the introduction of the Yee's staggered lattice [21], in the $70's$, the Maxwell' equations **E** and **B**-fields are staggered by rectangular unit cells along a cartesian grid. The **E**-field is set midway between a pair of **H**-field points, and the other way around for **H**-field and **E**-field points as shown in Fig. 10(b). By staggering the problem space, the $\mathbf{E}, \mathbf{H}$ grid points are out of phase, which might seems an additional complication w.r.t a collocated grid but at the end it is an effective approach. This former scheme is known as Yee's staggered lattice [21]. This technique is at the core of several FDTD package applications and gives robustness to the algorithm.

The following figures display what the Yee's staggered lattice [21] is; basically, the environment is discretised as $\Delta x, \Delta y$ in 2D and $\Delta x, \Delta y, \Delta z$ in 3D, respectively, to accomplish a computational stability advantage due to its unique configuration.

[13] The CDA is said to have second order accuracy or behaviour compare to the backward & forward approximation [13].



(a) Collocated grid.

(b) Staggered grid.

(c) Staggered grids from 2D and 3D views.

Figure 10. (a) Visualisation of a collocated grid where all the E and H-field components are simultaneously computed at a point for a given incoming wave source. **(b) Staggered grid where the field components are out of phase**, so each grid stores different field components [20][21]. **Figure (c) displays, from the leftmost, the 2D *Yee's lattice* where H-field is set midway between a pair of E-field components and vice versa**. H_z (*H-mode*) and E_z (*E-mode*) components stem from $\nabla \times \mathbf{E}$ and $\nabla \times \mathbf{H}$, accordingly. While in 3D, all three main field components are depicted for a single cubic grid voxel, volumetric picture element. In general, FDTD methods store different field components for different grid locations [20][21].

3.1.2 The simulation algorithm

Finite differences such as *forward, backward and central difference approximation* are popular in numerical methods; however the *central difference approximation* (CDA) is the one employed by the FDTD method to study EM wave simulation due to its smaller *error term*. We apply CDA to the Ampere's law:

$$\frac{\partial E(z,t)}{\partial t} \overset{def}{=} \lim_{\delta \to 0} \frac{E(z, t_0 + \frac{\delta}{2}) - E(z, t_0 - \frac{\delta}{2})}{\delta} \approx \frac{E(z, t_0 + \frac{\delta}{2}) - E(z, t_0 - \frac{\delta}{2})}{\delta}, \tag{32}$$

$$\frac{\partial H(z,t)}{\partial z} \overset{def}{=} \lim_{\delta \to 0} \frac{H(z_0 + \frac{\delta}{2}, t) - H(z_0 - \frac{\delta}{2}, t)}{\delta} \approx \frac{H(z_0 + \frac{\delta}{2}, t) - H(z_0 - \frac{\delta}{2}, t)}{\delta}. \tag{33}$$

Alternatively, one can obtain the Taylor expansion of $E(z, t_0 + \frac{\delta}{2})$ & $E(z, t_0 - \frac{\delta}{2})$, respectively. Then compute the difference between them, all divided by δ to obtain the CDA at t_0:

$$\begin{aligned} \frac{E(z, t_0 + \frac{\delta}{2}) - E(z, t_0 - \frac{\delta}{2})}{\delta} &= E'(z, t_0) + \underbrace{\frac{1}{3!}\frac{\delta^2}{2^2} E'''(z, t_0)}_{error\ \ term} + \ldots \\ &= E'(z, t_0) + \mathscr{O}(\delta^2) \\ \therefore \quad E'(z, t_0) &\approx \frac{E(z, t_0 + \frac{\delta}{2}) - E(z, t_0 - \frac{\delta}{2})}{\delta} \end{aligned} \tag{34}$$

which resembles what we are utilising during our EM simulation, as exemplified below in Eqs. (32-33). The ordo term (big "O"), $\mathcal{O}(\delta^2)$ in Eq. (34), indicates a sum of all the terms in the Taylor expansion that are not shown, starting at δ^2, and the *error term*.

Using the CDA as shown in Eqs. (32-33), one may re-write the Ampere's law and similarly the Faraday's law, see section 2.3.1, with finite differences. In addition, introducing the space-time *Yee*-lattice shown in Fig. 10, leads to an accurate, simple, and robust *algorithm*, which we will outline below, starting with the discretization of the time-space.
Hereafter, we use the following notation to indicate the location where fields are sampled in space & time

$$E_x(z,t) \rightarrow E_x(k\Delta z, n\Delta t) \rightarrow E_x^{k,n}, \qquad H_y(z,t) \rightarrow H_y(k\Delta z, n\Delta t) \rightarrow H_y^{k,n}, \tag{35}$$

where lattice position is indicated by superscripts for position k and time n, respectively. In the following, we rearrange the Ampere's and Faraday's law into the finite difference form as,

$$\frac{E_x^{k,n+1/2} - E_x^{k,n-1/2}}{\Delta t} = -\frac{H_y^{k+1/2,n} - H_y^{k-1/2,n}}{\epsilon_0 \Delta z},$$
$$\frac{H_y^{k+1/2,n+1} - H_y^{k+1/2,n}}{\Delta t} = -\frac{E_x^{k+1,n+1/2} - E_x^{k,n+1/2}}{\mu_0 \Delta z},$$

which explicate that the **E**-field is solved at integer space (k) steps, and half-integer ($n \pm 1/2$) time steps, while the **H**-field is solved by half-integer space ($k \pm 1/2$) steps, and integer time steps (n), by the assumption that both fields are interleaved in both space and time. Hence, with H, $k \pm 1/2$ means that H is situated between between the E field values, as shown in Fig. 10, similarly for E; this stems from the property of *Yee's lattice* or staggered grid lattice where both fields are offset in space and time, as depicted in Fig. 10.
Now we transform the above equations into two simultaneous recursive algorithms by solving for E_x at time $n + 1/2$ and H_y at time $n + 1$:

$$E_x^{k,n+1/2} = E_x^{k,n-1/2} - \frac{\Delta t}{\epsilon_0 \Delta z}\left(H_y^{k+1/2,n} - H_y^{k-1/2,n}\right), \tag{36}$$
$$H_y^{k+1/2,n+1} = H_y^{k+1/2,n} - \frac{\Delta t}{\mu_0 \Delta z}\left(E_x^{k+1,n+1/2} - E_x^{k,n+1/2}\right). \tag{37}$$

In Eq. (36) $E_x^{k,n+1/2}$ is the updated value based on the past step $E_x^{k,n-1/2}$ plus the finite difference, CDA, of the H_y derivative. Very similarly, the H_y is updated via Faraday's equation. In this way, the method computes recursively the fields for the entire time and space desired. These equations form the fundamental building blocks for how to understand the FDTD method.

After discretising the equations, one wants to renormalise Eqs.(36-37) to get the same dimensionality and order of magnitude between electric and magnetic fields. The main reason for this is that E_x and H_y differs by several orders of magnitude due to ϵ_0 and μ_0 [20]. The solution is to normalize by,

$$\widetilde{E} = \sqrt{\frac{\epsilon_0}{\mu_0}} E \qquad \text{Renormalisation.} \tag{38}$$

Substituting the change of variables Eq.(38) into Eqs.(36-37) gives

$$\widetilde{E}_x^{k,n+1/2} = \widetilde{E}_x^{k,n-1/2} - \beta\left(H_y^{k+1/2,n} - H_y^{k-1/2,n}\right), \tag{39}$$
$$H_y^{k+1/2,n+1} = H_y^{k+1/2,n} - \beta\left(\widetilde{E}_x^{k+1,n+1/2} - \widetilde{E}_x^{k,n+1/2}\right), \tag{40}$$

where $\beta = \frac{c}{\Delta z/\Delta t}$ is the ratio of the speed of light to grid velocity $\frac{\Delta z}{\Delta t}$ in vacuum and $c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}$ is the speed of light.

In Eqs.(39-40), the Ampere's and Faraday's law are discretised and staggered in time and space. Then, for given initial conditions and boundary conditions, we compute the corresponding **E** and **H**-fields recursively via the FDTD method till the space-time intervals are evaluated. The algorithm will be a first derivatives' system which makes it simple to implement, even though the intermixing of the electric and magnetic fields.

A simple step by step overview of the algorithm proposed by *Dr. Kane S. Yee* [21] follows:

1. Solve the finite difference Eqs. (39-40) to obtain the unknown updated fields in terms of the known past fields.
2. Evaluate the electric fields half-integer time step, $n+1/2$, into the future as illustrated in Eq. (36). Once they are known, they become past fields that can be used to update the magnetic field.
3. Evaluate the magnetic fields one-integer time step into the future at $n+1$ as shown in Eq. (37). These points can now be used to calculate the electric field of the next time step.
4. Repeat the two last steps till the fields are computed over the desired time duration.

3.1.3 Stability, *Courant stability condition*

A necessary condition for the stability of the algorithm is the *Courant condition*. An appropriate size of the cells like the space step Δz and time step Δt need to be carefully selected to fulfil this condition [20]. First, we define β, the *ratio of speed of light to the mesh velocity* $\Delta z/\Delta t$, to be

$$\beta \overset{def}{=} \frac{c}{\Delta z/\Delta t}$$

For numerical stability, $\beta < 1$, in the specific case of one dimensional problem. In the most general case of n dimension we instead have the inequality

$$\text{with } \beta \overset{def}{=} \frac{c}{\Delta z/\Delta t} \rightarrow \frac{\Delta t}{\sqrt{\epsilon_0 \mu_0} \cdot \Delta z} = c_0 \frac{\frac{\Delta z}{\sqrt{n}}}{\Delta z} \leq \frac{1}{\sqrt{n}} \tag{41}$$

with $n = 2, 3$ for specific dimensions such as 2D and 3D, respectively.
Once defined, β and the cell size Δz that characterise the accuracy of the computation, Δt follows according to Eq. (41) for stability scope.
In conclusion, the *Courant stability condition* $\Delta t = \frac{\Delta x}{\sqrt{n} c_0}$ is used to ensure stability during simulations by restricting the time step Δt once the discretisation Δz, is fixed.

3.1.4 Coded examples for 1D pulse propagation

The following examples are inspired from the work of *Professor Sullivan* [20], and implemented using `Python3`, and related libraries, i.e. `Numpy` and `Matplotlib`.
The *Yee's algorithm* [21], basically Eqs. (39-40), can be summarised as follows:

```
    Ex[k, 1] = Ex[k, 0] + beta * (Hy[k-1, 0] - Hy[k+1, 0]) #Ex[k, n+1/2] hence, Ex is f(x, t)
    Hy[k, 1] = Hy[k, 0] + beta * (Ex[k-1, 0] - Ex[k+1, 0])
```

for the discrete form of the Maxwell's equations along the desired time.
Alternatively,

```
    Ex[k] = Ex[k] + beta * (Hy[k-1] - Hy[k]) # same as x = x + 1 OR x += 1
    Hy[k] = Hy[k] + beta * (Ex[k] - Ex[k+1])
```

where this latter version is more slender and time dependency is implicit, such that the $n+1/2$ and $n-1/2$ superscripts from Eq.(39-40) are neglected, while space steps or positions are adjusted to integers as k and $k+1$ or $k-1$ for the sake of computation. Below follows a few examples of the pulse propagation on a finite number of *FDTD cells* where the **E** and **H**-fields are computed considering the staggered grid from *Yee's lattice*.

Free space

Here, we will illustrate how Gaussian pulse originated from the centre of the FDTD cells, at starting point `kc=100` as shown in Fig 11. The index `kc` represents the index of the electric field $E_x^{k,n}$ = `ex[kc]` in the middle of the simulation region, that travels outward in free space.

```
    # Gaussian pulse positioned in the middle of the problem space
    pulse = exp(-0.5 * ((t0 - time_step) / spread) ** 2)
    kc = 100 # index that represents the electric field Ex, in the middle of the cells
    ex[kc] = pulse # pulse located at centre where ex[k] = Ex(k)
```

`output plot:`

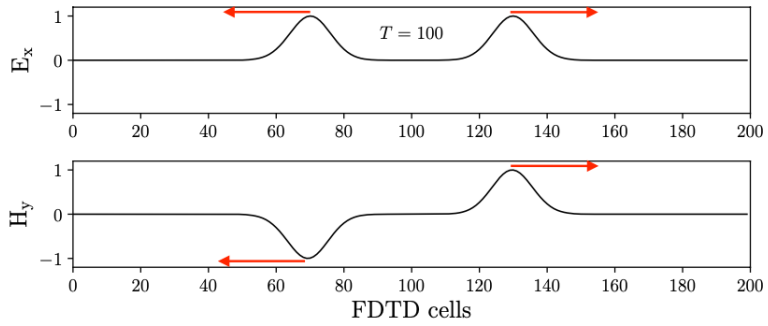


Figure 11. Illustration of a Gaussian pulse generated from the centre, `kc=100`, of the problem space where the electric field `ex[kc]`$=E_x^{k,n}$ propagates outward in free space with no medium in the simulation region in one dimension [20].

Notice that the pulse travels towards both end-points of the FDTD cells, see link to Python simulation. Thus, the E_x pulses are symmetric around the origin of the problem space, while H_y is both positive and negative based on the direction of propagation, as suggested by the *right hand-rule*.

Non-reflecting boundary condition in 1D

Here, we compute non-reflecting boundary conditions where the pulse originated from the centre of the FDTD region are absorbed in the end-points of the problem space without being reflected, see link to Python animation.

We have special points such as the boundary points `ke` and `0` where `ke` is the rightmost point and `0` is the leftmost point in the problem space, respectively as illustrated in Fig. 12. At these boundary points, we can't update their values with the algorithm. Instead, we need to implement a physical boundary conditions. Therefore, we need to treat $E_x^{0,n}$ in a special way, hence we assume that the flow of information is always originated from the centre of the simulation region outwards and no external source enters the problem space. Basically, any source is only going outward the FDTD space. To achieve this, one copies the value of the past $E_x^{1,n-2}$ into the first mesh cell at time n, taking into account that the **E**-field in one time step, Δt, travels distance

$$d = c_0 \cdot \Delta t = c_0 \cdot \underbrace{\frac{\Delta x}{2 \cdot c_0}}_{Courant} = \frac{\Delta x}{2}, \tag{42}$$

which shows that two time-steps are necessary for the **E**-field to cross one cell. In other words, we are at the boundary and we want

$$\boxed{E_x^{0,n} = E_x^{1,n-2}} \tag{43}$$

One possible way to implement this condition is to store the boundary data in two-element lists, i.e. `boundarylow` and `boundaryhigh`, see below implementation, from which one pops and appends the values to create the necessary two time-step delay. In particular, one pops the first element in the list and then append a new one, meaning that the newest element of the list spends exactly two steps inside the list before popping out.

To rephrase it, imagine a list of two elements such as `E=[0,0]`, then at the first time, one computes the algorithm shown in Eq. (39), while popping out the first element from this list. So now we have an old element left at position `E[0]` and a new one at `E[1]` in the list. In the second time step, the same algorithm is updated once again and the previous old existing element `E[0]` is popped out, so `E[1]` becomes `E[0]`, while a new `E[1]` is appended into the same list. Here we remind that two-time steps are necessary for the E_x-field to travel one cell. In this way, one emulates the extinction of the pulse at lower and upper boundaries by popping and appending elements from two lists, i.e. `boundarylow` and `boundaryhigh`.

A possible implementation of non-reflecting boundaries:

```
    # Non-reflecting Boundary Conditions
    ex[0] = boundarylow.pop(0) # Stack/Pile, Push-Pop method.
    boundarylow.append(ex[1])
    ex[ke - 1] = boundaryhigh.pop(0) # ke = number of FDTD cells.
    boundaryhigh.append(ex[ke - 2])
```

where `ke` is the number of FDTD cells and `kc=ke/2` is the centre of the cell where the pulse is triggered. Alternatively, once can employ `queue` module from `Python3` library.

`output plot:`

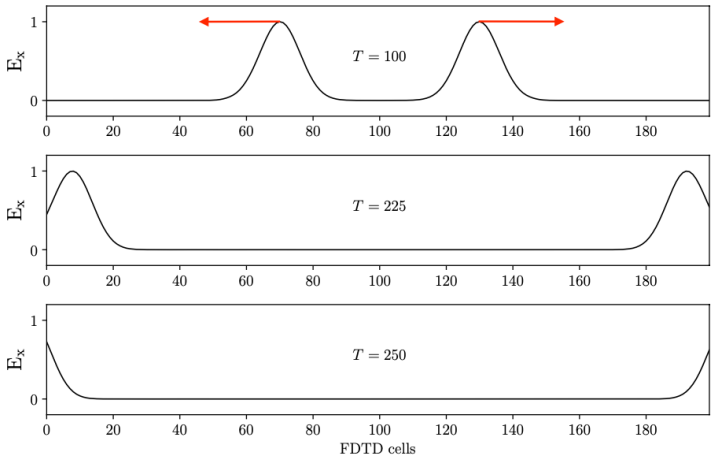


Figure 12. The Gaussian pulse, $G(t)$ generated via FDTD simulation from the centre of the problem space, propagates towards both the endpoints or edges of the FDTD space. At the edges, the E_x pulses vanish due to the non-reflecting boundaries; hence no reflecting components of the pulse bounces back into the FDTD simulation space [20].

Propagation towards a dielectric medium in 1D

In our final example, the pulse is now traveling from the left edge of the FDTD space towards a dielectric material starting from the leftmost simulation region and the effect of the collision is shown in Fig. 13.

`output plot:`

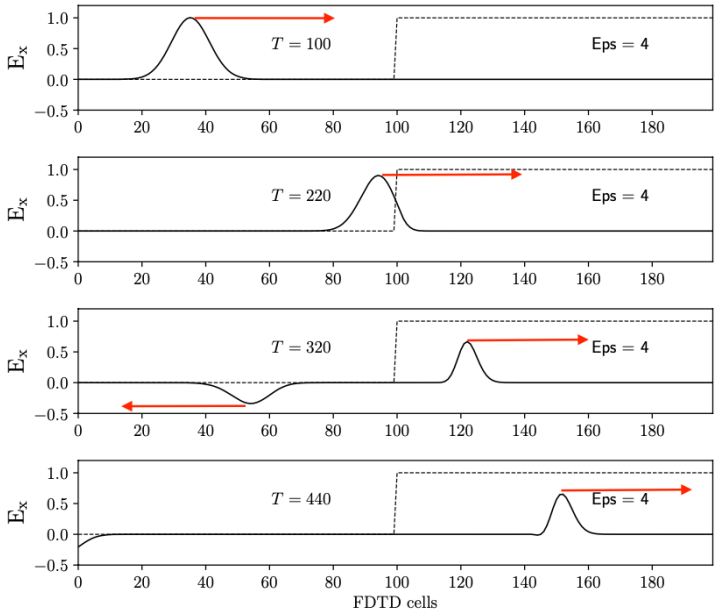


Figure 13. Representation of a pulse that travels from the left towards a medium with a relative dielectric constant, $\epsilon = 4$, where at time step $T = 220$ collides with a dielectric material. The collision yields one reflected and one transmitted part of the pulse. [20][22].

For simulations in the case of dielectric medium it is necessary to consider an extra factor $\epsilon_r = 4$ into Eqs. (8), such as

$$\frac{\partial \mathbf{E}}{\partial t} = \frac{1}{\epsilon_0 \epsilon_r} \boldsymbol{\nabla} \times \mathbf{H} \implies \frac{\partial E_x}{\partial t} = -\frac{1}{\epsilon_0 \epsilon_r} \frac{\partial H_y(z,t)}{\partial z}, \quad \text{Ampère's law} \tag{44}$$

$$\frac{\partial \mathbf{H}}{\partial t} = -\frac{1}{\mu_0} \boldsymbol{\nabla} \times \mathbf{E} \implies \frac{\partial H_y}{\partial t} = -\frac{1}{\mu_0} \frac{\partial E_x(z,t)}{\partial z}, \quad \text{Faraday's law} \tag{45}$$

with a similar factor needed to be added to the finite difference CDA, Eq. (39), which takes the final form in code

```
Ex[k] = Ex[k] + cb[k] * (Hy[k-1] - Hy[k])
Hy[k] = Hy[k] + beta * (Ex[k] - Ex[k+1])
```

where `cb[k] = 1/eps`, `eps` $= \epsilon$
As shown in Fig. 13, after collision with a medium, a Gaussian pulse, $G(t)$, is partially reflected and partially transmitted. However, notice a π phase shift of the reflected component, which is due to *energy and momentum conservation* [22], see link to Python animation.

In conclusion, these examples of pulse propagation in FDTD cells have demonstrated how a staggered grid can be implemented and explored in three realistic physical scenarios in 1D to familiarize ourselves with the basics of the FDTD simulation scheme.

3.2 FDTD in Lumerical

From this point on, we will exploit specific features from the Lumerical FDTD package [23], instead of implementing codes ourselves. Notice that all the above simulations are run in a staggered, discretised, and uniform environment which is not necessarily the one employed by Lumerical. Lumerical treats boundaries and space with different approaches such as perfectly matching layers (PML) , see section 3.2.2. In addition, it uses a non-uniform mesh technique to maximize the use of its computational resources, see section 3.2.4.
Lumerical adjusts the mesh, in accordance to the refractive index, n, of the material. Considering that frequency, $f = v/\lambda$, is conserved but not speed $v = c/n$ and $\lambda = \lambda_0/n$; it follows that n can be considered as a scaling factor by which v and λ are reduced and accordingly, the mesh size as shown in Fig. 14.

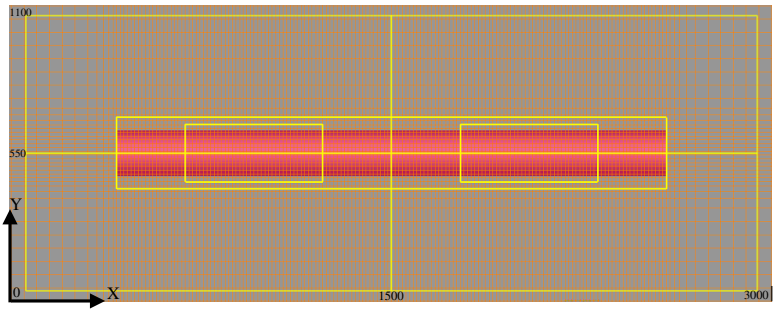


Figure 14. Grid representation of the FDTD simulation environment. Hovering the problem space, one notices Lumerical tends to make finer and dense grids close to the NW boundaries to keep some accuracy. On the other hand, it increases the grid sizes outside the NW domain to reduce computation time, following a non-uniform meshing. Figure generated from Lumerical simulation.

3.2.1 Initial conditions and sources

Lumerical offers several options of power sources such as beam source, dipole source, and total field scattering field (TFSF) [23]. In this project, the LED produces the radiation pattern that is modelled by introducing dipole sources along the $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ directions to be studied independently, as clarified in the Results section. The dipole sources are positioned in the centre of a *Transmitter* as is exemplified in Fig. 15 to observe and evaluate how the Directivity varies accordingly to different wavelengths and diameters, see link to Lumerical simulation 1. For the *Receiver*, we have mainly worked with a plane wave source which injects power along a plane into the FDTD simulation region as shown in Fig. 16 to monitor and study the absorption in the NW. We expect high power source absorption for small angles α, and low absorption for large angles, see link to Lumerical simulation 2.

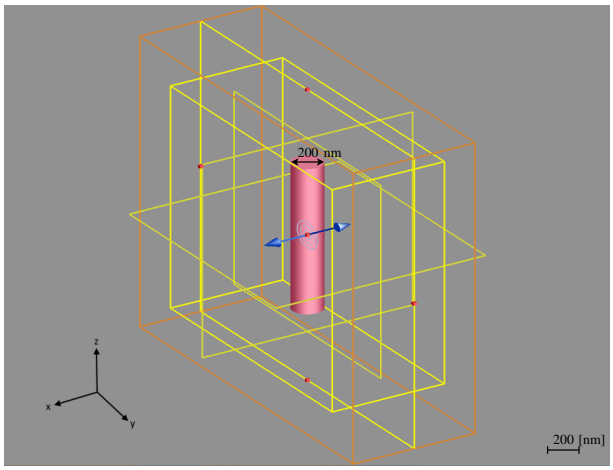


Figure 15. Dipole source is placed at the centre of the NW to model the LED radiation pattern. The movie monitor along the yz-plane, depicted with 4 red dots, records the intensity, $|\mathbf{E}|^2$ of **E**-fields. While, the FDTD region is depicted as orange box. Figure generated from Lumerical simulation.

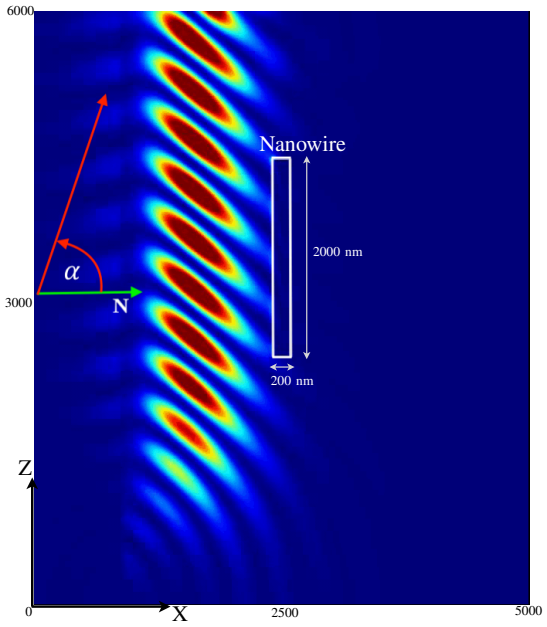


Figure 16. Wavefront of a *plane wave source* impinging on a nanowire within an FDTD simulation. The NW is centred at the origin of the simulation space, where the plane wave is injected with an angle α, with respect to the normal vector N of the leftmost interface. Figure generated from Lumerical simulation.

3.2.2 Boundary conditions

PML boundary conditions are necessary to keep the outgoing **E** and **H** fields from being reflected back into the problem space, see section 2.4. In Lumerical, the purpose of the perfect matching layers (PML) is to minimise the reflection noise and discard any possible numerical instability, without penalising computational time [20][24]. One can view the PMLs, as an *absorbing boundary medium* or "free flow" boundary conditions, as illustrated in Fig. 12 where the fields disappear at the end-facets of the problem space.

There are several options of PML profiles in the Lumerical software [19], and each of them has its own numerical behaviour customised for a specific task. However, we mainly focus on the *Standard and Stabilised profiles*, though exists other two profiles such as *Steep Angle and Custom* [24][25]. *Standard* profile works rather well for a limited number of layers and it performs the best when the PML boundaries are stretched along the simulation region. While, a material or device that cuts through a PML region might cause numerical instability. In this latter case, it is highly suggested to adopt a *Stability Profile* which is tailored purposely for this scenario, at the expense of a higher number of PML layers employed.

Overall, increasing the number of layers leads to a lower reflection at the expense of computational

efficiency; hence, for a reliable and good result, it is recommended to select an appropriate number of layers according to the computation time and memory available. Thus, the number of PMLs was mainly based on the computation time and quality of the results.

A simple scenario of a PML is demonstrated in Fig. 12 where a Gaussian pulse is absorbed at the endpoints or edges without reflecting anything back. An alternative example scenario is a perfect cavity/resonator with PML boundaries, where we may observe the behaviour in Fig. 17 [26], where the evanescent fields are effectively cut by the PML. This is also what we would expect from a similar simulation in Lumerical.

In Fig. 17, the source is set at the leftmost end, similarly to Fig. 13, and propagates along the cavity depicted in grey and filled with confined modes, blue and red bulbs in x-direction. While, in the y-direction, sufficient room is left to avoid that the outer PML boundaries can affect the cavity modes. PMLs are inside the edges of the simulation region, in order to absorb the confined modes at the end-point. The thickness of the PML layer is a crucial parameter to limit possible back-reflection from the source [26]. This latter FDTD simulation example was scripted using a powerful open-source `MEEP` package [26] which is a valid alternative to Lumerical.

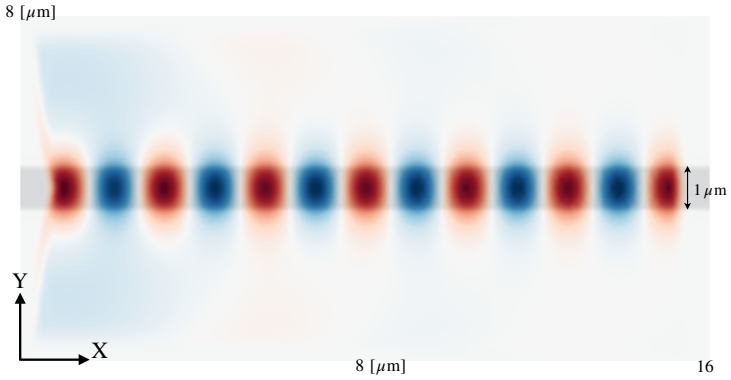


Figure 17. Representation of a perfect cavity which is the dark grey area in the centre of the 2D region. PMLs are employed to act as fictitious absorbing material added around the edges of the simulation region, so that modes or fields rapidly go to zero. The incoming source, from the leftmost side of the problem space, diffuses along x-y-directions forming cavity modes, blue and red bulbs, due to superposition of travelling waves. Figure generated using `MEEP` package [26] copyrighted by MIT under the GNU General Public Licence [27].

3.2.3 Storing the solution - Monitors

Frequency-domain field monitors are available in Lumerical to store data in the frequency domain from simulation results across some spatial region within the FDTD simulation region [28]. For the Receiver study, we are interested in storing information from two distinct transmission boxes and not over the whole nanowire. This is because we want to see the absorption difference between the two regions. A collection of data that are collected by the power monitors follows:

E: Electric field data as a function of position and frequency/wavelength.

H: Magnetic field data as a function of position and frequency/wavelength.

P: Poynting vector as a function of position and frequency/wavelength.

T: Transmission as a function of frequency/wavelength.

As an example of a relevant monitor for the study of the transmitter, notice in Fig. 15 a power monitor along yz-plane, depicted with 4 red dots, necessary to record the intensity, $|\mathbf{E}|^2$ of the **E**-fields, for instance. The extracted data such as **E**-fields versus angle, along different axes, are used to fit pulse peaks and define the standard deviation, σ, and directivity, D, as detailed in the Results section 4.2.2.

Figure 18 shows the simulation setup for case II for the Receiver NW studied. The NW is set at $x, y = [\frac{1}{4}, \frac{3}{4}]$ of the simulation space to provide the best conditions for a homogeneous illumination for both cases I and II. This method development, we refer to as *Homogenisation*, see link to Lumerical animation, case II as comparison to the initial setup as described in Fig. 16.

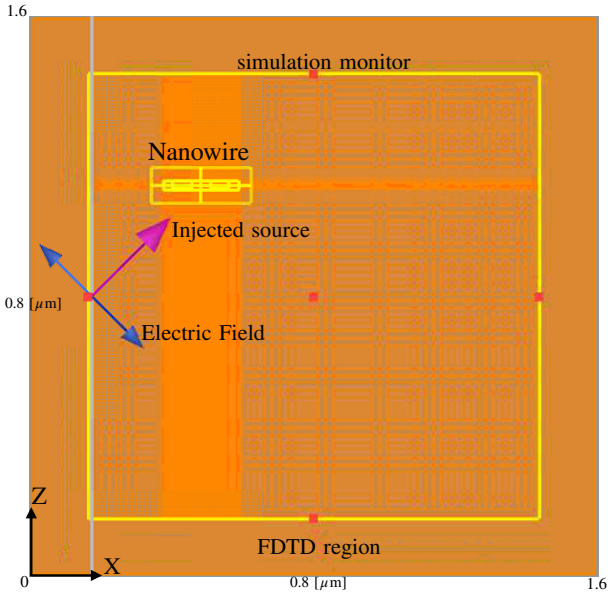
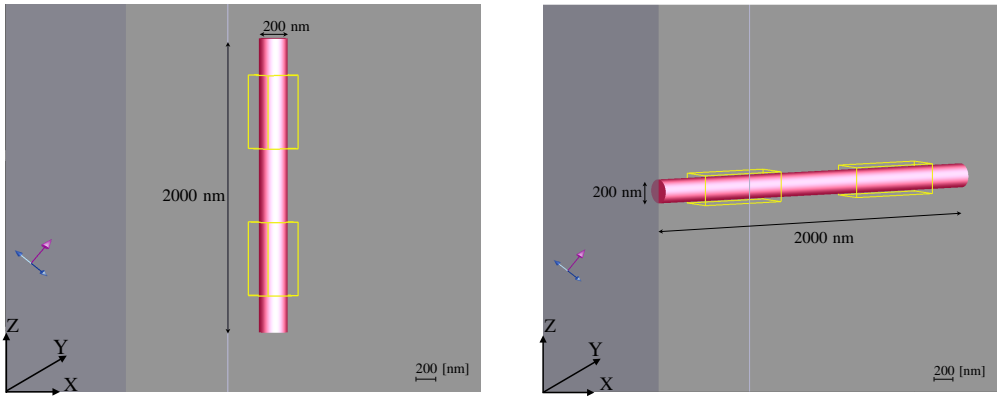


Figure 18. *Homogenisation* of the simulation environment where a monitor, depicted in yellow, is set along the whole xz-plane of the FDTD simulation region to record and monitor all the necessary events within the problem space such as absorption and intensity. Figure generated from Lumerical simulation.

Transmission box monitors around depletion regions

In Fig. 19, the transmission box monitors are used to record the flow of incoming electromagnetic waves to determine the difference in *in- and outflow* related to the absorbed power inside the volume of the boxes. The carrier difference on the center electrode, that non-equal absorption in the two regions will produce, charges the gate electrode of the FET[14] and thus turns it either on or off.



(a) Transmission box monitors *case I*.

(b) Transmission box monitors *case II*

Figure 19. Lumerical transmission box monitors employed around the depletion regions of semiconductors to measure the absorption in two specific regions. Figure generated from Lumerical simulation.

[14]FET is a type of transistor that use electric field to control the current flow of charges in a semiconductor.

3.2.4 Limitations

The FDTD method has a general difficulty with a large angle of incoming power source as displayed in Fig. 16, with respect to the PML boundary; [15] this may cause issues such as PML reflections of the source into the FDTD region and missing collision with a device. In Lumerical, the incoming source sweep angle is limited between `0°:50°` [16], which forced the experiment to be split into two cases, *case I* and *case II*, respectively, as shown in Fig. 20(a). Thereby, we sweep the source and not the NW rotation because we cannot rotate the transmission box monitors alongside the nanowires, accordingly, while keeping fixed the power source as illustrated in Fig. 20(b). The main problem of rotating the NW instead of the incoming source is that the box monitors as shown in Fig. 19, need to be rotated according to the NW and this feature is not available in Lumerical. The latter constrain is a compelling computational limitation rather than a physical one. Hence, the rest of the work and results are carried on using the configuration described in Fig. 20(a).

(a) Rotate the power source w.r.t a fixed NW into *case I* and *case II*, respectively.

(b) Rotate the NW w.r.t a fixed power source as unique case.

Figure 20. (a) Schematic representation of a fixed NW (case I & II) in the problem space where the incoming power source around the nanowire is swept. This scheme is the one employed during the entire experiment for the Receiver NW. In contrast, (b) represents a sketch of an alternative solution where instead a nanowire is rotated for a fixed source.

Further relevant limitations, with respect to the strengths and weaknesses of the FDTD method, are described below.

Advantages of FDTD modeling [29]:

- FDTD is a time-domain technique where a single simulation is sufficient to obtain a broad range of response system frequencies when a Gaussian pulse $G(t)$, is employed.
- FDTD allows to model easily a wide range of diverse linear dielectric and magnetic media.
- FDTD computes the E and H-fields over the entire meshed problem space in time which can be stored via monitors. Stored results can be manipulated for closer analyses and inspections.

Disadvantageous:

- FDTD computes the E and H-field overall problem space that requires be meshed; hence, this process can lead to a high computational time for large models.
- In addition, space discretisation can vary sensibly in spatial grid size, so this is an extra computational memory cost to take into account.
- A computational domain must be finite to be allocated in memory. So, often artificial boundaries are employed; however, they can cause errors. These errors can be prevented using PML to simulate an infinite and unbounded computational domain.

[15] The simulation radiation that hits the *Receiver* is approximated as a plane wave in the *far-field*, for distance $r \gg \lambda$ as mentioned in the Theory chapter, dipole sources.

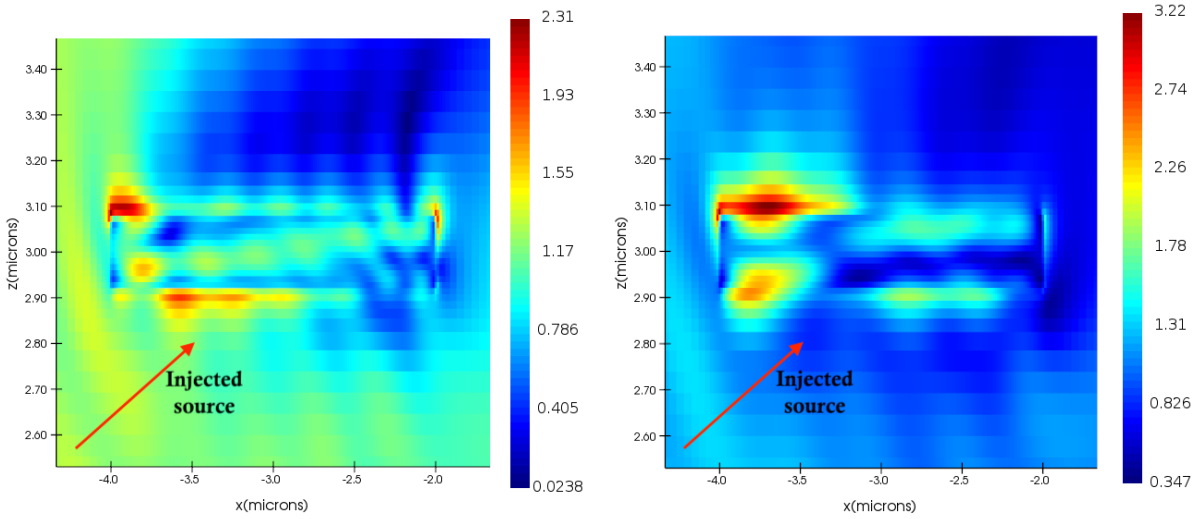
[16] It is not a general FDTD problem but a Lumerical one that cannot sweep the impinging source for angles $> 50°$.

4 Results and Discussion

This chapter is devoted to the study of electromagnetic simulations from Lumerical FDTD method to validate hypotheses 1) and 2), see section 1.1. The structure of this section is as follow:

- Firstly, we present results that show a drastic difference in absorption between the two photodiode regions in the Receiver NW's respective ends, by simulating certain illumination angles, wavelengths and NW diameters.
- Secondly, after selecting a proper power source, we show different communication wavelengths and NW diameters for the Transmitter NW to observe how the Directivity changes and how it can be controlled. Simulation data of Directivity are assembled in a table for comparison.
- Lastly, we pursue to establish the *Weight function* used to calculate the weights that we change by rotating nanowires of our neural network, which leads to our final result.

4.1 Receiver simulation via FDTD method in Lumerical



(a) $|\mathbf{E}|$, xz-plane $d = 160\,\text{nm}$, $\lambda = 650\,\text{nm}$. (b) $|\mathbf{E}|$, xz-plane $d = 160\,\text{nm}$, $\lambda = 800\,\text{nm}$.

Figure 21. Representation of an horizontal NW, in a FDTD simulation region, illuminated by an injected source at an angle $\alpha = 45°$ from the bottom-left corner for both figures as schematised by the red arrows. Each image displays the absolute electric field strength $|\mathbf{E}|$ $\left[\frac{\text{V}}{\text{m}}\right]$ along a NW, via a colour-map, for a single simulation, along an xz-plane view. For a fixed NW's diameter $d = 160\,\text{nm}$, **(a) featuring a shorter wavelength** λ**, we can see many different waveguide modes allowed**. While, in **(b) providing a longer wavelength** λ**, the patterns are much simpler**. Images extracted from a Lumerical simulation.

Before going into detailed simulations of the angle-resolved absorption, we study the physical promising parameters such as the wavelength, λ, and the NW's diameter, d. Indeed, in a recent study, *Van Dam et al.* [30] proved that it is possible to find a combination of wavelength and diameter, where the modes inside the cavities are dominated by the fundamental one, HE_{11}. Such a situation can be achieved for NW's diameter below 170 nm, which is optimal for our application.
We begin to fix the NW's diameter of $d = 160\,\text{nm}$ and look at the $|\mathbf{E}|$-field strength at various wavelengths. A plane wave is injected along the xz-plane, with a fixed angle of $\alpha = 45°$, from the bottom-left corner of the simulation environment, as shown in Fig. 21. We monitor the number of the cavity modes, which is expected to depend on the wavelength. *A priori*, we cannot analyse the different modes[17], but we can infer from the results if we have a few or many modes. More specifically, we observe in Fig. 21(a, b) the interference pattern between the different modes: the fewer the number of modes, the less interference pattern should be observed. In the example Fig. 21(a), for $\lambda = 650\,\text{nm}$, many modes are exhibited in the receiver NW, seen as periodic intensity modulations. On the contrary, in Fig. 21(b), many modes

[17] To define if the modes are high or low order ones, one may perform a signal decomposition, e.g. *Fourier series*.

have been suppressed by increasing the wavelength of the impinging power source to 800 nm. This result suggests that we are close to a single mode, in agreement with the previous work [30]. Alternatively, at a fixed wavelength, the suppression of modes can be done by shrinking the NW's diameter. Due to a time constraint, we chose a NW diameter of 160 nm because the wavelength of 800 nm suits the InP material well [30].

After determining the physical parameters, we want now to define the relative absorption[18] cross-section of the NW at $\lambda = 800$ nm of the photo-diode(+) and the photo-diode(−) as a function of the illumination angle as shown in Fig. 22. Unfortunately, Lumerical is known to have convergence issues, see section 3.2.4. So, we have to implement two distinct simulation cases as illustrated in Fig. 22 to overcome the power injection problem; it is cumbersome to inject power source for $\alpha > 60°$ [29]. Hence, we sweep an impinging power source, i.e. plane wave 1 between $\alpha = 0°\!:45°$ for one case, and then plane wave 2 from $\alpha = 45°\!:90°$ for the second case, accordingly.
The difference in absorbed carriers between the +/− photo-diodes represents the built-up potential on the V_{gate} contact that control the circuit current via a transistor towards the Transmitter LED, see section 1.1. Therefore, for small illumination angles α align with respect to the NW axis, we expect drastic differences in the relative absorption, between the photo-diode(+) and the photo-diode(−) of the NW. For small illumination angles, the photo-diode(−) is farther away from the power source, so in comparison, the relative absorption should be much smaller than photo-diode(+). On the other hand, for $\alpha = 90°$, the impinging power source is equally distributed along the NW from the plane wave 2 viewpoint. Ergo, at $\alpha = 90°$, the relative absorption pattern between +/− photo-diodes should be the same. While, for $\alpha = 45°$, the relative absorption[19] to an impinging power source should overlap for both plane wave 1 and plane wave 2 because 45° is a common angle. To this end, we can use the symmetry requirements to estimate how well we have constructed our simulation environment.

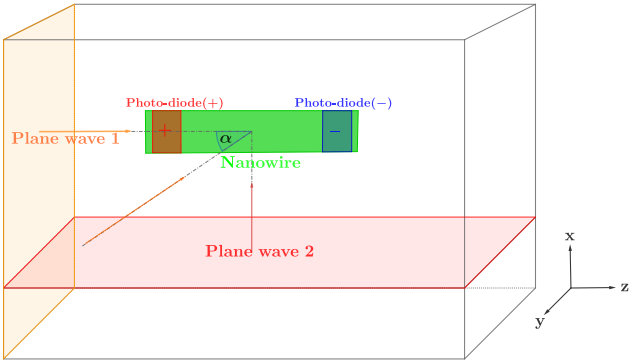
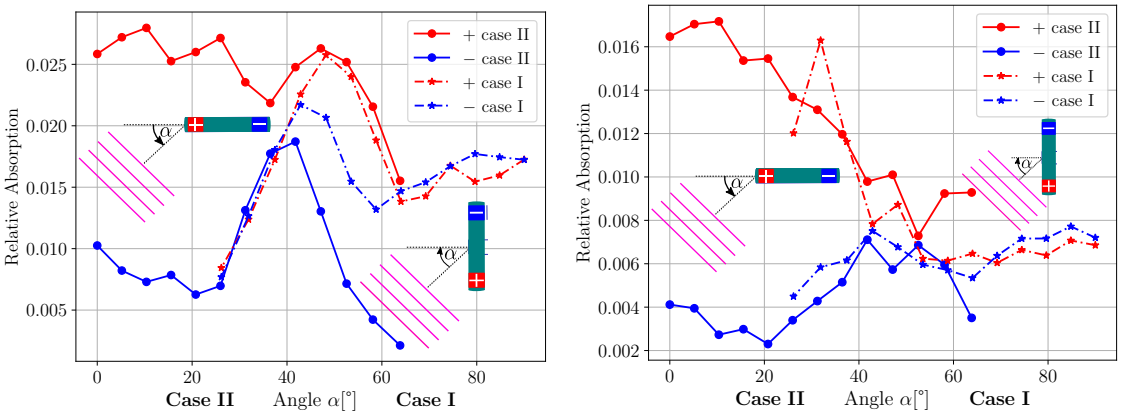


Figure 22. Overview of the simulation setup environment which shows two distinct sources, i.e. plane wave 1 and plane wave 2, injected from two different sides of the simulation cell. In our numerical experiment, we sweep the injection angle from 0° to 45° for one case and then from $45°\!:90°$ for the second case, accordingly. Then, we patch the full results, $0°\!:90°$, together to obtain a full simulation data of relative absorption, between two distinct +/− photo-diodes.

The acquired data from the FDTD simulations are four different datasets distinguished by two different setups, i.e. case I, II, and two distinct photo-diodes, red and blue per setup, as exemplified in Fig. 23. During each simulation, we sampled a finite number n of sweep angle points, e.g. $n = 13$ for the impinging power source, to save some computation time. Each sample point covers $\sim 5°$ of the total $0°\!:90°$.

We monitor the relative absorption pattern of the NW which depends on the illumination angle, α. More specifically, from Fig. 23(a), we can observe a significant difference in relative absorption between the red and blue data-points for small angles around 0° as expected, while we notice an overlapping tendency, for large illumination angles close to 90°. In contrast, around 45°, we can see a compelling gap between the red and blue relative absorption data-points. This gap is due to a lack of an homogeneous illumination with the current simulation environment, see section 3.2.3. This gap in the relative absorption cross-section needs to be addressed via a technique that we called *Homogenisation*, by repositioning the NW in the upper-left area of the FDTD region, see Fig. 18. This latter technique intends to solve issues due to geometrical effects. The *Homogenisation* results shown in Fig. 23(b), helped us to achieve the expected convergence of the data-points at $\alpha \sim 45°$.

[18] Relative absorption is meant with respect to the injected power source, i.e. plane wave.
[19] Transmission and absorption are interrelated to each other, e.g. $P_{abs} = A\,(I_{in} - I_{out}) = A\,I_0(1 - e^{-\alpha L}) = A\,I_0(T_L - T_0)$ where A is the injection area.



(a) Simulation with NW set at the centre of the FDTD region.

(b) Simulation with NW set at the upper-left of the FDTD region.

Figure 23. Collection of the main NW receiver simulations where the relative absorption of an injected plane wave is collected versus an illumination angle α, for both case II, 0° to 45°, and case I, 45° to 90°, respectively. **(b) has a better convergence at** 45° **w.r.t a) due to a reduction of geometrical effects.**

To investigate the simulations of the angle-resolved absorption in more depth, we patched, interpolated and scaled the sampled data-points as shown in Fig. 24, for the full range of illumination angles from 0°: 90°, to define two cross-section functions. We patched the red and blue lines from Fig. 23(b) respectively, at 45°. The interpolation of sampled data-points was computed via a piecewise smooth interpolation polynomial, e.g. *linear spline*[20]. Alternatively, the *Lagrange interpolation polynomial* is an effective approach; however, yields some evident oscillations at the edges of an interval, see Appendix A.3. Hence we dropped this latter method in favour of the *linear spline* one.

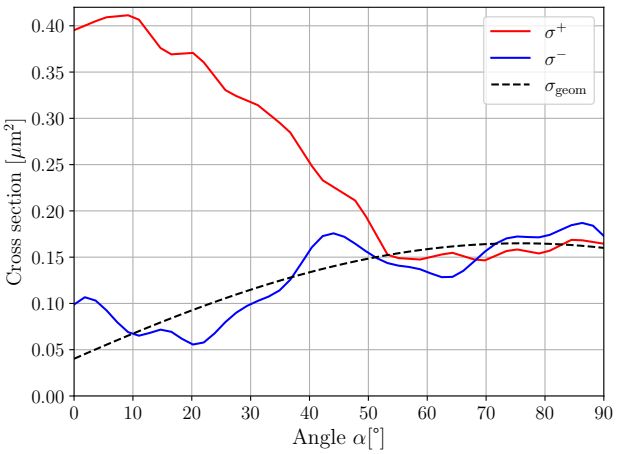


Figure 24. Comparison between simulated σ^+, σ^- cross-sections and the geometrical σ_{geom} used as theoretical reference versus α.

The resulting cross-sections are functions of α expressed by:

$$\sigma^+(\alpha) = \frac{P}{I} \cdot T^+(\alpha) = A_{\text{injection}} \cdot T^+(\alpha), \quad \left[\frac{\text{W}}{\text{W}/\mu\text{m}^2} = \mu\text{m}^2\right] \tag{46}$$

$$\sigma^-(\alpha) = \frac{P}{I} \cdot T^-(\alpha) = A_{\text{injection}} \cdot T^-(\alpha), \quad \text{I} = \left[\frac{\text{W}}{\mu\text{m}^2}\right] \tag{47}$$

where P is the power of the illumination source and I the intensity of the illumination, $A_{\text{injection}}$ the injection area and $T(\alpha)$ the transmission[21]. In addition, we compare the two simulated cross-sections σ^+

[20] Implementing a method, namely `scipy.interpolate.interp1d` from `Python3`.

[21] The absorption and transmission are interlinked to each other, e.g. $Abs = A(I_{\text{in}} - I_{\text{out}}) = A(I_0 - I_0 e^{-\kappa L}) = A\, I_0(T_L - T_0)$ where A = area of injection.

and σ^- with the geometrical σ_{geom} used as theoretical reference to shed light on the physical properties of Receiver NW. The σ_{geom} is given by:

$$\sigma_{\text{geom}}(\alpha) = \sigma_{\text{top}} \cdot \cos\alpha + \sigma_{\text{side}} \cdot \sin\alpha \tag{48}$$

where, if one visualises a NW as a cylinder, σ_{top} is the facet area of the NW ends, whereas σ_{side} are the added projected rectangular areas of the photo-diodes for $\alpha = 90°$ as exemplified in Fig. 22.

The simulated σ^+ (red curve) in Fig. 24 is remarkably higher than σ^- (blue curve), especially for small illumination source angles. Whereas, the σ^- curve well follows the σ_{geom} which highlights that we obtained consistent results in comparison with the theoretical one. The oscillation of σ^- around σ_{geom} is related to numerical instability. While, for $\alpha \sim 50°$ to $90°$ both σ^+, σ^- tend to clearly converge which is what we expect.

In summary, this final result comparing the simulated cross-sections for a wavelength of $\lambda = 800\,\text{nm}$ verifies our hypothesis 1), stating that we can tune the local absorption in a NW by changing the illumination angle.

4.2 Transmitter simulation via FDTD method in Lumerical

The purpose of these studies is to test hypothesis 2) stated in the Introduction. We started first by choosing the nanowire material to support the proper emission wavelength by tailoring the band-gap of the Transmitter such that we get the desired wavelength as output. We assume that the emission from the electron-hole recombination in the LED of the Transmitter produces light with an energy corresponding to the communication wavelength that we desire, namely 800 nm, as concluded in the Receiver section. Thereby, we want a band-gap above this latter energy value for the rest of the Transmitter NW (that is, outside of the small LED region) in order for it to be transparent to the emitted light. To fulfill these requirements, we create a material alloy GaInP with 30 % gallium and 70 % Indium, which corresponds to a band-gap of $\lambda = 750$ nm when converted to wavelength. As the size of the source that emits the radiation is small compared to the length of the GaInP nanowire as well as the energy is lower than our NW band-gap; we can neglect the variation of the refractive index and model the whole nanowire as consisting purely of our GaInP alloy. To sum up, the Transmitter NW model is made up of a single type of material that acts as a waveguide enough wide to support at least one mode [30].

4.2.1 Dipole source along x-axis

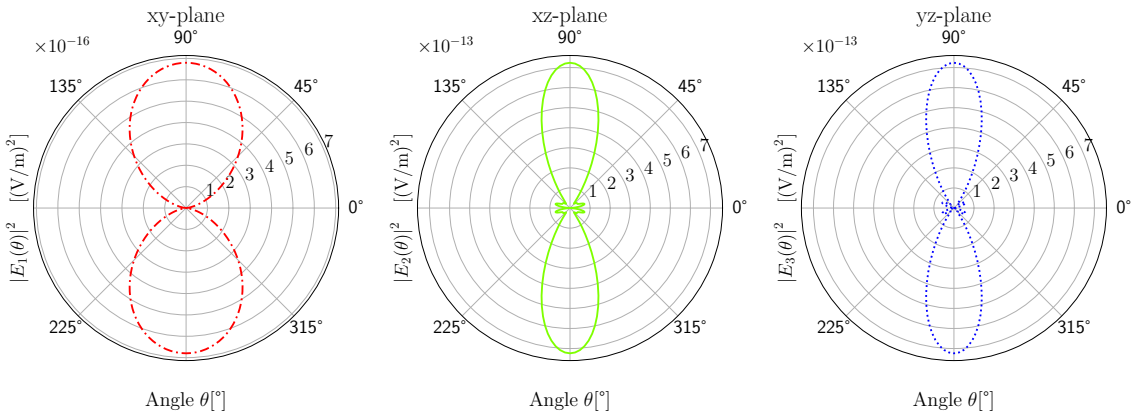


Figure 25. Emission intensity $|\mathbf{E}|^2$ patterns on the xy, xz and yz-plane for $\lambda = 800$ nm and diameter $d = 160$ nm.

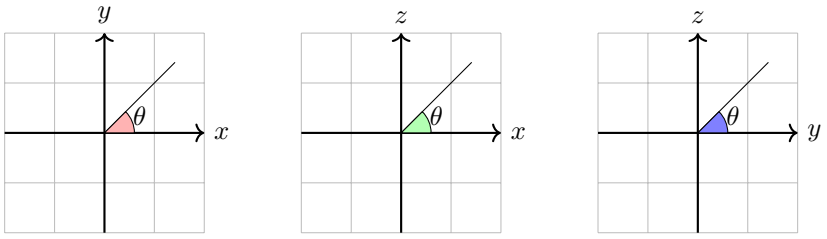


Figure 26. Polar plot references for the xy, xz, yz-plane.

The dipole source placed in the middle[22] of the NW creates an electromagnetic field that couples with the modes of the cavity, and these subsequently couple out to the free space by shaping and setting the far-field emission pattern of the LED along $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ axes. The NW will act as a waveguide where modes are partially travelling waves and partially standing waves.

To analyse and study the LED emission pattern, we have recorded the intensity of the electric fields $|E_i(\theta)|^2$ for $\lambda = 800\,\text{nm}$ and NW diameter $d = 160\,\text{nm}$ of dipole sources along all the three axes, i.e. $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$ with respect to the xy, xz, yz-plane, accordingly. Figure 25 is a decomposition with respect to the xy, xz, yz-plane of the dipole source along $\hat{\mathbf{x}}$ axis where the light is mainly emitted from the big lobes at $90\,^\circ$ and $270\,^\circ$ along the xz and yz-plane. For simplicity and intuitive visualisation, the Cartesian coordinates $[x, y, z]$ are mapped to spherical coordinates $[\rho, \theta, \phi]$, e.g. the radial distance, polar angle, and azimuthal angle, respectively. A prescription on how to interpret the angle in these plots is given in Fig. 26. The light emission pattern in the xy-plane describes the light emitted orthogonal to the NW. Compared to the emission in the xz and yz-plane, it is relatively marginal and negligible, about three orders smaller i.e. 10^{-16} versus 10^{-13}. Therefore, we mainly focus on the results from the xz and yz-plane. Similar results for dipole sources oriented along the $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ axes.

4.2.2 Fitting of the results

From the analysis of the simulation data presented in Fig. 27(a), we try to fit the signal peaks to acquire values for the standard deviation σ and the Directivity D. To properly fit the extracted data $|E_i|^2$, there exists a variety of fitting functions that can suit the task. However, after a close analysis of the shape of the peaks, as displayed in Fig. 27(a), we selected the Gaussian function as the best option. The Gaussian function is the probability density function of the normal distribution,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \tag{49}$$

where σ is the standard deviation and μ is the mean value, while $1/\sigma\sqrt{2\pi}$ is the amplitude factor. In short, we focus on the exponential part of $f(x)$ during the fitting process.
For all planes and diameters of interests, we:

- unwrapped the intensity pattern of the polar plot, as shown Fig. 27,
- fitted the Gaussian function to the emission peak between 0° : 180° as represented in Fig. 27(a), using the method `scipy.curvefit` implemented in `Python3`,
- extracted the standard deviation σ,
- computed the Directivity D from the data, as defined in Eq. (25) and the results are gathered in the Tab. 1.

There are also non-ideal scenarios such as the one denoted in Fig. 28(a) where the emission signals are populated by several unwanted peaks or side lobes in the folded polar plot illustrated in Fig. 28(b) which makes the transmission rather dispersive.

[22] The dipole source is set at the centre of the Transmitter NW for simplicity and practical reason. One wants to place the contact pads as far as possible from the LED; otherwise, the electrons and holes might recombine via the metal surfaces.

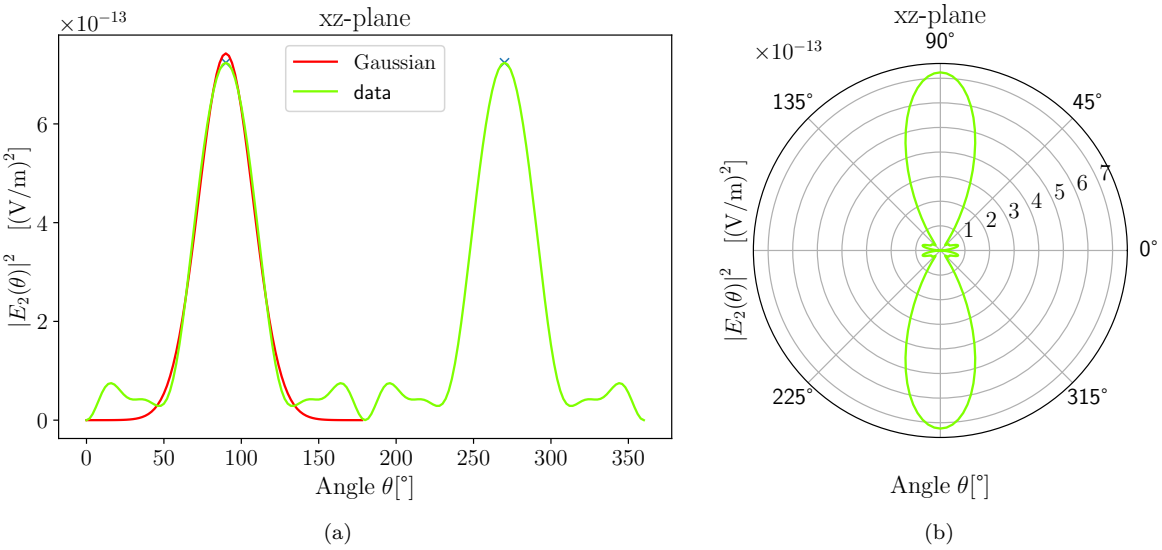


Figure 27. (a) Visualisation of the curve fitting of $|E_2(\theta)|^2$ along the xz-plane, for a NW diameter of **160 nm**, using the Gaussian function. (b) Same data-points as in (a) but for a folded plot. Both plots are displayed in polar coordinates; however, (a) unfolded and (b) folded, respectively. **In (a), we mainly focus on fitting the best function to the peak signals to gain promising values for the standard deviation σ, and directivity D.**

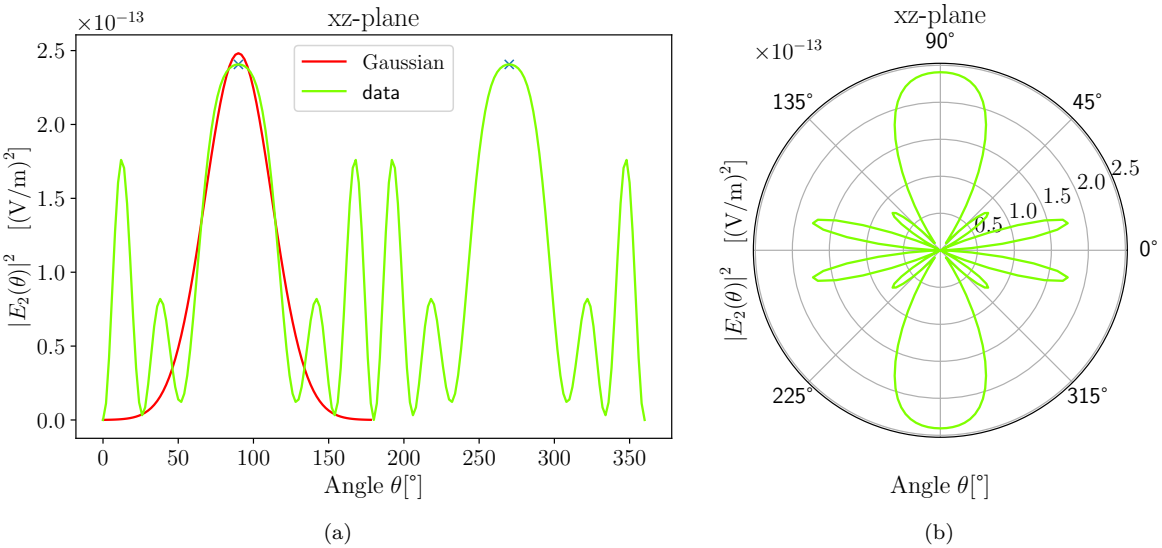


Figure 28. (a) Representation of the curve fitting procedure of the most intense peak of the radiation pattern at 90°, for a NW diameter of **200 nm** along the xz-plane as contrast to the prior Fig. 27. Note the symmetry of the two main peaks around angle 180°, which is expected due to the system geometry. (b) Same peaks as in (a) but folded as a polar plot.

Table 1. Assembled values of the NW diameter d, standard deviation σ and directivity D for the xz and yz-plane, respectively. The unit for the standard deviation is radians while Directivity is dimensionless. "–" means that no reasonable value has been found during the curve fitting procedure.

x-source

d (nm)	σ_{xz}	D_{xz}	σ_{yz}	D_{yz}
220	0.5437	1.8976	0.5316	2.1839
200	0.4966	2.2953	0.4860	2.7169
180	0.3226	2.522	–	1.8396
160	0.4313	3.6571	0.4248	3.8258
140	0.5981	2.9118	0.6164	2.7405
120	0.7016	2.49395	0.8633	2.0273

y-source

d (nm)	σ_{xz}	D_{xz}	σ_{yz}	D_{yz}
220	0.5315	2.1866	0.5449	1.9039
200	0.4860	2.7195	0.4968	2.3053
180	–	1.8488	0.3209	2.5286
160	0.4253	3.8231	0.4319	3.6636
140	0.6173	2.7395	0.5975	2.9125
120	0.8597	2.032	0.7011	2.4934

The summary of our work is given in Tab.1 where we want to present results for dipole sources oriented along the $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$-axes. While, $\hat{\mathbf{z}}$-oriented source is neglected due to non-dipole behaviour, see in Fig. 29(b), does not resemble our expectations. We found an optimal diameter for sharp pulses and high Directivity at 160 nm. We observe that σ can be controlled by tuning slightly the diameter between 120 -160 nm. On the other hand, we encounter also a critical diameter, e.g. 180 nm where a nanowire was populated by double signal peaks that were hard to fit. After a close analysis, we can conclude that the Transmitter has a spectrum of values that allow us to tune some parameters in our NN. For instance, if we want one-to-one node communication, we need a sharp and precise emission signal which can be obtain for $d = 160$ nm and gives a high Directivity value. In contrast, if we want to simulate a Fully-connected neural network with connections to a range of nodes, then we want a diffuse emission pattern using a small diameter, e.g. $d = 120$ nm and D small as well. Therefore, the change of Directivity will consequently alter the strength of the connections of our neural network.

To this point, we tested and validated our hypothesis 2), stating that we can direct and control the light cone emission through a NN by tuning the Transmitter parameters.

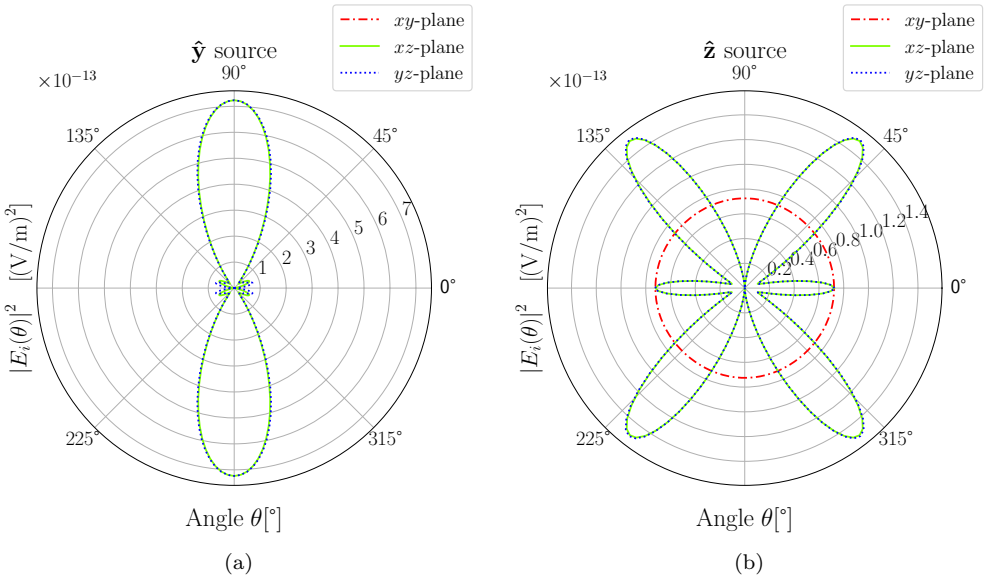


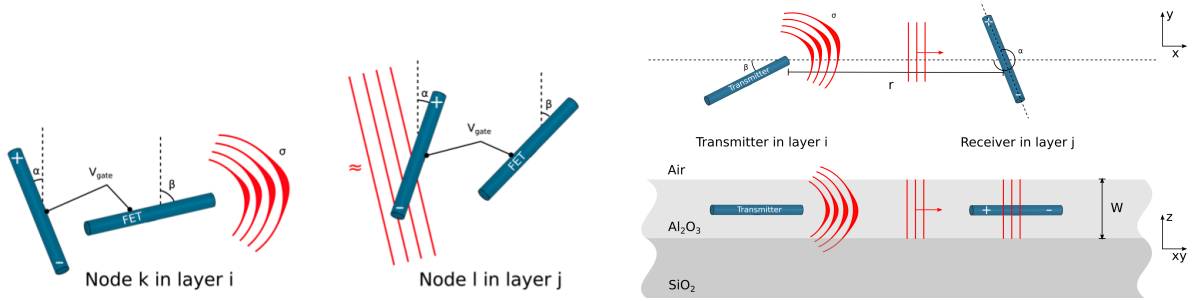
Figure 29. (a) Radiation intensity along the three planes for the $\hat{\mathbf{y}}$-source equivalent to the one shown in Fig. 25 for the $\hat{\mathbf{x}}$-source. While (b) is the dipole $\hat{\mathbf{z}}$-source as comparison to the $\hat{\mathbf{y}}$-source in (a).

4.3 Realisation of the network

Here, we briefly introduce a simple scheme and overview of the communication between a pair of artificial neurons as exemplified in Fig. 30, and we will explain the main features that characterise each component.

The layer-to-layer transmission occurs via a Transmitter that injects a radiation pulse onto the next layer thus acts as a light-emitting diode (LED). On the other hand, a Receiver absorbs the incident radiation source in two distinct regions labelled with (+/−), where each acts basically as a nanowire solar cell [8]. This latter process is repeated between layers and nodes during communication.

The thin line in Figure 30(a) marked V_{gate} represents an electrical coupling of a Receiver and Transmitter NW via a metal contact set at the centre of the receiver NW. The transmitter is assumed to be electrically connected, while the V_{gate} acts as a switcher for the emitter NW, see section 1.1. In addition, in Fig. 30(b) represents a sketch of the NWs' communication confined in a waveguide structure of materials such as air, SiO_2 and Al_2O_3. This scheme allows light to be trapped by exploiting the fact that air and the substrate SiO_2 have a lower refractive index with respect to the Al_2O_3.
Ergo, the higher refractive index of Al_2O_3 surrounded by two lower ones helps to confine propagating modes during node-to-node communication.



(a) *Receiver* and *Transmitter* electrically coupled via a V_{gate} which forms a neuron or antenna.

(b) Optical communication between nodes confined within a waveguide structure.

Figure 30. Neurons are electrically connected devices in (a) and sandwiched inside a waveguide structure in (b), that communicate with each other optically between hidden layers, forming a neural network. Figure courtesy of David Winge, licensed under CC BY [31].

4.4 Constructing the weight function from simulations

After determining the optimal NW diameter of $d = 160\,\text{nm}$ for the Receiver and Transmitter; we are interested in looking at the Weight Function that describes the change in voltage on the gate due to the difference in absorption between two distinct absorption regions, P^{+}_{abs} and P^{-}_{abs}, respectively, of the Receiver. The absorption difference is proportional to the ΔV_{gate} and this is what switches the transistor ON or OFF and controls the current through the LED, see section 1.1. We used the cross-sections, σ^{+} and σ^{-} from Eqs. (46, 47) respectively, necessary to compute the absorption[23], P^{+}_{abs} and P^{-}_{abs} in units of power, for any incoming intensity $I(r, \beta)$:

$$P^{+}_{\text{abs}} = \sigma^{+}(\alpha) \cdot I(r, \beta) \tag{50}$$

$$P^{-}_{\text{abs}} = \sigma^{-}(\alpha) \cdot I(r, \beta) \tag{51}$$

where the intensity is assumed to be originated from a Transmitter located at distance r from the Receiver and rotated by an angle β with respect to the connecting vector as exemplified in Fig. 31(b). To be able to express this intensity, the Transmitter Directive gain $G(\beta)$ from Eq. (24),

$$G(\beta) = \frac{U(\beta)}{2 \cdot \frac{1}{\pi} \int_0^{\frac{\pi}{2}} d\beta \cdot U(\beta)}, \tag{52}$$

can be used together with the total emitted power P_{tot}. Keep in mind that there exists a phase shift of 90° between the Gaussian function shown Fig. 27(a) and the one from Eq. (52). Fourthly, the intensity w.r.t (r, β):

$$I(r, \beta) = \frac{G(\beta) \cdot P_{\text{tot}}}{2\pi \cdot r \cdot W} \tag{53}$$

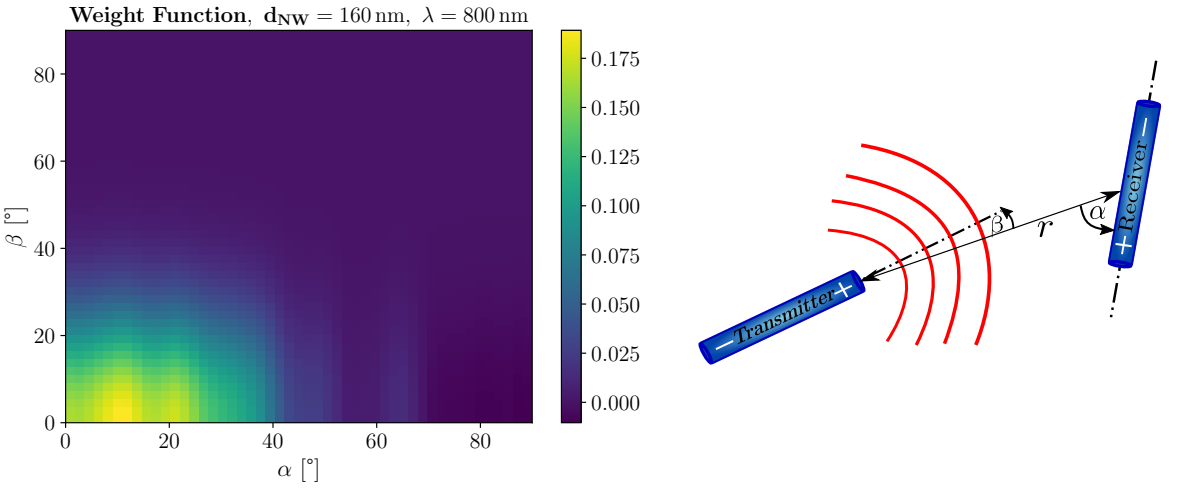
[23] Transmission and absorption are interrelated to each other, e.g. $P_{\text{abs}} = A\,(I_{\text{in}} - I_{\text{out}}) = A\,(I_0 - I_0 e^{-\alpha L}) = A\,I_0(T_L - T_0)$ and $(P^{+}_{\text{abs}} - P^{-}_{\text{abs}}) \propto \Delta V_{\text{gate}}$ where A = Area and P^{+}_{abs} and P^{-}_{abs} are the absorption in the +/− regions.

to express the intensity pattern where W is the width of waveguide structure as illustrated in Fig. 30(b). Finally, by relating the voltage difference to the difference in absorption we reach

$$W(r,\alpha,\beta) = \frac{(P^{+}_{\text{abs}} - P^{-}_{\text{abs}})}{P_{\text{tot}}} = \frac{(\sigma^{+}(\alpha) - \sigma^{-}(\alpha)) \cdot G(\beta) \cdot P_{\text{tot}}}{2\pi \cdot r \cdot W \cdot P_{\text{tot}}} \tag{54}$$

as our final expression for the Weight Function $W(r,\alpha,\beta)$. This represents a direct connection between the absorption in the + and − regions of the Receiver to the resulting weights. The Weight Function describes how the gate voltage changes for a fixed output power of a connected transmitter, if we rotate either the receiver or transmitter NW, or change the distance between the nodes. It is essentially the fraction of power that can be converted to charges on the gate electrode.

We illustrate the Weight Function in Fig. 31 which is our main achievement. One can notice that there is a big contrast in the 2D map where it is possible to "switch off" a connection in two way, by rotating the Receiver or the Transmitter for $\alpha, \beta > 50°$. While, for $\alpha, \beta < 30°$ there is a strong directivity in the optical signals that suggest that one can actually send signals from one node to another. However, if a Transmitter shines onto several Receiver NWs, any change of the Transmitter parameters would then influence an entire column of a weight matrix for a network with more than two neurons, see section 1.1. These correlation effects could be studied with the help of the Weight Function produced in this work. Therefore, the weights quantify the strength of node connection; however, if the connection is one-to-many neurons then we may consider a weight matrix, made up of a set of weight function evaluations.



(a) Visualisation of NN communication strength.

(b) Transmitter-Receiver optical communication.

Figure 31. (a) Visualisation of the expected Weight Function, $W(\alpha,\beta)$ for NW $\mathbf{d = 160}$ nm which describes the weights of our Neural Network as a function of our two main parameters α, β with the distance $r = 3000\,\text{nm}$ kept fixed. **(b) Framework of the optical communication between *Transmitter and Receiver* nanowires** with angles β and α set w.r.t the distance r.

5 Conclusion and Outlook

Our study shows that one can send information to either side of a InP Receiver NW in a controlled manner, simply by adjusting the illumination angle. This causes a drastic variation of relative absorption between its two ends. This result was obtained via Lumerical by creating a FDTD simulation region and studying the local absorption by the use of transmission box monitors around both ends of a NW, see section 3.2.3. In addition, we adjusted the NW diameter and incoming light wavelength. Hence, the main point is that one can tune the absorption difference between the plus and minus regions by simply changing the injected source angle.

For the Transmitter NW, we established that the width of the resulting light cone, originating from internal photo-emission, can be tuned by changing the nanowire diameter. To do so, we set a dipole source at the centre of the GaInP nanowire to study how the emission pattern changed with diameter. The emission peaks were fitted and we found an optimal range of Directivity and standard deviation σ values for a tunable emitter. In this way, one can tune the diameter based on the requested communication scheme: a wide spread signal for a Fully-connected neural network or a sharp and directed one for one-to-one node communication. Hence, the diameter size of the Transmitter was the key point to define the desired emission pattern.

By combining the results of the Receiver and Transmitter we provide the theoretical foundation of a functional artificial neuron able to communicate optically in a neural network. The communication between nodes are set by the relative characteristic angles (α, β) and the weights quantifying the strength of these connections are described by the Weight Function, displayed in Fig. 31. From Fig. 31, one can see that there is a compelling contrast in the 2D map where it is possible to "turn off" the connection in two ways, by rotating a Receiver or a Transmitter for $\alpha, \beta > 50°$ as depicted in the violet area. With the Weight Function as our main result, we laid the foundation to test hypothesis 3), see section 1.1.

As final remark , the Weight Function can be used to design networks based on artificial neurons similar to the ones studied here. One proposal would be to setup the requested weights and unknown angles as a nonlinear system of equations. Using the Weight Function from this work, one can then solve the system to find the optimal rotation angles of the Receivers and Transmitters of such a network. In other words, the solution to this latter mathematical system gives the rotations and positions of the nanowires in a neural network that best mimic a given set of trained weights.

A. Appendix

A.1 d'Alembert's solution to the wave equation

Plenty of interesting and relevant phenomena that arise frequently in mathematics and physics sciences such as fluid flows and electric and magnetic potential fields, are modelled by functions of several variables that satisfy certain *partial differential equations*, PDE. One of these PDE is the **Laplace's equation**, special case of Poisson's equation, where $\nabla^2\mathbf{u}(x_i) = 0$ must fulfilled, with function $u(x_i) \in \mathbb{R}^n$ be C^2-*smooth* as axiom. Solutions of the equation, are *harmonic functions*, within the domain, if and only if, they satisfied the Laplace's equation.

The PDE, we are interested in is the *wave equation* that can degenerate to the Laplace's equation within certain conditions; however, the function u needs to be C^2-*smooth* and not C^∞-*smooth* as the Laplace one. The wave equation is a second order linear PDE that can be considered as vector value function v: $\mathbb{R}^n \to \mathbb{R}^m$; however, can be also a scalar wave equation, special case of the vector one that describe waves as vectors, e.g. electrical field, **E**. Hence, each (E_x, E_y, E_z) components of an electric vector field wave must satisfy the scalar wave equation, same for the magnetic field, **B**.

We will describe a special case, the scalar wave equation in free space without any sources or sinks as follows the general form:

$$\frac{\partial^2 u(\vec{\mathbf{x}}, t)}{\partial t^2} = c^2 \left(\frac{\partial^2 u(x_1, t)}{\partial x_1^2} + \frac{\partial^2 u(x_2, t)}{\partial x_2^2} ... + \frac{\partial^2 u(x_n, t)}{\partial x_n^2} \right) = c^2 \nabla^2 u(x, t) \tag{55}$$

Solution to the wave equation in one dimension, by the *d'Alembert theorem*, can be seen as linear combination of sinusoidal plane waves, a possible simple solution, with various directions and wavelength of propagation; nevertheless all with the same propagation speed c. The wave equation intrinsic linearity shows that any multiple of a solution or sum of any two solutions is a solution. *Superposition principle* follows by setting *initial conditions* at time t, such as standing waves or harmonics.

General solution to the wave equation in 1D:

$$u(x, t) = f(x - ct) + g(x + ct) \tag{56}$$

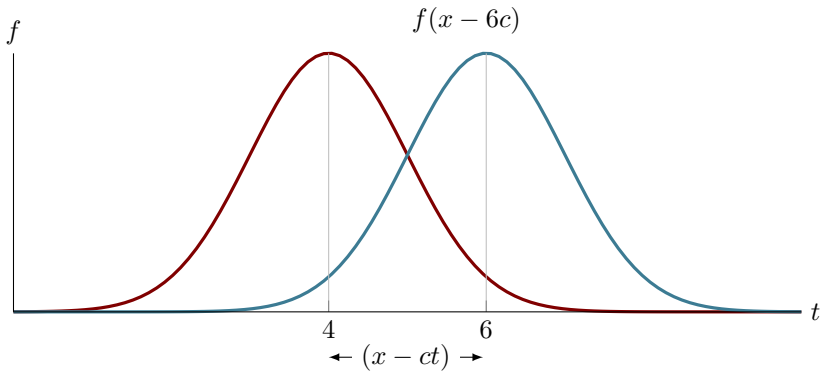


Figure 32. Wave form $u(x, t) = f(x - ct)$ travelling towards the right at t=4 and t=6, respectively, while towards the left for $u(x, t) = g(x + ct)$.

where t is the time and $f(x - ct)$ defines a waveform travels to the right along the x-*axis* with speed c, while $g(x + ct)$ is the component of the waveform travelling to the left side as shown in Eq. (56).

Taking into account the one dimensional case:

$$\frac{\partial^2 u(x, t)}{\partial t^2} = c^2 \frac{\partial^2 u(x, t)}{\partial x^2} \qquad \text{1D case} \tag{57}$$

We want to prove that Eq. (56) is a solution to the scalar wave equation, in addition, we want to prove that it is also valid for a general case where f and g are arbitrary functions w.r.t new variables as follows:

$$\left.\begin{aligned}\xi &= x + ct\\ \eta &= x - ct\end{aligned}\right\} \quad \text{two new variables w.r.t } \xi \text{ and } \eta \tag{58}$$

We express the wave equation by using derivatives w.r.t ξ and η in place of x and t. first partial derivatives follows:

$$\left.\begin{aligned}\frac{\partial u(x,t)}{\partial t} &= \frac{\partial u}{\partial \xi}\frac{\partial \xi}{\partial t} + \frac{\partial u}{\partial \eta}\frac{\partial \eta}{\partial t}\\ \frac{\partial u(x,t)}{\partial x} &= \frac{\partial u}{\partial \xi}\frac{\partial \xi}{\partial x} + \frac{\partial u}{\partial \eta}\frac{\partial \eta}{\partial x}\end{aligned}\right\} \quad 1^{st} \text{ partial derivatives}$$

Chain rule for both 1^{st} and 2^{nd} partial derivatives.

$$\left.\begin{aligned}\frac{\partial}{\partial t}\left(\frac{\partial u(x,t)}{\partial t}\right) = \frac{\partial^2 u(x,t)}{\partial t^2} &= c^2\left(\frac{\partial^2 u(x,t)}{\partial \xi^2} - 2\frac{\partial^2 u(x,t)}{\partial \eta \partial \xi} + \frac{\partial^2 u(x,t)}{\partial \eta^2}\right)\\ \frac{\partial}{\partial x}\left(\frac{\partial u(x,t)}{\partial x}\right) = \frac{\partial^2 u(x,t)}{\partial t^2} &= c^2\left(\frac{\partial^2 u(x,t)}{\partial \xi^2} + 2\frac{\partial^2 u(x,t)}{\partial \eta \partial \xi} + \frac{\partial^2 u(x,t)}{\partial \eta^2}\right)\end{aligned}\right\} \quad 2^{nd} \text{ partial derivatives}$$

considering the wave equation and *Schwarz-Clairaut's theorem*, we replace the above second partial derivatives into Eq. (57),

$$\begin{aligned}\frac{\partial^2 u(x,t)}{\partial t^2} &= c^2\frac{\partial^2 u(x,t)}{\partial x^2}\\ 2\frac{\partial^2 u(x,t)}{\partial \eta \partial \xi} &= -2\frac{\partial^2 u(x,t)}{\partial \eta \partial \xi}\\ 4\frac{\partial^2 u(x,t)}{\partial \eta \partial \xi} = 0 &= \frac{\partial}{\partial \eta}\left(\frac{\partial u(x,t)}{\partial \xi}\right) \iff\\ \iff \frac{\partial u(x,t)}{\partial \xi} = 0 &\iff u(x,t) = constant\\ \therefore \quad &\textbf{wave equation holds} \quad \square\end{aligned}$$

Hence, we proved that the equation is satisfied and the equality holds for the general form

$$u(x,t) = f(\xi) + g(\eta) = f(x - ct) + g(x + ct)$$

,

A.2 Index of Refraction, n

In optics, the refractive index or refraction index $n = c/v$ of a material is a dimensionless number that describes how fast light travels through the material, where c is the speed of light in vacuum and v is the phase velocity of light in the medium. An example is the refractive index of water, 1.333, meaning that light travels 1.333 times slower in water than in a vacuum. Increasing the refractive index corresponds to decreasing the speed of light in the material.

The refractive index determines how much the path of light is bent, or refracted, when entering a material. This is described by Snell's law of refraction as in Fig.5, $n_1 \sin\theta_1 = n_2 \sin\theta_2$, where θ_1 and θ_2 are the angles of incidence and refraction, respectively, of a ray crossing the interface between two media with refractive indices n_1 and n_2. The refractive indices also determine the amount of light that is reflected when reaching the interface, as well as the critical angle for total internal reflection, their intensity (Fresnel's equations) and Brewster's angle [14].

For an incoming EM light source, refraction implies the Snell's Law, which states that, for a given a pair of diverse media, the ratio of the sines of the incident θ_1 and reflected θ_2 source is equal to the ratio of phase velocities $\frac{v_1}{v_2}$ in the two media, or the ratio of refraction indices, $\frac{n_2}{n_1}$ of the two media. In other word [16],

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{v_1}{v_2} = \frac{n_2}{n_1}$$

Index of refraction can be seen as the factor by which the speed and the wavelength of the radiation are reduced with respect to their vacuum values: the speed of light in a medium is $v = c/n$, and similarly the wavelength in that medium is $\lambda = \lambda_0/n$, where λ_0 is the wavelength of that light in vacuum. This implies that vacuum has a refractive index of 1, and that the frequency ($f = v/\lambda$) of the wave is not affected by the refractive index. As a result, the perceived colour of the refracted light to a human eye, which depends on the frequency, is not affected by the refraction or the refractive index of the medium [16].

Index of refraction varies with wavelength; this causes white light to split into constituent colours when refracted, known as dispersion. This dispersion effect can be observed in prisms and rainbows, and as chromatic aberration in lenses. Light propagation in absorbing materials can be described using a complex-valued refractive index. The imaginary part then handles the attenuation or transmission, while the real part accounts for refraction. For most materials the refractive index changes with wavelength by several percent across the visible spectrum. Nevertheless, refractive indices for materials are commonly reported using a single value for n, typically measured at 633 nm [16].

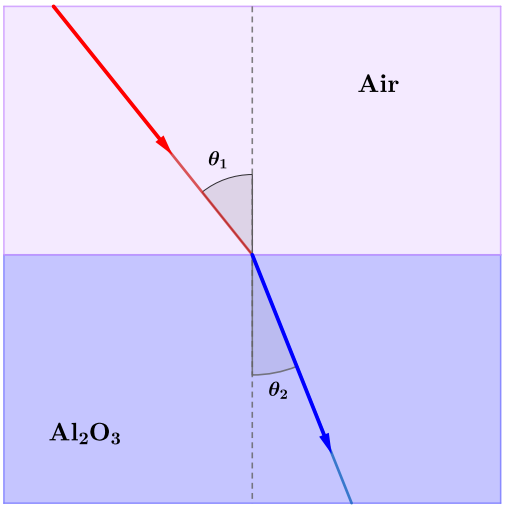


Figure 33. Representation of light travels via two distinct media, e.g. air to Al_2O_3, where its speed changes causes the rays to bend. This phenomenon, called refraction, is governed by Snell's Law, which states that the ratio of the sine of the angle of incidence to that of the angle of refraction is a constant, namely the index of refraction, n.

A.3 Lagrange Polynomial

Lagrange interpolation has been used to interpolate the extracted data-points from the receiver simulation. These data-points are used to study the absorption difference between two distinct depletion regions of a Receiver NW. Follows definition and a simple example of the Lagrange interpolation polynomial employed during our simulation process.

Problem Determine a polynomial p of degree smaller or equal to three that fulfils $p(-1) = 2$, $p(0) = 6$, $p(2) = 4$ and $p(3) = 30$ in the Lagrange basis.

Solution After proving the existence and uniqueness of the Lagrange theorem as suggested in *Mayers book* [13], we can construct the polynomial as follows:

$$p_n(x) = \sum_{k=0}^{n} L_k(x) y_k$$

where the L_k(x) is the Lagrange interpolation polynomial, namely:

$$L_k(x) = \prod_{\substack{i=0 \\ i \neq k}}^{n} \frac{(x - x_i)}{(x_k - x_i)}$$

We know that our nodes or interpolation points are x_0 = -1, $x_1 = 0$, $x_2 = 2$, and $x_3 = 3$.

As n=2 from p_n, we observe that

$$\begin{aligned} L_0(x_0) &= \frac{(x - x_0)}{(x_0 - x_0)} \frac{(x - x_1)}{(x_0 - x_1)} \frac{(x - x_2)}{(x_0 - x_2)} \frac{(x - x_3)}{(x_0 - x_3)} \\ &= \frac{(x - 0)}{(-1 - 0)} \frac{(x - 2)}{(-1 - 2)} \frac{(x - 3)}{(-1 - 3)} \\ &= -x \frac{(x - 2)}{3} \frac{(x - 3)}{4} \end{aligned}$$

Hence, if we evaluate L_0 for x_0 = -1 then we get 1, as follows [13]:

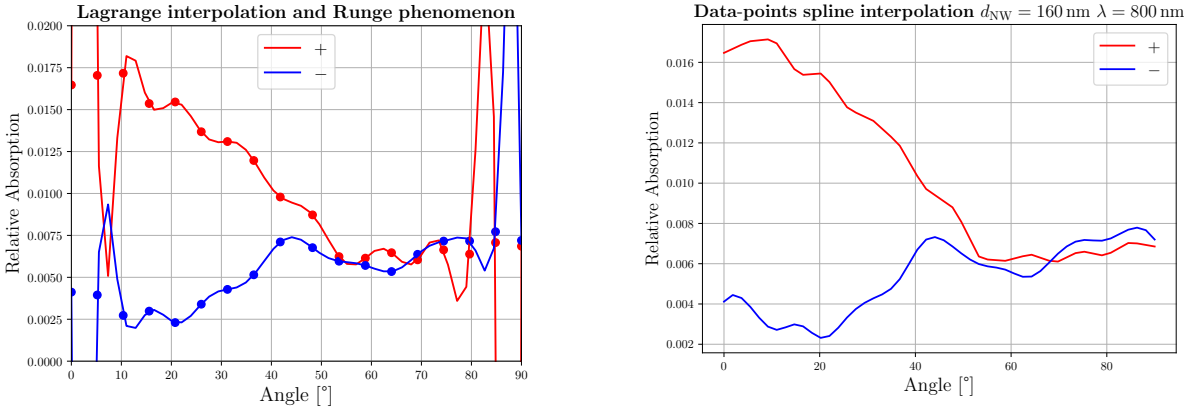
$$L_k(x_i) = \begin{cases} 1, & i = k, \\ 0, & i \neq k, \end{cases}$$

In a similar way is computed the other $L_k(x_i)$ which we show the final result:

$$\begin{aligned} L_1(x_1) &= \frac{(x + 1)(x - 2)(x - 3)}{6} \\ L_2(x_2) &= \frac{-x(x + 1)(x - 3)}{6} \\ L_3(x_3) &= \frac{x(x + 1)(x - 2)}{12} \end{aligned}$$

The resulting Lagrange polynomial is defined by

$$p_2(x) = \frac{-x(x - 2)}{3} \frac{(x - 3)}{4} \cdot 2 + \frac{(x + 1)(x - 2)(x - 3)}{6} \cdot 6 + \frac{-x(x + 1)(x - 3)}{6} \cdot 4 + \frac{x(x + 1)(x - 2)}{12} \cdot 30 \quad \blacksquare$$



(a) (b)

Figure 34. A comparison between two diverse interpolation polynomials for NW d=160 nm. **(a) the Lagrange interpolation polynomial & Runge phenomenon**, where high order polynomials yield two compelling oscillations at the side wings of the polynomials. **(b) A piecewise smooth interpolation polynomial** opted as alternative to the prior Lagrange to interpolate our experimental data-points.

Bibliography

[1]F. B. Fitch, "Mcculloch warren s. and pitts walter. a logical calculus of the ideas immanent in nervous activity. bulletin of mathematical biophysics, vol. 5, pp. 115–133", Journal of Symbolic Logic **9** (1944).

[2]M. Minsky, "Computers and thought, chapter steps towards artificial intelligence", (1963).

[3]C. Ramos, J. C. Augusto, and D. Shapiro, "Ambient intelligence—the next step for artificial intelligence", IEEE Intelligent Systems **23**, 15–18 (2008).

[4]C. Mead, "Neuromorphic electronic systems", Proceedings of the IEEE **78**, 1629–1636 (1990).

[5]N. Geddes, "The large hadron collider and grid computing", Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **370**, 965–977 (2012).

[6]T. Netter and N. Francheschini, "A robotic aircraft that follows terrain using a neuromorphic eye", in Ieee/rsj international conference on intelligent robots and systems, Vol. 1 (2002), 129–134 vol.1.

[7]J. W. Schnupp and A. J. King, "Neural processing: the logic of multiplication in single neurons", Current Biology **11**, R640–R642 (2001).

[8]J. Wallentin, N. Anttu, D. Asoli, M. Huffman, I. Åberg, M. H. Magnusson, G. Siefer, P. Fuss-Kailuweit, F. Dimroth, B. Witzigmann, et al., "Inp nanowire array solar cells achieving 13.8% efficiency by exceeding the ray optics limit", Science **339**, 1057–1060 (2013).

[9]S. Orfanidis, *Electromagnetic waves and antennas; sophocles j*, 2004.

[10]D. O. Winge, S. Limpert, H. Linke, M. T. Borgström, B. Webb, S. Heinze, and A. Mikkelsen, "Implementing an insect brain computational circuit using iii–v nanowire components in a single shared waveguide optical network", ACS photonics **7**, 2787–2798 (2020).

[11]A. R. Hambley, N. Kumar, and A. R. Kulkarni, *Electrical engineering: principles and applications* (Pearson Education (New Jersey), 2011).

[12]A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks (alexnet) imagenet classification with deep convolutional neural networks (alexnet) imagenet classification with deep convolutional neural networks",

[13]U. o. O. Endre Süli David F. Mayers, "An introduction to numerical analysis", in (Cambridge University Press, 2003).

[14]D. J. Griffiths, *Introduction to electrodynamics*, 2005.

[15]D. Poljak and M. Cvetković, "Chapter 2 - theoretical background: an outline of computational electromagnetics (cem)", in *Human interaction with electromagnetic fields*, edited by D. Poljak and M. Cvetković (Academic Press, 2019), pp. 21–52.

[16]F. L. Pedrotti, L. M. Pedrotti, and L. S. Pedrotti, *Introduction to optics* (Cambridge University Press, 2006).

[17]Lumerical, *Units and normalization*, https://kb.lumerical.com/solvers_optical_units.html, May 2021.

[18]J.-P. Bérenger, "A perfectly matched layer for the absorption of electromagnetic waves", Journal of computational physics **114**, 185–200 (1994).

[19]Lumerical, *Pml boundary conditions in fdtd and mode*, https://support.lumerical.com/hc/en-us/articles/360034382674-PML-boundary-conditions-in-FDTD-and-MODE, May 2021.

[20]D. M. Sullivan, *Electromagnetic simulation using the fdtd method* (John Wiley & Sons, 2013).

[21]K. Yee, "Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media", IEEE Transactions on antennas and propagation **14**, 302–307 (1966).

[22]L. Ibbotson, *The fundamentals of signal transmission: optical fibre, waveguides and free space* (Butterworth-Heinemann, 1998).

[23]Lumerical, *Maxwell's solver for 2d/3d nanophotonic devices*, https://www.lumerical.com/products/fdtd/, May 2021.

[24]J.-P. Bérenger, "Perfectly matched layer (pml) for computational electromagnetics", Synthesis Lectures on Computational Electromagnetics **2**, 1–117 (2007).

[25]S. D. Gedney and B. Zhao, "An auxiliary differential equation formulation for the complex-frequency shifted pml", IEEE Transactions on Antennas and Propagation **58**, 838–847 (2009).

[26]A. F. Oskooi, D. Roundy, M. Ibanescu, P. Bermel, J. D. Joannopoulos, and S. G. Johnson, "Meep: a flexible free-software package for electromagnetic simulations by the fdtd method", Computer Physics Communications **181**, 687–702 (2010).

[27]*Gnu general public license, version 3*, http://www.gnu.org/licenses/gpl.html, Last retrieved 2021-12-02, 2021.

[28] Lumerical, *Frequency-domain profile and power monitor - simulation object*, https://support.lumerical.com/hc/en-us/articles/360034902393-Frequency-domain-Profile-and-Power-monitor-Simulation-object, May 2021.
[29] Lumerical, *Understanding injection angles in broadband simulations*, https://support.lumerical.com/hc/en-us/articles/360034382894-Understanding-injection-angles-in-broadband-simulations, May 2021.
[30] D. van Dam, D. R. Abujetas, R. Paniagua-Dominguez, J. A. Sánchez-Gil, E. P. Bakkers, J. E. Haverkort, and J. Gómez Rivas, "Directional and polarized emission from nanowire arrays", Nano letters **15**, 4557–4563 (2015).
[31] *Creative commons*, https://creativecommons.org/licenses/by-sa/4.0/, Last retrieved 2021-12-02, 2021.