Attenuation and Photodetection of Sub-Bandgap Slow Light in Silicon-on-Insulator Photonic Crystal Waveguides

Attenuation and Photodetection of Sub-Bandgap Slow Light in Silicon-on-Insulator Photonic Crystal Waveguides

by

John L. Gelleta

B.Eng. University of Saskatchewan, 2010

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Master of Applied Science

McMaster University

MASTER OF APPLIED SCIENCE (2013) McMaster University
Hamilton, Ontario

TITLE: Attenuation and Photodetection of Sub-Bandgap Slow Light in Silicon-on-Insulator Photonic Crystal Waveguides

AUTHOR: John L. Gelleta

SUPERVISOR: Dr. Andrew P. Knights

NUMBER OF PAGES: 13, 146

Abstract

A glass-clad, slow-light photonic-crystal waveguide is proposed as a solution to sub-bandgap light detection in silicon photonic circuits. Such detection in silicon is perceived as a challenge owing to silicon's indirect band gap and transparency to 1550nm wavelengths, yet is essential for achieving low-cost, high-yield integration with today's microelectronics industry. Photonic crystals can be engineered in such a way as to enhance light-matter interaction over a specific bandwidth via the reduction of the group velocity of the propagating wave (i.e. the slowing of light). The interaction enhanced for light detection in the present work is electron-hole pair generation at defect sites. The intrinsic electric field of a *p*-*i*-*n* junction enables light detection by separating the electron-hole pairs as a form of measurable current. The photonic-crystal waveguides are designed to have bandwidths in the proximity of a wavelength of 1550nm. Refractive indices of over 80 near the photonic-crystal waveguide's Brillouin zone boundary are measured using Fourier transform spectral interferometry and are found to correspond to numerical simulations. Defect-induced propagation loss was seen to scale with group index, from 400dB$\cdot$cm^{-1} at a group index of 8 to 1200dB$\cdot$cm^{-1} at a group index of 88. Scaling was sublinear, which is believed to be due to the spreading of modal volume at large group index values. Photodetectors were measured to have responsivities as high as 34mA$\cdot$W^{-1} near the photonic-crystal waveguide's Brillouin zone boundary for a reverse bias of 20V and a remarkably short detector length of 80μm. The fabrication of each device is fully CMOS-compatible for the sake of cost-effective integration with silicon microelectronics.

Acknowledgements

This work would not have been possible were it not for the help of many people. First and foremost, I am deeply indebted to my supervisor, Dr. Andrew Knights, for accepting me as his graduate student, and for providing technical assistance to help me move forward in my work, financial assistance so that I could live with food in my stomach and a roof over my head, and moral assistance when I began to doubt my own capabilities. Above all, I thank him for his humour, which made working under his tutelage a most enjoyable experience.

I am indebted to many more people for helping me surmount technical challenges involved in my work. In no particular order, I thank Doris Stevanovic for sharing her knowledge and experience with fabrication procedures and for showing general concern for my safety, Zhilin Peng for instructing me on the use of various equipment – especially the sputtering system, whose operating procedure I forgot on multiple occasions – Matthew Minnick and Rob Vandusen for assistance with reactive-ion etching, and Jack Hendriks for assistance with ion implantation. I thank Dr. William Whelan-Curtin and Dr. Kapil Debnath for hosting me in beautiful St. Andrews for a spell, dispensing valuable knowledge about photonic crystals, and assisting me with the fabrication and characterisation of my devices. I owe a special thanks to Dr. Édgar Huante-Cerón and Jason Ackert, who guided me all throughout my work with knowledge of silicon photonics, experimental procedures, and vendors, to name a few.

The most lasting memories of my experience at McMaster University, next to the long, frustrating nights spent in the clean room, are the friends with whom I made contact. Thus, I would like to thank Dr. Édgar Huante-Cerón and Dixon Paez for exposing me to fragments of Latino culture, Jason Ackert for bonds forged over a mutual taste in music and for revealing that I, a Canadian, had no idea who Neil Young was, and Abhi Rampal for teaching me the value of not being overly formal in correspondence. I would like to further thank these individuals, as well as Dr. Chris Brooks, Dr. Dylan Logan, Thomas Jacob, Kyle Murray, Ross Anthony, Jiamin Zheng, and Zhao Wang, for the warm conversations we shared. Apart from Dixon's

solar simulator, our conversations were the brightest things present in Dr. Knights' crowded, windowless laboratory.

Finally, I must thank my family for their selfless emotional and financial support throughout life, especially my mother, who has always been my biggest supporter, and my father for leading by example and for constantly informing me, from an early age, that – yes, indeed – the square root of two is one-point-four-one-four.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis explores the slow-light enhancement of absorption of light of around 1550nm wavelengths in silicon-based glass-clad photonic-crystal waveguides. The photonic-crystal waveguides are designed to support guided modes with low group velocities at around 1550nm wavelengths. The absorption spectrum of silicon is extended to include 1550nm-wavelength photons via the controlled implantation of defects. Two photonic devices are considered in the study of slow-light enhancement: a waveguide attenuator, featuring a photonic-crystal waveguide implanted with defects, and a photodetector, featuring a photonic-crystal waveguide implanted with defects and integrated with a *p*-*i*-*n* diode. The waveguide attenuator is used to study the slow-light enhancement of total absorption, whereas the photodetector is used to study the slow-light enhancement of photocurrent. Each device is sensitised to 1550nm-wavelength photons to demonstrate its applicability with telecommunications. In an effort to exploit the low-cost, high-yield advantages of modern silicon manufacturing technology, fabrication of each device is fully CMOS-compatible.

1.1 The Need for Silicon Photonics

Photonics is the technology concerned with the generation, modulation, processing, transmission, and detection of photons, or light. It is the analogue of electronics with electrons. One of the most successful examples of photonics to date is fiber-optic systems. Fiber optics permit low-loss transmission of information at high bandwidth (gigabits per second) across hundreds of kilometres without need for signal regeneration, to an extent with which metal wires cannot compete. Low-loss transmission is enabled by the transparency of the material (silica) to the transmitted photons, the guiding mechanism (total internal reflection), and the general lack of optical cross-talk, while high bandwidth is enabled by light's high frequency (THz) and ability to be multiplexed. These properties give photonics distinct advantages over electronics.

Recent research endeavours to transfer these advantages to smaller systems. Anticipated applications on small scales include lab-on-a-chip biosensing, bio-medicine, and optical networking [1–3]. Perhaps the most anticipated application of all, however, is chip-level optical interconnection [3–7]. The unprecedented improvement of high-performance computational systems over the last few decades has been driven by scaling down the feature size of these systems' electrical components. Doing so reduces the RC constants and power consumption of logical elements and permits a higher bandwidth density, all the while keeping the systems compatible with existing VLSI fabrication techniques. However, since the 1980s, it has been evident that electrical interconnects, which, at the time of writing, are still used in modern high-performance processors, are faced with a serious scaling limitation: as an electrical wire is scaled down in size, its RC constant does not change at all [8]. This spells trouble for the fully electrical paradigm, for it places a hard limit on bit-rate capacity of electrical interconnects irrespective of the system's architectural aspect ratio [9]. This scaling limit, known as the *communication bottleneck*, originates from the resistive properties of electrical wires. Optical interconnection is seen as a solution to this bottleneck because the propagation of information along optical waveguides experiences no such resistivity.

Compounding the bottleneck problem is the problem of power dissipation. In addition to scaling, performance improvements were traditionally driven by increasing the system's clock speed. However, with an increase in clock speed comes a concomitant rise in power dissipation due in part to the more frequent discharge of the system's capacitive elements. By 2005, on-chip power density approached $100\mathrm{W}\cdot\mathrm{cm}^{-2}$ [5]. The challenge this alarming increase in heat generation posed to packaging technology became so severe that the microprocessor industry shifted away from increasing clock speed in favour of distributing the computational workload among a larger number of cores [10]. Unfortunately, doing so places greater strain on chip interconnection. Maintaining computational synchronicity between n cores requires $n(n-1)/2$ channels of communication. In fact, it is estimate that already 40% of the power a chip consumes is used to move information on and off the chip; increasing the number of cores is expected to push this value higher [7]. The communication requirements of this hitherto unseen parallelisation of workload serve only to further underscore the importance of shifting interconnection technologies to an optical paradigm.

Optical interconnection must meet certain milestones in order to be superior to electrical interconnection [6]. For superior bandwidth density, its elements must be densely integrated. Electrical interconnection holds a distinct advantage over optical interconnection in this regard due to the latter being diffraction limited. Fortunately, optical interconnection can get around this difficulty via multiplexing – a technique unique to the optical paradigm. For a superior power-delay product, its elements must make efficient use of power (which goes hand-in-hand with alleviating the power dissipation problem) and must convert electrical signals to optical signals and vice-versa within hundreds of picoseconds. Optical signals propagate much faster within waveguides than electrical signals do through wire due to smaller impedance. However, there is a fixed time delay associated with generating and detecting optical signals. Thus, the superiority of optical interconnection in terms of the power-delay product is felt most strongly over long, global-scale interconnects, as opposed to local-scale ones. Finally, in order to share the low-cost, high-yield benefits of VLSI technology, the fabrication of optical interconnects must conform to the same monolithic techniques used in microelectronics. Such conformity is often described as *CMOS compatibility*, named after the fabrication techniques employed in making complementary metal-oxide (CMOS) transistors.

Conforming to VLSI fabrication techniques may well be the most important hurdle for making optical interconnection affordable. This hurdle is what stimulates the wealth of research devoted to silicon-based photonics. By sharing a common material with microelectronics, silicon-based photonics readily integrates with the microelectronics industry's production line. Silicon also has attractive material properties for photonic devices, including high thermal conductivity, a high optical damage threshold, high third-order nonlinearities, and high transparency to photons of wavelengths between 1.1μm and nearly 7μm [3]. Silicon also has a stable oxide, which may be used to protect devices from performance-degrading surface contamination [11], reduce propagation losses in waveguides [12], and provide general mechanical stability. In addition, silicon wafers are the cheapest to manufacture and are of the highest crystal quality among all semiconductors. To its detriment, silicon has several noticeable drawbacks. It has an indirect bandgap, so creating a silicon-based light source is a tremendous challenge; it does not exhibit the Pockel's effect, so silicon-based modulators must exploit the less-efficient thermo-optic or plasma dispersion effects; also, there exists no conventional means of detecting photons of wavelength in the telecommunication band (1.3μm-1.6μm) [13, 14].

The biggest limitation on the success of photonics might be laws not of physics, but of economics. Indeed, research into silicon photonics is challenged to produce devices that are not only efficient, but also economical in terms of fabrication, packaging, production volume, and market demand [15]. Nevertheless, this very challenge is what makes silicon attractive in spite of being technically inferior to other popular materials, most noticeably III-V semiconductors such as gallium arsenide (GaAs) and indium phosphide (InP). By borrowing the decades of capital invested in the microelectronics industry, silicon photonics may deliver affordable solutions to existing problems in the shortest time frame.

1.2 Silicon-on-Insulator (SOI)

The term "Silicon-on-Insulator" (SOI) describes a material structure consisting of a thin layer of crystalline silicon (Si) overlaying an insulating layer, which is typically silicon dioxide (SiO_2). An illustration of SOI's cross-section is provided in figure 1.1. Photonic devices are formed by selectively etching away the top silicon layer, also known as the *device layer*, using lithography techniques. The buried-oxide (BOx) layer is at least 1μm thick to prevent evanescent coupling between the device layer

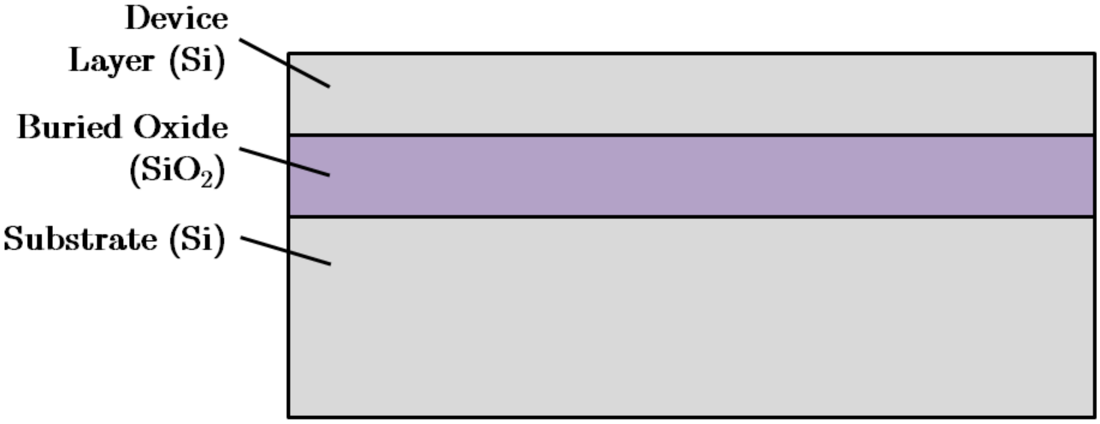


Figure 1.1: Illustrated cross-section of a silicon-on-insulator (SOI) wafer. The thickness of the device layer typically ranges from hundreds of nanometres to a few micrometres. The thickness of the buried oxide layer is typically one to two micrometres.

and the underlying silicon substrate. SOI is attractive for photonics because of the transparency of silicon to infrared photons and the high refractive index contrast between silicon and its oxide (3.48 versus 1.45, respectively). This high contrast strongly confines optical modes within the device layer, which allows optical waveguides with cross-sectional areas on the order of $0.1\mu m^2$ to successfully guide infrared photons. Waveguides of such small cross-sectional areas are important for increasing bandwidth density in applications. Notably, SOI is also attractive for high-performance microelectronics [16]. The coincidence of these attractive properties makes SOI a preferable material for realising competitive, CMOS-compatible integrated circuits.

Three common methods for manufacturing SOI are *Separation by Implantation of Oxygen* (SIMOX), *Bond and Etch-Back* (BESOI), and the *Smart-Cut*® process [17]. SIMOX involves the implantation of a high dose ($> 10^{18}cm^{-2}$) of oxygen ions below the surface of a silicon wafer. The wafer is kept at a high temperature (~600 °C) during implantation to prevent the formation of an amorphous silicon overlayer. The insulating SiO_2 layer is then formed by annealing the implanted wafer at 1300 °C for several hours. This annealing step also helps remove crystal dislocations that invariably form during implantation. By implanting at a high dose, the concentration profile of the implanted oxygen ions saturates, ensuring that the SiO_2 layer forms with abrupt edges and uniform density. The oxide layer's depth and thickness are controlled by varying the implantation energy. There is a technical limit to how deeply oxygen ions can be implanted into silicon, so if a thicker device layer is desired, then annealing may be followed by epitaxial growth of additional silicon.

BESOI involves bonding two thermally oxidised silicon wafers together, then thinning one wafer to form the device layer. Wafer bonding is performed by coating the surface of each wafer in hydroxyl groups, then placing the wafers in contact at room temperature. The hydroxyl groups chemically bond to residual water molecules on each surface to hold the two wafers together. Heating the bonded wafers up to 1100 °C diffuses away the water and forms siloxane bonds, which fuse the wafers together. The process is finished by etching away the silicon of one wafer to form a device layer of a target thickness. For thick (10μm-100μm) device layers, etching via chemical mechanical polishing (CMP) suffices. For thinner layers, a combination of CMP followed by selective etching is used, due to the difficulty in maintaining thickness uniformity with polishing alone. Selective etching requires the introduction of an *etch-stop*, typically boron or germanium, which unfortunately contaminates the final device layer.

The *SmartCut*® process combines steps in both SIMOX and BESOI. A high dose ($\sim 10^{17}\text{cm}^{-2}$) of hydrogen ions are implanted into the silicon of a thermally oxidised silicon wafer, known as a *seed wafer*. Implantation depth beneath the thermal oxide is controlled by the implantation energy; the depth is typically between a few hundred nanometres and a few microns. The presence of hydrogen, together with implantation-induced crystal dislocations, weakens bonds in the silicon lattice. Following the implant, a second wafer, known as a *handling wafer*, is chemically bonded to the thermal oxide layer using techniques employed in BESOI. An initial thermal process at 600 °C splits the first wafer along the peak of the hydrogen concentration profile. The resulting SOI structure consists of the handling wafer, thermal oxide layer, and a thin layer of silicon beyond which hydrogen was implanted and which becomes the device layer. A second thermal process at 1100 °C fuses the handling wafer to the thermal oxide layer as in BESOI. Subsequent CMP reduces the surface roughness of the device layer. *SmartCut*®, like SIMOX, provides flexible control of the device layer's thickness, but it also has the added advantage of being able to reuse its seed wafer. Since no part of the handling wafer forms the device layer, the handling wafer may be of lower crystalline quality than the seed wafer, allowing for significant reduction in production costs. Also, *SmartCut*® does not involve as much CMP as in BESOI, so device layers are of more uniform thickness than BESOI. For even more uniform thickness, CMP may be substituted with selective etching [18]. Due to these advantages, *SmartCut*® is the preferred method for manufacturing SOI.

1.3 Integrated Sub-bandgap Photodetection in SOI

The focus of this thesis is the investigation of novel methods for photodetection in a silicon waveguide. Photodetectors are devices that measure the intensity of incident light by converting photons into an electrical signal. In semiconductors such as silicon, the conventional way of performing this conversion is by having photons excite electrons across the semiconductor's electrical band gap, whereupon the photon is absorbed and the electron becomes mobile under an applied electric field. For compatibility with telecommunications, this conversion must be possible for photons of 1.3μm-1.6μm wavelengths. Unfortunately for silicon, photons of these wavelength have insufficient energy to enable the excitation: A photon of 1.3μm wavelength has an energy of 0.95eV, whereas the band gap energy of silicon at a temperature of 300K is 1.12eV. (Photons of longer wavelengths have even lower energies.) Photons whose energies lie below silicon's band gap energy are fittingly called "sub-bandgap" photons. Indeed, silicon's inability to absorb sub-bandgap photons is the very reason why silicon waveguides may guide these photons with low propagation loss [19]. For efficient photodetection, however, this transparency clearly must be broken.

There are several methods for detecting sub-bandgap photons in silicon waveguides. One method is to integrate a silicon waveguide with a III-V semiconductor and have the III-V semiconductor perform the detection [20, 21]. III-V semiconductors such as GaAs and InP make for suitable photodetectors because of their tunable band gap energy and relatively high carrier mobility. However, they are regarded as an expensive solution because they do not naturally occur in nature and are difficult to integrate with the CMOS production line. In hope of making silicon photonics more affordable, a considerable amount of research over the past decade has looked past III-V hybrids in search of a photodetector that is both CMOS-compatible and monolithic. This research includes the development of strained Si-Ge heterostructures [22–24], the introduction of optical dopants, such as Erbium, into silicon [25], and the introduction of defects with mid-bandgap energy levels into silicon [26, 27]. The third approach is used here.

Since the 1950s, it has been known that irradiated silicon absorbs photons of wavelengths longer than those prescribed by its band gap energy [28]. When silicon is irradiated with high-energy ions, silicon atoms are dislodged from lattice sites, forming various vacancy- and interstitial-type defects. A subset of these defects are stable at room temperature. These stable defects introduce foreign energy states into the electronic band structure of silicon. The states that happen to lie within silicon's band gap permit resident electrons to exist at energy levels that are normally forbidden. The energy required of a photon to excite electrons to and from these foreign energy states is smaller than the energy required for conventional band-to-band excitation. Thus, mid-bandgap defect states serve as a conduit through which sub-bandgap photons may create mobile electrons.

Following the irradiation of silicon, various absorption peaks indicating the presence of mid-bandgap defect states are seen. Of these, peaks centred at $1.8\,\mu$m, 3.3μm, and 3.9μm are correlated with a divacancy-type defect, which is a pair of adjacent empty lattice sites [29]. Photoconductivity [30, 31], infrared absorption [29, 32], and deep-level transient spectroscopy [33, 34] studies indicate that this defect has three mid-bandgap energies: 0.31eV above the valence (lower) band, 0.41eV below the conduction (upper) band, and 0.23eV below the conduction band, with each energy corresponding to a separate charge state of the divacancy (neutral, singly negative, and doubly negative, respectively). The 3.3μm and 3.9μm absorption peaks are attributed to the excitation of electrons to divancancies of these charge states [30,31,34], while the 1.8μm absorption peak is suggested to be an internal transition within the neutrally charged state [32]. A separate study has shown that the divacancy is stable at room temperature and is rapidly annealed out at temperatures above 300 °C [35]. Thus, the divacancy is expected to play a prominent role in practical

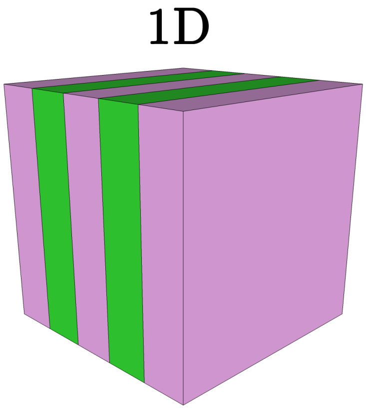


Figure 1.2: Simple examples of photonic crystals with one-, two-, and three-dimensional periodicity. The two colours represent materials with different dielectric constants.

room-temperature absorption of sub-bandgap photons.

1.4 Photonic Crystals

A photonic crystal is a low-loss periodic dielectric medium that affords control of light's modal dispersive properties – that is, the frequencies ω and wave vectors $\mathbf{k}$ of light propagating through the medium. A photonic crystal may be periodic in one, two, or three dimensions, and it must consist of at least two materials with different dielectric constants. Simple examples of photonic crystals are shown in figure 1.2. Every medium supports a number of modes describing the frequencies and wave vectors, or directions, at which light may propagate. In freespace, light may propagate at any frequency and in any direction. In a photonic crystal, however, light is typically prohibited from propagating at all but at a discrete set of frequencies in a given direction. The reason for this prohibition is that the light wave is partially reflected at each of the crystal's periodic dielectric interfaces, and these multiple reflections interfere destructively to eliminate the forward-propagating wave. This phenomenon has been known since the nineteenth century, when Lord Rayleigh published one of the first analyses of the optical properties of multilayer films – a one-dimensional photonic crystal. Thus, a photonic crystal is a medium that constrains light's propagation via the phenomenon of *coherent scattering*.

For certain periodic arrangements, light within a certain frequency range may be prohibited from propagating in *any* direction. By way of analogy with solid-state theory, these arrangements are said to exhibit a **photonic band gap** (PBG). No photon is supported within a PBG, just as no electron is supported within an electronic band gap. If light of a frequency within a crystal's PBG is incident on the crystal at *any* angle, then the light will be totally reflected. This phenomenon lies in stark contrast with that of total internal reflection, which occurs only if light approaches an interface from within the higher dielectric medium and at an incident angle larger than a given critical angle. Note that photonic crystals do not exhibit PBGs in directions along which there is no periodicity, as there exists no coherent scattering phenomenon to constrain modes that propagate along these directions.

The PBG is a novel feature of photonic crystals and may be exploited for a variety of applications. Commercialised examples include fiber Bragg gratings, which are a type of one-dimensional crystal. Fiber Bragg gratings are used in fiber-optic systems to add (drop) light of specific frequencies to (from) a fiber-optic cable via PBG-assisted reflection. Also, a fiber-optic cable itself may consist of a two-dimensional photonic crystal with a solid or hollow core. Light of frequencies within the PBG of this crystal are constrained to propagate along the core, just as a conventional fiber constrains light along its silica core. The novelty of a photonic-crystal fiber, however, is that its core may be of a lower refractive index than its cladding, which may curtail unwanted material-dependent effects, and that light may be more tightly confined to its core via appropriate design of the crystal's dispersive properties.

For complete control of light's dispersive properties, a photonic crystal must be periodic in three dimensions. Unfortunately, fabricating three-dimensional crystals with minimal defects is a huge technical challenge. As a compromise, many research groups are making use of the enormous capital invested in the microelectronics industry to fabricate **photonic-crystal slabs** out of SOI or III-V materials using standard lithographic techniques. Photonic-crystal slabs are hybrid structures that combine PBG confinement in-plane with total internal reflection (TIR) out-of-plane. The cost of relying on TIR for confinement in the third dimension is that modes of sufficiently

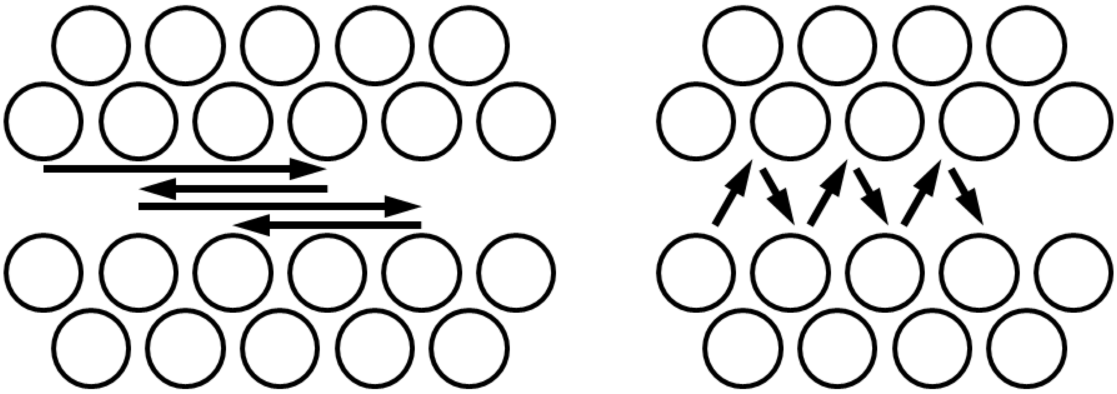


Figure 1.3: Two mechanisms for slow-light propagation in a photonic crystal waveguide: (left) coherent back-scattering, and (right) near-normal propagation. Adapted from [40].

high frequencies extend out-of-plane, or, alternatively, *leak* out of the crystal. Photonic circuits can be made out of crystal slabs by perturbing the crystal's dielectric periodicity along lines to make waveguides or around points to make resonant cavities. Via appropriate perturbation, waveguides may support modes that propagate with extremely low group velocities [36–39]. This so-called *slowing of light* is a consequence of one of two mechanisms: one, light coherently scatters backward nearly as much as it does forward, and, two, light propagates at near-normal incidence. These mechanisms are illustrated in figure 1.3. The consequential compression of modal energy density may enhance light-matter interactions within the waveguide [40, 41]. Resonant cavities, on the other hand, have been demonstrated to have ultra-high Q values [42,43], which help enhance photoluminescence in silicon [44] and achieve ultra-low-threshold quantum-dot lasing in III-V material systems [45–47]. Dynamic tuning of a resonant cavity's Q value leads to high-order functionality such as modulation of quantum-dot emission [48] and all-optical RAM [49]. What makes each of these functions possible in spite of imperfect out-of-plane confinement is careful design of the photonic crystal's dielectric pattern.

The novelty of this thesis is the integration of an SOI-based photonic-crystal waveguide (PCW) with a sub-bandgap photodetector. The structure, which is illustrated in 1.4, consists of a *p*-*i*-*n* junction straddling the PCW. The *p*-*i*-*n* junction

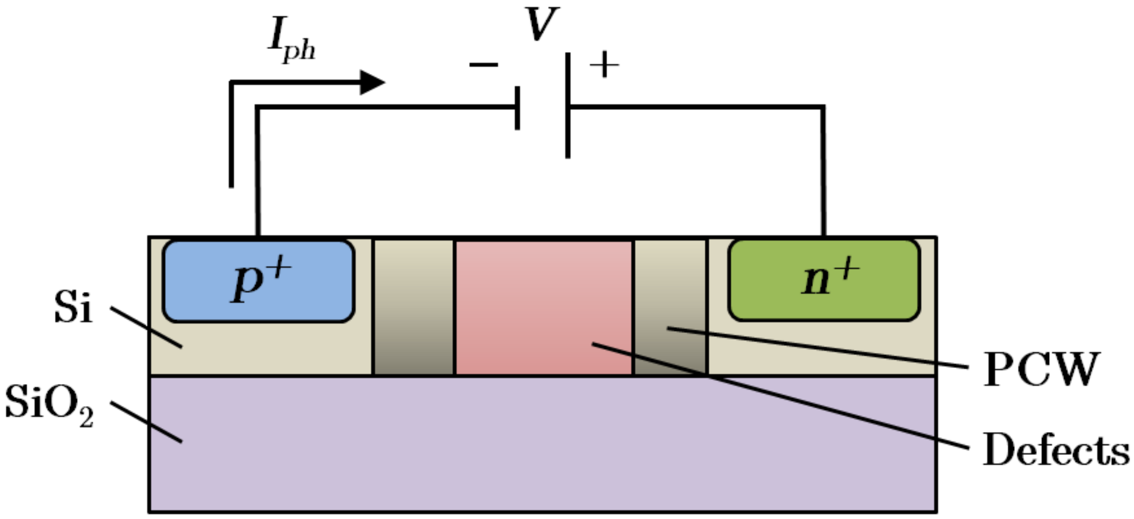


Figure 1.4: Illustrated cross-section of the SOI-based photodetector to be studied in this thesis. A photocurrent I_{ph} is generated by light propagating through the photonic-crystal-slab waveguide (PCW) and is increased in magnitude by bias potential V. p^+ and n^+ denote p- and n-regions that are heavily doped.

exhibits a built-in electric field between the p- and n-doped regions. This electric field is aligned perpendicularly to the PCW's axis. The photodetector is sensitised to sub-bandgap photons via the controlled implantation of defects into an area concomitant with that of the guided mode. Charge carriers generated at defect sites are swept into the p- and n-regions by the electric field and, thereby, form a measurable current. Current may be enhanced by applying a bias across the p-i-n junction so as to strengthen the electric field. The PCW is designed to support modes with low group velocities in attempt to enhance the responsivity of the device via spatial compression of modal energy density. By enhancing the responsivity so, this photodetector may be made shorter than conventional rib-waveguide photodetectors, such as those featured in [26], upon which this work is based. Reducing waveguide length may prove useful for dense integration of photodetectors with microelectronic circuits.

1.5 Thesis Outline

This thesis demonstrates two phenomena: the slow-light enhancement of propagation loss in a waveguide attenuator and of responsivity in a sub-bandgap photodetector. Chapter 2 begins with an introduction to photodetection in silicon. Discussions of electronic bands in semiconductors and photo-excitation of electrons between these bands are presented. These discussions are followed by a brief treatment of photodiode theory involving *p*-*n* junctions. Central to the operation of the photodetectors featured in this work is the mechanism of defect-mediated carrier generation, which is explained next. Chapter 2 concludes with a list of performance metrics for photodetectors.

Chapter 3 is devoted to photonic crystal theory. Explanations are provided from the ground up, starting with a treatment of Maxwell's equations for mixed dielectric media, then moving on to an application of these equations to systems with one-dimensional and two-dimensional periodicity, in that order. Maxwell's equations are cast as a Hermitian problem, from which useful properties, such as the orthogonality of modes, the electromagnetic variational theorem, and scaling laws, are demonstrated. A one-dimensional system is presented to discuss the physical origin of the photonic bandgap (PBG), a novel feature of photonic crystals. A two-dimensional system is presented to introduce the concept of polarisation sensitivity and to lead the reader toward a discussion of the system featured in this work: a hybrid slab structure combining PBG confinement in two dimensions and total internal reflection in the third. Particular attention is paid to the *band structure*, which summarises a system's dispersive properties, throughout the discussion of each of these systems. The slowing of light is introduced as a natural consequence at particular regions within these band structures. Chapter 3 concludes with an explanation of Fourier transform spectral interferometry, an experimental technique for measuring the group velocity of modes in waveguides.

Chapter 4 documents the designs and fabrication processes of the waveguide attenuator and photodetector. For each device, the design process, which involves the prediction of photonic crystal properties using MIT Photonic-Bands simulation software, is provided in detail. Section 4.1 focuses on the design and fabrication of the waveguide attenuator. Section 4.2 focuses on the design and fabrication of the photodetector.

Experimental results are provided in chapter 5. The optical response of the waveguide attenuator is presented first. The electrical, optical, and photocurrent responses of the photodetector are presented second. Prefacing the discussion of each device is a description of the experimental setup used to measure data.

Chapter 6 concludes this thesis with a summary of important results from the preceding chapter. A discussion of possible improvements to the photodetector are included.

Chapter 2

Photodetector Theory

This chapter provides theory critical to understanding the operation of a silicon-based photodetector. Section 2.1 discusses the mechanism by which and the rate at which silicon converts photons into charge carriers. Section 2.2 provides a detailed treatment of p-n junction theory as it relates to detecting these charge carriers. Section 2.3 discusses the electrical characteristics of p-n junctions in the presence of light, in what is known as *photodiode* operation. Section 2.4 discusses how carrier generation via deep-level trapping can be exploited to sensitise silicon photodiodes to sub-bandgap photons. Section 2.5 concludes this chapter with a summary of some important photodetector performance metrics. The theory presented in this chapter is well-established and can be found in [50–53].

2.1 Photons in Semiconductors

Central to the discussion of photodetection in semiconductors is the concept of *energy bands*. Atoms in solid-state materials are bonded to each other at quantum-length proximity, meaning that they cannot be treated as individual entities. Valence electrons are not bound to any single atom. Rather, they belong to the ensemble of atoms, hereafter referred to as the *crystal lattice*. When solving the Schrödinger equation for electron energy, the periodic potential of the crystal lattice results in a splitting of atomic energy levels and the formation of energy bands. Each band

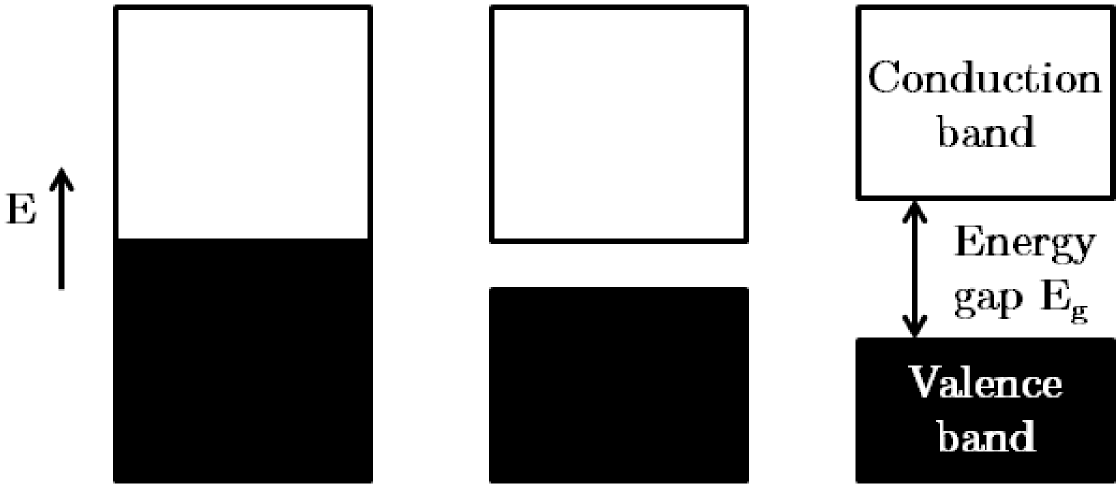


Figure 2.1: One-dimensional electronic band structures of a metal (left), semiconductor (middle), and insulator (right) at 0K. E is energy. Black-shaded areas represent occupied states by electrons. White-shaded areas represent unoccupied states. The upper and lower bands, commonly referred to as the conduction and valence bands, respectively, overlap in a metal, whereas there is no overlap in a semiconductor nor insulator.

contains a large number of finely-spaced energy levels such that the band can be approximated as a continuum. Bands are separated from one another by energy gaps within which no electron state is supported. The distribution of electrons within each band and the energy separation of each band, known as the **bandgap energy**, determines the material's electrical properties. Materials whose lower band is completely filled with electrons and whose bandgap energy is large (> 3eV) are electrical insulators. Those for which the bandgap energy is small or nonexistent are electrical conductors. Finally, those for which the bandgap energy lies between these two extremes (0.1eV$\sim$3eV) are semiconductors. A summary of these materials is provided in figure 2.1, and a list of bandgap energies for several common semiconductors is provided in table 2.1.

Semiconductors are unique in that they are naturally insulating, yet can become conductive when subjected to external stimuli. In accordance with the Pauli Exlusion principle, no two electrons can occupy the same quantum state. In the absence of external stimuli, states of lower energy are filled first. For a semiconductor such as

Material	Bandgap Energy, E_g (eV)	Bandgap Wavelength, λ_g (μm)
Si	1.11	1.15
Ge	0.66	1.88
GaAs	1.42	0.87
InP	1.35	0.92

Table 2.1: Electronic bandgap energies for a selection of group IV and group III–V semiconductors at 300K [50]. Bandgap wavelength corresponds to a photon of energy equal to the bandgap energy. To wit, $\lambda_g = hc/E_g$, where h is Planck's constant and c is the speed of light in vacuum.

silicon, one band is completely filled with electrons while the band immediately above is completely vacant. These bands are called the *valence* and *conduction* bands, respectively. With no vacant states available within the valence band, motion of electrons is impossible. However, in the presence of an external stimulus, such as a lattice vibration or an incident photon of energy greater than the bandgap energy, an electron may be excited into the conduction band. There, the electron behaves as a mobile carrier, as it is free to move among the abundance of unoccupied states. Under the influence of an applied electric field, this electron drifts in the crystal lattice and, therefore, contributes to the electric current. The naming of the conduction band is attributed to this effect. Furthermore, the departure of the electron to the conduction band provides an empty state to the valence band, allowing the remaining electrons in the valence band to change places with each other under the influence of an electric field. This collective motion may be equivalently regarded as the motion of a single, positively-charged *hole* in the opposite direction. Thus, an external stimulus may result in the formation of two charge carriers: a negatively-charged electron in the conduction band and a positively-charged hole in the valence band.

A photodetector exploits the formation of electron-hole pairs via the absorption of photons. The energy of an individual photon, E_{ph}, is given by

$$E_{ph} = h\nu \tag{2.1}$$

where h is Planck's constant and ν is the photon's frequency. If this energy is greater

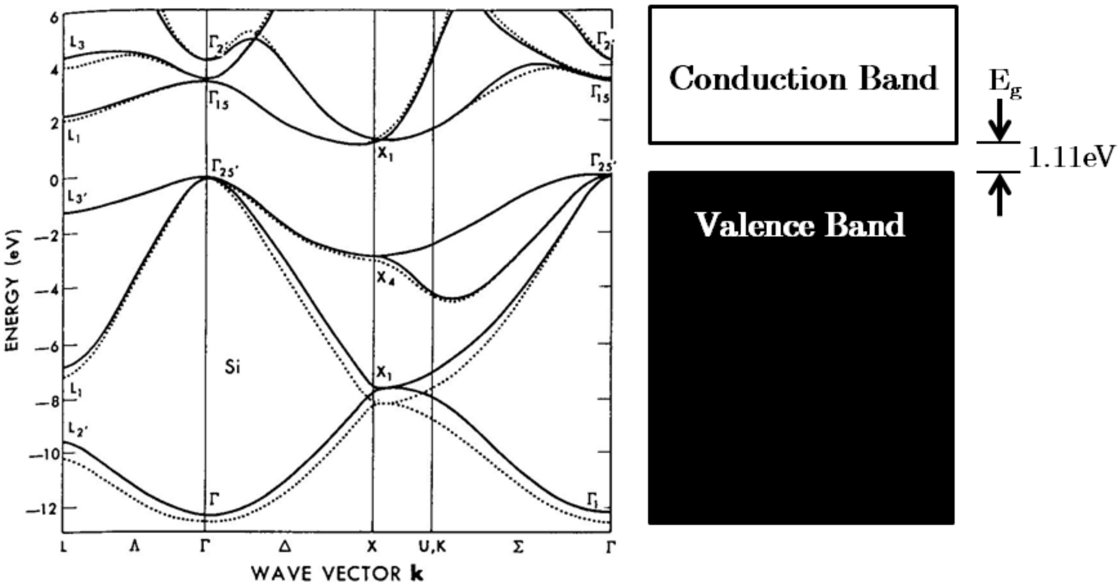


Figure 2.2: Electronic band structure of silicon. The valence band maximum is taken as zero energy by convention. Symbol Γ represents a zero k-vector. Symbols L, X, and K represent k-vector values π/a_1,π/a_2, and π/a_3, respectively, along crystal lattice directions Λ (111), Δ (100), and Σ (110), respectively. The E-k diagram on the left is taken from [54].

than the semiconductor's bandgap energy, the photon can be absorbed to yield an electron-hole pair. The efficiency of this absorption is strongly dependent on the semiconductor material's energy-momentum dynamics. In general, the energy E of a quantum state in the semiconductor is a function of wave-vector $\mathbf{k}$. For an electron in freespace, the E-k relationship is an isotropic parabola ($E = \hbar^2 k^2/2m_o$, where $\hbar$ is the reduced Planck's constant, k is $|\mathbf{k}|$, and m_o is electron mass). This follows from the electron being free of electric potential. For an electron in a semiconductor, however, the E-k relationship is a subject of different dynamics. Namely, the relationship is governed by the periodic potential of the crystal lattice, which involves the structural arrangement of atoms and their relative distances as dictated by bond strength. The E-k relationship of silicon is illustrated in figure 2.2. The energy is a periodic function of the components (k_1,k_2,k_3) of vector $\mathbf{k}$, with periodicities (π/a_1,π/a_2,π/a_3), where a_1, a_2, and a_3 are the crystal lattice constants. Thus, the energy of an electron is a function of not only the magnitude of its momentum, but also on the direction in which it is travelling in the crystal. If the least energetic state in the conduction band and the most energetic state in the valence band occur at the same momentum, then the semiconductor is called a **direct-gap** material. Otherwise, it is called a **indirect-gap** material. Evidently, silicon is an indirect-gap material.

Electrons in indirect-gap materials require a substantial change in their momentum in order to be promoted from the top of the valence band to the bottom of the conduction band. Photons do not carry enough momentum to facilitate this change, so excitation must be mediated by a third entity, typically a phonon, which is a quantised vibration of the crystal lattice. This requirement of coincidence of three entities (electron, photon, and phonon) is what makes carrier generation in indirect-gap materials unlikely, and, thus, inefficient. It should be noted that silicon may also absorb photons of energy sufficient for a direct-gap transition, after which the promoted electron will rapidly *thermalise*, or release energy in the form of a phonon, down to the bottom of the conduction band. Unlike the indirect-gap transition, this process does not require the presence of a phonon to be initiated, so it is more efficient. However, the incident photon is required to have a much greater energy (4.17eV for the direct-gap transition as opposed to 1.11eV for the indirect-gap transition [54]), making the exchange impractical in comparison to true direct-gap semiconductors such as gallium arsenide. The indirect-gap and direct-gap formations of an electron-hole pair in silicon are shown in figure 2.3.

The rate of photo-excitation of charge carriers depends on the concentration of electrons in the conduction band and holes in the valence band. At finite temperature, there is always a non-zero concentration of both electrons and holes in either band. These concentrations are a product of each band's density of quantum states and the probability of each of these states being occupied. The density of states for electrons, $\rho_c(E)$, and holes, $\rho_v(E)$, are given by

$$\rho_c(E) = \frac{(2m_c)^{3/2}}{2\pi^2\hbar^3}(E - E_c)^{1/2}, \quad E \geq E_c \tag{2.2}$$

$$\rho_v(E) = \frac{(2m_v)^{3/2}}{2\pi^2\hbar^3}(E_v - E)^{1/2}, \quad E \leq E_v \tag{2.3}$$

where, respectively, m_c and m_v are the effective electron and hole masses, and E_c and E_v are the energies of the bottom of the conduction band and top of the valence band. The probability of occupancy is given by the **Fermi-Dirac distribution** (or Fermi function). For electrons, this distribution is

$$f_c(E) = \frac{1}{e^{\frac{(E-E_f)}{k_B T}} + 1} \tag{2.4}$$

where k_B is the Boltzmann constant (1.38×10^{-23}J·K^{-1}) and T is temperature. E_f is an energy constant known as the **Fermi level** which, upon inspection of (2.4), is seen to represent the energy at which probability of occupancy is 0.5. The distribution for holes is simply the complement of (2.4):

$$\begin{aligned} f_v(E) &= 1 - f_c(E) \\ &= \frac{1}{e^{\frac{(E_f-E)}{k_B T}} + 1} \end{aligned} \tag{2.5}$$

In intrinsic silicon, the Fermi level lies roughly in the middle of the bandgap. At $T = 0$K, one can conclude that the probability of an electron occupying states above the Fermi level (or, rather, in the conduction band) is precisely zero. Thus, (2.4) upholds the rule that electrons occupy the lowest energy levels first. Photo-excitation

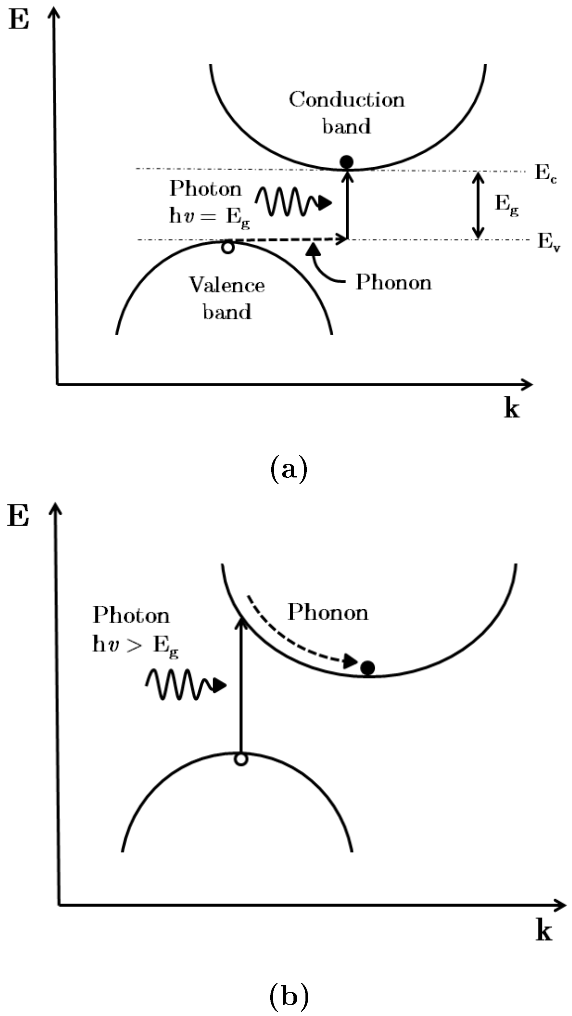


Figure 2.3: Illustrations of electron-hole pair formation in silicon: (a) Indirect-gap formation involving the simultaneous interaction of a photon and phonon; (b) direct-gap formation involving the absorption of a photon followed by the emission of a phonon (ie. thermalisation). E_c and E_v correspond to the energy value of the quantum state at the bottom of the valence band and the top of the conduction band, respectively. Bandgap energy E_g is the difference of these two quantities.

of an electron from the valence band to the conduction band may occur only when a conduction-band state is free of an electron and a valence-band state is filled with an electron.

Before the rate at which photo-excitation yields electron-hole pairs is quantified, it pays to express density of states and occupancy probability in terms of the incident photon's frequency. To first order, the top of the valence band and the bottom of the conduction band may be approximated with parabolas. Then, a conduction-band state has energy E_2 given by

$$E_2 = E_c + \frac{\hbar^2 k^2}{2m_c} \tag{2.6}$$

and a valence-band state has energy E_1 given by

$$E_1 = E_v - \frac{\hbar^2 k^2}{2m_v} \tag{2.7}$$

A photon of energy $h\nu$ must interact with an electron and hole such that energy is conserved. Namely, the photon must interact with an occupied state of energy E_1 and an unoccupied state of energy E_2, so

$$E_2 - E_1 = h\nu \tag{2.8}$$

By substituting equations (2.6) and (2.7) into (2.8) and noting that $E_g = E_c - E_v$, energies E_2 and E_1 can be recast as a function of the photon's frequency. To wit,

$$E_2 = E_c + \frac{m_r}{m_c}(h\nu - E_g) \tag{2.9}$$

$$E_1 = E_v - \frac{m_r}{m_v}(h\nu - E_g) \tag{2.10}$$

where

$$\frac{1}{m_r} = \frac{1}{m_c} + \frac{1}{m_v} \tag{2.11}$$

The one-to-one correspondence between state energies and the photon's frequency permits the density of states to be recast in terms of frequency, as well. Using the incremental relation $\rho_c(E_2)dE_2 = \rho(\nu)d\nu$, substituting in equations (2.2) and (2.9), then solving for $\rho(\nu)$ gives

$$\rho(\nu) = \frac{(2m_r)^{3/2}}{\pi\hbar^2}(h\nu - E_g)^{1/2} \tag{2.12}$$

Lastly, the occupancy probability for photoabsorption, $f_a(\nu)$ must be determined. A hole occupying state of energy E_2 and an electron occupying state of energy E_1 are independent events, so this probability is simply

$$f_a(\nu) = [1 - f_c(E_2)]\,[1 - f_v(E_1)] \tag{2.13}$$

where $f_c(E)$ and $f_v(E)$ are given by (2.4) and (2.5), respectively.

For a pair of quantum states separated by energy $E_2 - E_1 = h\nu_o$, the photo-excitation rate is obtained as follows. The density of states $\rho(\nu_o)$ is multiplied by the occupation probability $f_a(\nu_o)$ and the transition probability density $W_i(\nu)$. The transition probability density represents the likelihood that a photon initiates an excitation event given that the occupancy condition is satisfied. This density is a product of the semiconductor material's transition cross-section $\sigma(\nu)$ and the mean photon-flux spectral density ϕ_ν (photons per second per hertz per cm^2). For the absorption of one photon within the narrow frequency band between ν and $\nu + d\nu$, $W_i(\nu)$ is given by

$$\begin{aligned} W_i(\nu)d\nu &= \phi_\nu \sigma(\nu) d\nu \\ &= \phi_\nu \frac{c^2}{8\pi\nu^2\tau_r} g(\nu) d\nu \end{aligned} \tag{2.14}$$

where τ_r is the radiative electron-hole recombination lifetime and g(ν) is the lineshape function. The overall photo-excitation rate, $r_{ex}(\nu)$ (photons per second per hertz per cm^3), is found by integrating the final product over ν_o, giving

$$r_{ex}(\nu) = \phi_\nu \frac{c^2}{8\pi\nu^2\tau_r} \rho(\nu) f_a(\nu) \tag{2.15}$$

where $g(\nu)$ was approximated as $\delta(\nu - \nu_o)$. Note that the photo-excitation rate is proportional to the photon-flux spectral density. In section 3.5, it will be shown how photonic-crystal waveguides can be used to increase spectral density and, thereby, increase the photo-excitation rate. Also note that this equation accounts for direct-gap transitions *only*, as it does not take into account the presence of phonons. Finally, not all absorption mechanisms are considered; photons may also be absorbed in free-carrier, impurity-to-band, and excitonic transitions. Nevertheless, the dependence of photo-excitation on photon-flux, carrier concentration, and photon frequency has been established.

2.2 The *p*-*n* Junction

Left to their own devices, photo-excited electrons and holes in an electrically neutral semiconductor move in random directions until they recombine. Photodetection is achieved only when these electrons and holes are separated to form a measurable current. Separation is possible via the application of an electric field. In this work, the built-in electric field of a *p*-*n* junction is used to this end. This section provides a brief summary of the physics of *p*-*n* junctions.

The concentration of mobile electrons and holes in silicon can be significantly altered by incorporating specific impurities, or **dopant**, into silicon's crystal lattice. Dopant of atomic group III or lower have fewer valence electrons than silicon. If a dopant of these groups replaces a silicon atom, the dopant will replace electrons in the valence band with mobile holes, one for every fewer valence electron the dopant has, provided that there is enough thermal energy within the material to ionise the dopant site. Dopant with fewer valence electrons than silicon is said to be an **acceptor**, since it removes, or accepts, electrons from the valence band. Conversely, dopant of groups V or higher have more valence electrons than silicon. Replacing a silicon atom with a dopant of these groups will replace holes in the conduction band with mobile electrons, again provided that the dopant site can be ionised. A dopant with more valence electrons than silicon is said to be a **donor**, since it donates electrons to the conduction band. Silicon with an abundance of mobile holes is said to be a *p*-type material, whereas silicon with an an abundance of mobile electrons is said to be an *n*-type material. Common donors and acceptors are phosphorus and boron, respectively, because both are miscible with silicon and are easily ionised at room temperature. Both contribute one mobile electron or hole to silicon. At even modest dopant concentrations, the number of mobile carriers attributed to the presence of these dopant overwhelms the number attributed to thermal distribution as described in (2.4) and (2.5). Thus, at room temperature, where all dopant may be assumed to be ionised, the mobile electron or hole concentration roughly equals the dopant concentration. It must be noted that whereas charge carrier concentration changes with the introduction of dopant, silicon as a whole remains electrically neutral, as the charge of a carrier is counterbalanced by the charge of the dopant site from which it originated.

A change in carrier concentration with doping may be mathematically described using the concept of Fermi energy. When doped with acceptors, silicon's Fermi energy shifts toward the valence band. Conversely, when doped with donors, silicon's Fermi energy shifts toward the conduction band. Let $n(E)\Delta E$ and $p(E)\Delta E$ be the number of mobile electrons and holes per unit volume, respectively, with an energy lying between E and $E + \Delta E$. Using (2.2)–(2.5), $n(E)$ and $p(E)$ are simply

$$\begin{aligned} n(E) &= \rho_c(E) f_c(E) \\ p(E) &= \rho_v(E) f_v(E) \end{aligned}$$

respectively. The concentrations of electrons, n, and holes, p, (populations per unit volume) are obtained by evaluating the following integrals.

$$\begin{aligned} n &= \int_{E_c}^{\infty} n(E)\,\mathrm{d}E \\ p &= \int_{-\infty}^{E_v} p(E)\,\mathrm{d}E \end{aligned}$$

Upon doing so, the following is approximated.

$$n = n_i e^{-\frac{(E_f - E_i)}{k_B T}} \tag{2.16}$$

$$p = n_i e^{-\frac{(E_i - E_f)}{k_B T}} \tag{2.17}$$

where

$$\begin{aligned} n_i &= \sqrt{N_c N_v} e^{-\frac{E_g}{2k_B T}} \\ N_c &= 2\left(\frac{2\pi m_c k_B T}{h^2}\right)^{3/2} \\ N_v &= 2\left(\frac{2\pi m_v k_B T}{h^2}\right)^{3/2} \end{aligned}$$

E_i is the Fermi energy of undoped, or intrinsic, silicon. n_i is the mobile electron and hole concentrations in their respective bands in intrinsic silicon ($n_i = 1.45 \times 10^{10}\mathrm{cm}^{-3}$ at 300K). Note that (2.16) and (2.17) are valid irrespective of the doping type; the effect of doping is reflected in the difference $\pm(E_f - E_i)$. Mobile electron and hole concentrations are also related by the product $np = n_i^2$. The constancy of this product underscores the tradeoff of doping: increasing the concentration of one carrier type necessarily decreases the concentration of the other. If the dopant's ionisation energy is around k_BT, such as for boron (45meV) or phosphorus (44meV) [55], it is reasonable to approximate n and p with donor and acceptor concentrations N_D and N_A, respectively.

A p-n junction is formed by placing p-type and n-type silicon in contact. When these two types are brought into contact, electrons and holes diffuse toward areas of low concentration. Thus, electrons diffuse from the n-type silicon into the p-type silicon, where they recombine with the abundant holes. Conversely, holes diffuse from the p-type silicon into the n-type silicon, where they recombine with the abundant electrons. As a consequence, a narrow region on both sides of the contact becomes depleted of mobile charge carriers. This region is called the **depletion layer**. With the elimination of mobile charge carriers within this layer, the positively-ionised donors on the n-side and the negatively-ionised acceptors on the p-side, which are both incorporated in the crystal lattice and, thus, cannot diffuse, become exposed. No longer counterbalanced by their mobile charge carriers, the distribution of these fixed charges leads to the creation of an electric field which points from the n-side toward the p-side of the junction. Equilibrium is reached when this built-in field obstructs the diffusion of additional mobile carriers through the junction region. Figure 2.4 illustrates one such p-n junction. In thermal equilibrium, the entire junction must have a single Fermi energy. The junction's conduction and valence bands bend across the depletion layer to accommodate this requirement. The bending corresponds to a contact potential V_o given by

$$\begin{aligned} V_o &= q(E_{cp} - E_{cn}) \\ &= q(E_{vp} - E_{vn}) \end{aligned}$$

in the band diagram, where q is elementary charge, E_{cp} and E_{cn} are the conduction band energies of the p-side and n-side of the junction, respectively, and E_{vp} and E_{vn} are the valence band energies of the p-side and n-side of the junction, respectively.

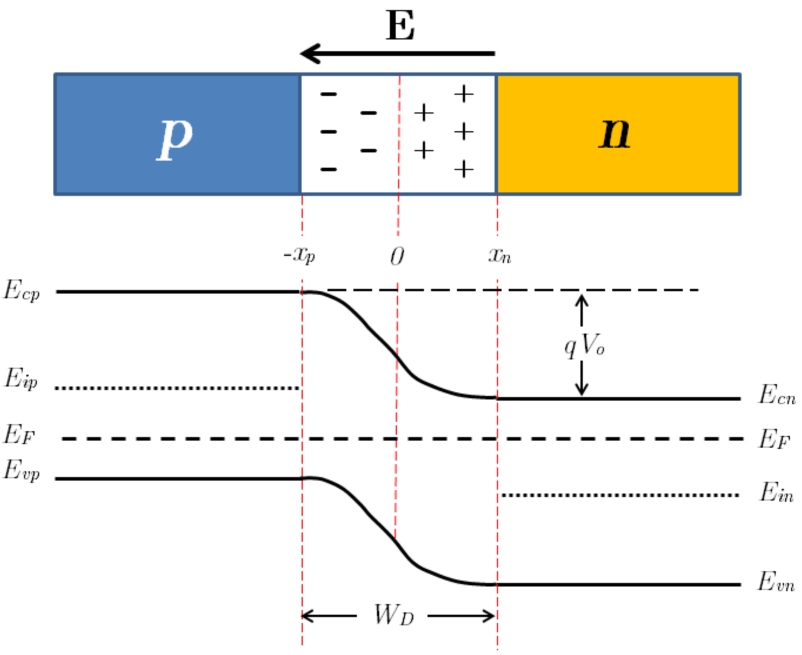


Figure 2.4: A p-n junction in thermal equilibrium showing fixed charges, electric field **E**, and depletion layer (top) and energy band diagram with built-in junction potential V_o (bottom). x_p and x_n are the depletion layer widths in the p-side and n-side, respectively. W_D is the total depletion layer width ($W_D = x_p + x_n$). E_{ip} and E_{in} are the intrinsic Fermi energies of the p-side and n-side, respectively. The Fermi energy E_F is below this intrinsic Fermi energy on the p-side and above it on the n-side in accordance with charge carrier concentrations.

If both dopant concentrations are known, the built-in voltage, V_o, and the depletion width, W_D, can be calculated. The built-in voltage is given by

$$V_o = \frac{k_B T}{q} \ln\left(\frac{N_A N_D}{n_i^2}\right) \tag{2.18}$$

Although finite in value, this voltage cannot be measured. This is because the metal-semiconductor contact voltages that would be established at either end of the junction when performing the measurement would counterbalance it. This must be so, otherwise the *p*-*n* junction would act as a battery in clear violation of conservation of energy. The depletion width is calculated from consideration that an equal amount of positive or negative fixed charge must exist on either side of the junction. This charge-equality condition can be expressed as

$$|x_p| N_A = |x_n| N_D \tag{2.19}$$

where $|x_p|$ and $|x_n|$ are the magnitudes of the widths of the depletion region in the p-side and n-side, respectively. (2.19) assumes that the junction has a uniform cross-section. Using Poisson's equation, voltage drop as a function of position within the depletion layer may be found. Noting that the total drop must equal the built-in voltage expressed in (2.18), the following expression for depletion width is obtained.

$$\begin{aligned} W_D &= x_n + x_p \qquad (2.20) \\ &= \sqrt{\frac{2\epsilon V_o}{q}\left(\frac{1}{N_A} + \frac{1}{N_D}\right)} \qquad (2.21) \end{aligned}$$

ϵ is the electrical permittivity of the material ($1.04 \times 10^{-12}\mathrm{F{\cdot}cm^{-1}}$ for silicon at 300K). It is evident from (2.19) that the depletion layer width lies primarily on the side with lighter doping concentration.

Current across a *p*-*n* junction is comprised of two components: diffusion and drift. Diffusion current is the motion of electrons in response to a concentration gradient Here, diffusion current flows from the *p*-side of the junction to the *n*-side. Drift current is the motion of electrons under the influence of an electric field, such as the built-in electric field across the depletion layer. Here, drift current originates from the thermal generation of charge carriers around the depletion region. Charge carriers generated within the depletion region proper are duly separated by the electric field. Minority charge carriers (i.e. electrons within the p-side and holes within the n-side) generated within one diffusion length of the depletion layer may also enter the depletion layer and contribute to the drift current. In equilibrium, these two components are equal in magnitude and in opposite directions, so no net current flows across the junction.

Under an applied bias, the junction is no longer in equilibrium. In reverse bias, a negative voltage $-V$ is applied to the p-side of the junction. This augments the potential energy inhibiting the diffusion of majority carriers (ie. holes in the p-side and electrons in the n-side) to opposing sides by a value of qV. Appropriately, diffusion current is weakened and drift current is allowed to dominate. A finite amount of current, which points from the n-side to the p-side, now exists across the junction. Mobile holes are led from the p-material into the external circuit supplying the bias potential. Similarly, mobile electrons are led from the n-material. As a result, more fixed charge is uncovered in the depletion layer, causing the depletion layer to *widen* to a value prescribed by (2.20) with V_o replaced by $V_o + V$. As stated before, drift current originates from the thermal generation of charge carriers. Thermal generation is independent of barrier potential, so net current across the junction remains constant with V.

In forward bias, a positive voltage V is applied to the p-side of the junction. This produces an electric field which opposes the built-in field. Accordingly, the potential energy inhibiting the diffusion of majority carriers is weakened by a value of qV. A greater number of majority carriers now have sufficient energy to surmount the energy barrier and diffuse to the opposite side. As a result, diffusion current becomes larger than drift current, and a finite current pointing from the p-side to the n-side exists across the junction. Majority carriers contributing to the diffusion current are supplied by the external circuit applying the forward bias. The presence of these externally-supplied carriers neutralise some of the fixed charge in the depletion layer, causing the depletion layer width to *narrow* to a value prescribed by (2.20) with V_o replaced by $V_o - V$. Unlike in reverse-bias mode, the magnitude of diffusion current is strongly dependent on V. An increase in V lowers the energy barrier further and permits more majority carriers to contribute to the diffusion process, which thereby

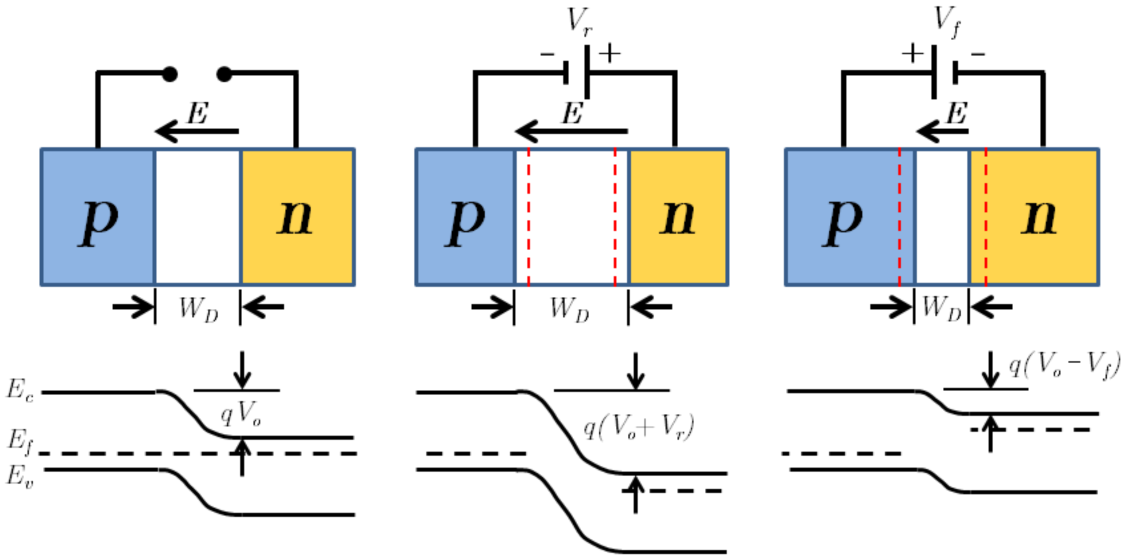


Figure 2.5: Electric field, depletion width, and energy band diagram of a p-n junction under (left) zero bias, (middle) reverse bias $V = V_r$, and (right) forward bias $V = V_f$.

increases current. At sufficiently high V, this current can become large enough to burn out the p-n junction's contacts via Joule heating. The effect both reverse- and forward-bias modes have on the junction's built-in potential may be described as a misalignment of the Fermi energies in the p- and n-regions. A summary of these bias modes is illustrated in figure 2.5.

The junction's current-voltage (I-V) characteristics resemble those of a diode. In forward bias, current is found to be proportional to a factor $\exp(qV/k_BT)$. In reverse bias, current is a small, constant value I_s. At zero bias, current is zero. These characteristics are captured in the following equation.

$$I = I_s\left(e^{\frac{qV}{k_BT}} - 1\right) \tag{2.22}$$

The corresponding I-V curve is plotted in figure 2.6. (2.22) is known as the **ideal diode equation**, since it does not take into account second-order features such as turn-on voltage, forward resistance, and Zener breakdown. I_s is the drift current and is given by

$$I_s = Aqn_i^2\left(\frac{D_p}{L_pN_D} + \frac{D_n}{L_nN_A}\right) \tag{2.23}$$

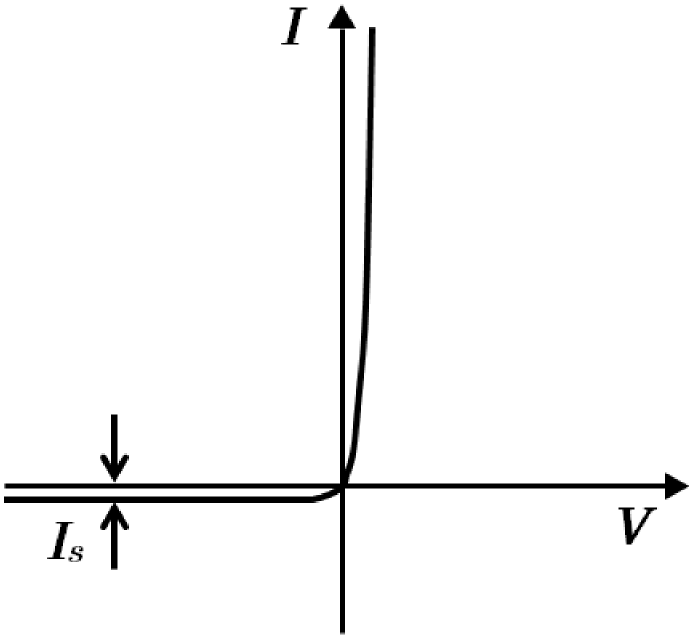


Figure 2.6: *I*-*V* characteristics of an ideal *p*-*n* junction.

where A is junction area, D_p is the diffusion constant of minority holes in the n-side, and D_n is the diffusion constant of minority electrons in the p-side. L_p and L_n are the diffusion lengths of these same minority carriers and are related to the diffusion constant D and minority-carrier lifetime τ as follows:

$$L = \sqrt{D\tau} \tag{2.24}$$

τ is the average time it takes for a minority carrier to recombine with a majority carrier of the opposite type. The diffusion constant is proportional to carrier mobility μ as follows:

$$\frac{D}{\mu} = \frac{k_B T}{q} \tag{2.25}$$

A large diffusion constant increases the diffusion length, which allows more thermally generated minority carriers to drift into the depletion region and contribute to the drift current. On the other hand, a large minority-carrier lifetime, which also increases diffusion length, entails that carriers swept through the depletion layer take longer to recombine with minority carriers of the opposing type on the other side. In the steady state, reverse-bias current consists of the external circuit resupplying carriers lost to this recombination. Thus, minority-carrier lifetime limits drift current. These competing processes are reflected in the quotient $D/L = \sqrt{D/\tau}$ in (2.23). For a large drift current, minority carriers must quickly diffuse to enter the depletion layer (large D), yet quickly recombine with carriers of the opposite type that approach from the other side (small τ).

2.3 Photodiodes

When illuminated with photons of energy greater than the bandgap, additional electron-hole pairs are generated within a p-n junction. If these pairs are generated inside of or within one diffusion length of the depletion layer, they will be separated by the built-in field and contribute to the drift current. The I-V equation of a p-n junction can be modified to accommodate this contribution as follows.

$$I = I_s\left(e^{\frac{qV}{k_BT}} - 1\right) - I_{ph} \tag{2.26}$$

I_{ph} is the **photocurrent**, or the contribution of photo-excited electron-hole pairs. It it of negative polarity so that it adds to the drift current. Its magnitude is proportional to the incident photon flux and the drift and diffusion of electrons and holes. Let the electron-hole pair (EHP) generation rate of a uniformly illuminated p-n junction of area A be g_{ph} (units EHP·cm^{-3}·s^{-1}). The number of separable holes created per second on the n-side is AL_pg_{ph} Similarly, the number of separable electrons created per second on the p-side is AL_ng_{ph}. Additionally, AW_Dg_{ph} electron-hole pairs are generated within the depletion layer, of which all are separable. Assuming that the drift time is much shorter than the recombination lifetime, every one of these electron-hole pairs contributes to the photocurrent. The photocurrent is thus

$$I_{ph} = qAg_{ph}(L_p + L_n + W_D) \tag{2.27}$$

The resulting change to the junction's I-V curve is illustrated in figure 2.7. When the device is short-circuited (ie. $V = 0$), a finite, negative current equal to I_{ph} exists across

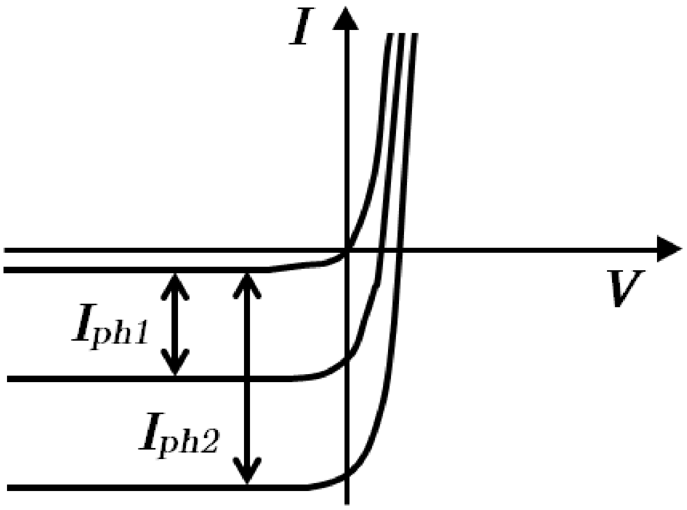


Figure 2.7: *I-V* characteristics of an ideal *p-n* junction under illumination. The junction experiences an increase in current proportional to incident photon flux Φ. In this case, $\Phi_2 > \Phi_1$, giving $I_{ph2} > I_{ph1}$.

the junction. Thus, the I-V characteristics cross the current-axis at negative values. When the device is open-circuited (ie. $I = 0$), a finite, positive voltage exists across the junction. The appearance of a forward voltage across an illuminated junction is known as the *photovoltaic effect*. Contrast these characteristics with those of a *p-n* junction that is not illuminated (equation 2.22). In the case of an unilluminated junction, voltage is zero when current is zero and vice versa.

Photodetectors are commonly operated under reverse bias (ie. in the third quadrant of figure 2.7) for several reasons. First, under reverse bias, current is roughly independent of voltage, but remains proportional to the optical generation rate. This allows a one-to-one correspondence between current and incident photon flux to be established. Second, reverse bias strengthens the electric field in the depletion layer, which serves to increase carrier drift velocity. Increasing this drift velocity improves the photodetector's transient response, which permits the resolution of photon signals that are more closely spaced in time. Third, reverse bias widens the depletion layer, which increases A in (2.27), leading to a larger photocurrent. Furthermore, a wider depletion layer allows a greater fraction of carriers to be generated within this region

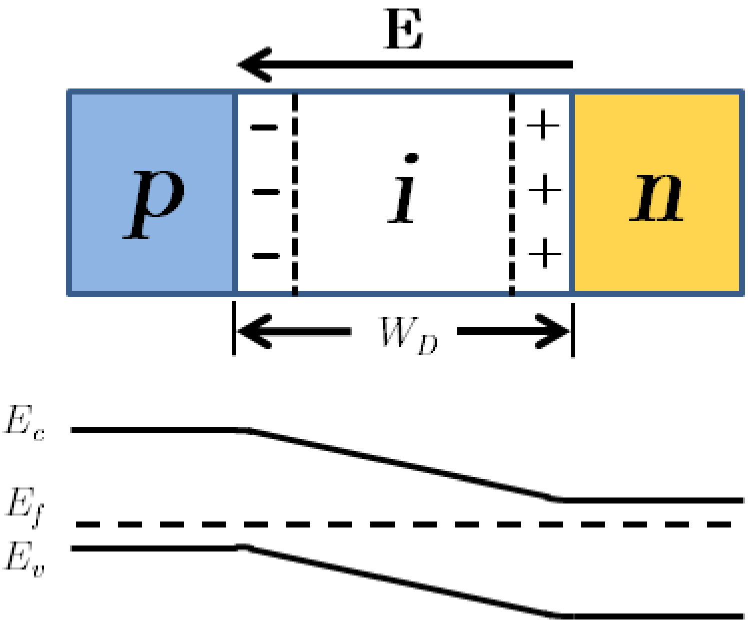


Figure 2.8: Illustration of a *p*-*i*-*n* junction and its corresponding energy band diagram.

as opposed to the *p*- and *n*-regions. Carriers generated within the depletion layer are immediately separated by the drift process. In contrast, carriers generated within the *p*- and *n*- regions must first diffuse into the depletion layer before being separated by the drift process. Diffusion is a much slower process than drift, so increasing the proportion of carriers generated within the depletion layer further improves the photodetector's transient response. Lastly, widening the depletion layer increases the distance by which the majority electrons on the *n*-side and the majority holes on the *p*-side are separated. A permanent circuit capacitance C_j is attributed to this separation of majority carriers. C_j is given by

$$C_j = \epsilon A_j \sqrt{\frac{q}{2\epsilon(V_o - V)} \frac{N_D N_A}{N_D + N_A}} = \frac{\epsilon A_j}{W_D} \tag{2.28}$$

where A_j is the junction's cross-sectional area. A larger W_D value leads to a smaller C_j value, which lowers the RC constant, and thus the response time, of the photodetector circuit.

A convenient method for controlling the width of the depletion layer is to place

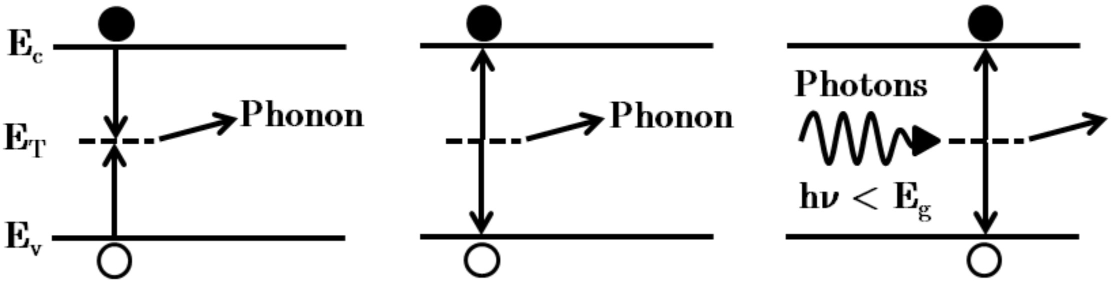


Figure 2.9: Electron-hole recombination and generation via deep-level traps in silicon: (left) recombination, (center) generation, and (right) generation via absorption of a photon with sub-bandgap energy. Since silicon is an indirect bandgap semiconductor, all three processes are mediated by a phonon.

an intrinsic region between the *p* and *n* regions. This kind of structure, called a *p*-*i*-*n* junction, is shown in figure 2.8. The intrinsic region may even be substituted with a region that is lightly doped, in which case the structure is called a p-ν-n junction. With its relative absence of free carriers, this middle region is easily depleted of carriers when placed in contact with the *p*- and *n*-regions. As a result, the depletion layer and accompanying electric field extend across this entire region. There is a limit as to how wide the intrinsic region can be. A region too wide will not be fully depleted of carriers, especially if it is doped. In this case, two depletion regions will form – one near the *n*-region and one near the *p*-region. Also, widening the depletion layer increases the time required for carrier drift. If the depletion layer is made too wide, this time requirement will bottleneck the aforementioned benefits to transient response.

2.4 Carrier Generation via Deep-Level Traps

Apart from direct-gap transitions as explained in section 2.1, electrons and holes can also be recombined and generated via transitions to or from deep-level traps. Deep-level traps are energy states that lie approximately in the middle of a semiconductor's band gap and are introduced by certain impurities such as gold, or by lattice defects such as interstitials and vacancies. Traps facilitate several transition processes, each of which is illustrated in figure 2.9. For the case of recombination (left illustration in figure 2.9), an electron or hole relaxes to the trap and remains there. Separately, a charge carrier of the opposite type relaxes to the trap, then the two carriers recombine. The case of generation (centre illustration in figure 2.9) is identical to recombination, only now the two charge carriers are excited, rather than relaxed, to the trap. In an indirect-gap semiconductor such as Silicon, a phonon must still mediate these transitions. However, here the trap serves to hold a charge carrier in place, so the spatial coincidence of charge carrier and phonon is more probable. Accordingly, deep-level transitions dominate over bandgap transitions in indirect-gap semiconductors.

Clearly, recombination and generation are dual processes. The net transition rate, U, is described by Shockley-Read-Hall statistics:

$$U = \frac{\sigma_n \sigma_p v_{th} N_T (np - n_i^2)}{\sigma_n \left[n + n_i \mathrm{exp}\left(\frac{E_T - E_i}{kT}\right)\right] + \sigma_p \left[p + n_i \mathrm{exp}\left(\frac{E_i - E_T}{kT}\right)\right]} \tag{2.29}$$

σ_n and σ_p are, respectively, the electron and hole capture cross sections, N_T is the trap density, and E_T is the trap energy. U is maximum when the trap energy E_T equals the intrinsic Fermi energy E_i. For a neutrally-charged divacancy in silicon (V_2^0), $E_T = E_c - 0.41\mathrm{eV}$ [56], giving $E_T - E_i \approx 0.15\mathrm{eV}$. A positive value of U represents net recombination; a negative value represents net generation. Whether there is net recombination or generation is seen, upon inspection, to depend on the difference $np - n_i^2$. With respect to the intrinsic region of p-n junctions, there may be net recombination in forward-bias mode, and there is net generation in reverse-bias mode. For a semiconductor in equilibrium, $np = n_i^2$, so the net transition rate is zero.

The denominator of U can be simplified if the semiconductor is predominantly n- or p-type. If the semiconductor is n-type, then the denominator can be approximated as $\sigma_n n$, which allows U to be condensed as follows.

$$U = \frac{\sigma_n \sigma_p v_{th} N_T (np - n_i^2)}{\sigma_n n} \tag{2.30}$$

$$= \frac{\sigma_p v_{th} N_T (np - n_i^2)}{n} \tag{2.31}$$

$$= \frac{np - n_i^2}{\tau_p n} \tag{2.32}$$

τ_p $(= 1/\sigma_p v_{th} N_T)$ is the **lifetime** of holes, or the "time taken" for holes to participate in the transition. Similarly, if the semiconductor is p-type, U is condensed to read

$$U = \frac{np - n_i^2}{\tau_n p} \tag{2.33}$$

where τ_n $(= 1/\sigma_n v_{th} N_T$ is the lifetime of electrons, or the "time taken" for electrons to participate in the transition. Taken together, and assuming that $\sigma_n \approx \sigma_p$, U may be formally rewritten as

$$U = \frac{np - n_i^2}{\tau(n + p + 2n_i \cosh\left(\frac{E_T - E_i}{kT}\right))} \tag{2.34}$$

where

$$\tau = \frac{1}{\sigma v_{th} N_t} \tag{2.35}$$

is the lifetime of the minority carrier. Thus, minority carrier lifetime limits the transition rate. This agrees with intuition, since traps are less likely to capture minority carriers, as these carriers are fewer in number, yet must capture minority carriers to complete the transition.

Deep-level trapping sensitises a photodiode to sub-bandgap photons. Whereas before a photodiode could only absorb sub-bandgap photons via two-photon absorption, which is an improbable process, trapping enables a photodiode to also absorb sub-bandgap photons via the two-stage trapping process (right illustration in figure 2.9). Here, photons need only have energies equal to or greater than the energy difference between the trap and a single band edge to participate. Electron-hole pairs generated at traps are swept out of the intrinsic region by the built-in electric field to keep the region depleted of charge carriers and, therefore, ensure that the polarity of U remains negative. Most importantly, the trapping mechanism promotes the pairing of photons with phonons to excite electron-hole pairs in semiconductors with indirect bandgaps. Thus, traps may be exploited to surmount the barrier silicon's indirect bandgap poses to the detection of sub-bandgap photons.

2.5 Photodetector Performance Metrics

There are several common performance metrics by which a photodetector is qualified. These include quantum efficiency, responsivity, gain, response time, and signal-to-noise ratio. In this work, photodetectors are operated in non-avalanche mode and are illuminated solely with continuous-wave sources, so gain and response time are outside the scope of investigation. This section briefly discusses the remaining three metrics: quantum efficiency, responsivity, and signal-to-noise ratio.

Quantum Efficiency

Quantum efficiency, η, is defined as the probability that a single photon incident on a photodetector generates an electron-hole pair that contributes to the detector's current. When many photons are incident, it is simply the quotient of the flux of generated electron-hole pairs that contribute to the detector's current over the incident photon flux. Its value ranges from 0 (no photon yields a detectable electron-hole pair) to 1 (every photon yields a detectable electron-hole pair). Quantum efficiency is typically less than unity for a number of reasons. First, some incident photons are simply reflected from the detector's surface. Second, some of the photons that are not reflected fail to be absorbed due to the probablistic nature of the absorption process. Third, some of the electron-hole pairs that are generated from absorbed photons quickly recombine and are therefore unable to contribute to the detector current. Carrier recombination is especially likely at structural anomalies such as surfaces, grain boundaries, and lattice defects – implanted or otherwise. The following equation for quantum efficiency accounts for these adverse effects.

$$\eta = (1 - \mathcal{R})\xi \left[1 - e^{-\alpha L}\right] \tag{2.36}$$

$\mathcal{R}$ is the optical power reflectance at the surface, ξ is the fraction of generated electron-hole pairs that contribute to the detector current, α is the absorption coefficient of the material (cm^{-1}), and L is the photodetector length. The factor $(1 - \mathcal{R})$ accounts for the fraction of photons that successfully pass into the material. ξ accounts for the fraction of photo-generated electron-hole pairs that successfully avoid recombination.

In this work, ξ can be increased by applying a stronger reverse bias voltage to impart the carriers with a higher drift velocity for separation. Lastly, the factor $\left[1 - e^{-\alpha L}\right]$ accounts for the fraction of photons absorbed in the material.

Quantum efficiency is dependent on photon wavelength due to the wavelength dependence of α. For intrinsic material, α is minimal for wavelengths longer than the bandgap wavelength (ie. $\lambda > hc/E_g$), so η becomes small. On the other hand, α is exceedingly large at short wavelengths, so most of the photons are absorbed near the material's surface. The recombination lifetime of electron-hole pairs generated near this surface is small due to the large concentration of dangling bonds, so ξ, and therefore η, becomes small. A photodetector should be operated within the spectral window between these two extremes. This spectral window is determined by the characteristics of the chosen material.

Responsivity

Responsivity, $\Re$ (A/W), is defined as the quotient of detector current over incident optical power. It represents the electrical signal strength to be expected from a detector in operation. Given incident photon flux Φ (s^{-1}), incident optical power is $P = h\nu\Phi$ (W) at frequency ν. Of these photons, a fraction η generates electron-hole pairs that contribute to the detector's current. Current is therefore $i_{ph} = \eta q \Phi$ (A). Responsivity is then

$$\begin{aligned}\Re &= \frac{i_{ph}}{P} \\ &= \frac{\eta q}{h\nu} \qquad (2.37)\end{aligned}$$

Responsivity is independent of optical power until saturation is reached, beyond which point it reduces in value. Note that since $\Re$ is proportional to η, responsivity is subject to the same dependencies on wavelength as quantum efficiency. Typically, detectors with higher responsivity also have higher sensitivity – that is, they are able to detect a smaller number of photons at a time – since the detector's photo-generated current exceeds its noise limit at lower input powers.

Noise

Noise is the random fluctuation of a photodetector's electric current from its mean value with time. Sources include photon noise, photo-electron noise, and receiver circuit noise. Photon noise is associated with the random arrivals of the photons themselves. Photo-electron noise is associated with the random generation of electron-hole pairs that contribute to the electric current. Recall that only a fraction, η, of photons yields separable electron-hole pairs. Even if photons arrive at a precisely constant rate - that is, if photon noise is zero - absorption of the photons that comprise this fraction occurs at random times. Receiver circuit noise is associated with resistive elements and active devices present in external circuitry. These sources of noise place a limit on the minimum-detectable current signal. The minimum-detectable signal is usually given in terms of the so-called **signal-to-noise ratio** (SNR) given by

$$\mathrm{SNR} = \frac{\overline{I}_{ph}}{\sigma_I^2} \tag{2.38}$$

where $\overline{I}_{ph}$ and σ_I^2 are the mean and variance of the photodetector's current, respectively. The minimum-detectable signal is defined as the mean signal at which SNR = 1. The presence of noise serves to increase the value of the variance, and thus the value of the minimum-detectable signal, which is generally undesirable.

Of the three sources of noise mentioned, receiver circuit noise is generally the largest. Thus, unadorned photodetectors tend to be circuit-noise-limited. The random thermal motion of charge carriers, known as **Johnson noise**, within the receiver circuit gives rise to a signal variance σ_J^2 given by

$$\sigma_J^2 = \frac{4k_B T B}{R} \tag{2.39}$$

where T is the circuit's temperature, B is the frequency bandwidth the detector is operating within, and R is the circuit's resistance. Johnson noise can be reduced either by placing a large resistor across the photodiode (ie. converting photocurrent into photovoltage) at the expense of increasing the circuit's RC constant or by cooling the circuit down. A transimpedance amplifier may also be used to convert photocurrent into photovoltage while maintaining a fast response time.

Chapter 3

Photonic Crystal Theory

In this work, a **photonic crystal** is integrated with a photodetector for the purpose of enhancing the photodetectors' responsivity per unit length. True to its name, a photonic crystal is the optical analogue of an atomic crystal. Both materials are comprised of a pattern, known as a **unit cell**, that is repeated over a set of discrete positions, known as a **lattice**, in space. For the case of an atomic crystal, this unit cell consists of an arrangement of atoms. The periodic potential that arises from this arrangement constrains the energies at which and directions along which an electron gas can propagate. Importantly, there can exist a gap in the crystal's energy band structure, meaning that electrons are forbidden to propagate with certain energies in certain directions. If this gap extends to cover all directions, then the crystal is said to exhibit a **complete bandgap**. The subset of crystals that exhibit such bandgaps include semiconductors, of which silicon is a member. In the case of a photonic crystal, the atoms are replaced by macroscopic media with different dielectric constants, and the periodic potential is replaced by a periodic dielectric function. If absorption is low and the dielectric contrast is sufficiently high, then scattering at the dielectric boundaries can produce the same phenomena for photons as the atomic potential does for electrons. Particularly, the crystal can exhibit a complete **photonic bandgap** (PBG) within which the propagation of photons, or *modes*, of certain frequencies is forbidden. Unlike with an atomic crystal, the photonic crystal's unit cell and lattice may be designed to forbid propagation at arbitrary frequencies. Furthermore, a defect – that is, a perturbation of the lattice at a specific location – may be introduced to support modes localised around the defect. Taken a step further, a line of such defects may be formed to support modes that propagate along an axis. If the propagation of these modes is normally forbidden by the PBG, then the modes experience lossless waveguiding. A photonic crystal with such a line defect is referred to as a photonic crystal waveguide (PCW). PCWs can support propagation in anomalous ways. Here, they are used to support propagation at low group velocities and, thereby, compress a propagating mode's electromagnetic energy density. By compressing the mode's energy density, modal intensity upon which the carrier generation rate of the photodetector depends is magnified, leading to an enhancement of the photodetector's responsivity.

This chapter provides a detailed description of the operating principles of photonic crystals. It begins with an overview of the macroscopic Maxwell equations as they apply to dielectric media, followed by some basic concepts of solid-state physics. Subsequently, a photonic crystal with one-dimensional periodicity is introduced to discuss the physical origin of the PBG. This discussion leads into the cases of a photonic crystal with two-dimensional periodicity and, finally, the PCW featured in this work: a hybrid "slab" structure that utilises bandgap confinement in one direction and total internal reflection in the other. This chapter concludes with the explanation of a technique for measuring the group velocity of a propagating mode in waveguides. Unless otherwise stated, all theory presented here is taken from [60].

3.1 The Maxwell Equations

The propagation of light in a photonic crystal is described by the four macroscopic Maxwell equations. They are

$$\nabla \cdot \mathbf{D} = \rho \tag{3.1}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{3.2}$$

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 \tag{3.3}$$

$$\nabla \times \mathbf{H} - \frac{\partial \mathbf{D}}{\partial t} = \mathbf{J} \tag{3.4}$$

where, respectively, $\mathbf{D}$ and $\mathbf{B}$ are the electric displacement and magnetic induction fields, $\mathbf{E}$ and $\mathbf{H}$ are the macroscopic electric and magnetic fields, and ρ and $\mathbf{J}$ are the free charge and current densities. For weak fields, the electric and magnetic fields are related to the electric displacement and magnetic induction fields by the equations

$$\mathbf{D} = \varepsilon_o \varepsilon \mathbf{E} \tag{3.5}$$

$$\mathbf{B} = \mu_o \mu \mathbf{H} \tag{3.6}$$

where, respectively, ε_o and μ_o are the permittivity and permeability of freespace, and ε and μ are the relative permittivity and permeability of the medium.

The Maxwell equations can be written in a more descriptive form under the following assumptions. First, for a mixed dielectric medium, ε and μ are a function of position vector $\mathbf{r}$. Assuming that the medium is isotropic, does not change with time, and is minimally dependent on frequency ω (which is a valid assumption for silicon over the C band frequency range), $\varepsilon = \varepsilon(\mathbf{r})$ and $\mu = \mu(\mathbf{r})$. By extension, all fields are also a function of $\mathbf{r}$. Second, this work deals with materials in the absence of free charge and current, so ρ and $\mathbf{J}$ are zero. Third, the material is non-magnetic (as is true for silicon and its oxide), so $\mu(\mathbf{r})$ is approximately unity. Fourth, the material is transparent, which means that $\varepsilon(\mathbf{r})$ is purely real and positive. Taking these assumptions together, the Maxwell equations become

$$\nabla \cdot [\varepsilon(\mathbf{r})\mathbf{E}(\mathbf{r},t)] = 0 \tag{3.7}$$

$$\nabla \cdot \mathbf{H}(\mathbf{r},t) = 0 \tag{3.8}$$

$$\nabla \times \mathbf{E}(\mathbf{r},t) + \mu_o \frac{\partial \mathbf{H}(\mathbf{r},t)}{\partial t} = 0 \tag{3.9}$$

$$\nabla \times \mathbf{H}(\mathbf{r},t) - \varepsilon_o \varepsilon(\mathbf{r}) \frac{\partial \mathbf{E}(\mathbf{r},t)}{\partial t} = 0 \tag{3.10}$$

In general, both **E** and **H** are complicated functions of time and space. However, since the Maxwell equations are linear, the time dependence may be separated from the spatial dependence by expanding the fields into a set of harmonic modes:

$$\mathbf{E}(\mathbf{r},t) = \mathbf{E}(\mathbf{r})e^{-i\omega t} \tag{3.11}$$

$$\mathbf{H}(\mathbf{r},t) = \mathbf{H}(\mathbf{r})e^{-i\omega t} \tag{3.12}$$

These harmonic modes are simply a sinusoid modulated by an amplitude. The real part of these modes may be taken at any time t to obtain the physical fields. Following the rules of Fourier analysis, any solution to the Maxwell equations may be built with an appropriate combination of these modes.

With temporal- and spatial-dependence separated, (3.11) and (3.12) may be substituted into Maxwell equations (3.7)-(3.10) to eliminate the temporal-dependence:

$$\nabla \cdot [\varepsilon(\mathbf{r})\mathbf{E}(\mathbf{r})] = 0 \tag{3.13}$$

$$\nabla \cdot \mathbf{H}(\mathbf{r}) = 0 \tag{3.14}$$

$$\nabla \times \mathbf{E}(\mathbf{r}) - i\omega\mu_o \mathbf{H}(\mathbf{r}) = 0 \tag{3.15}$$

$$\nabla \times \mathbf{H}(\mathbf{r}) + i\omega\varepsilon_o\varepsilon(\mathbf{r})\mathbf{E}(\mathbf{r}) = 0 \tag{3.16}$$

The divergence equations have a simple physical interpretation: there are no sources of electric or magnetic fields in the medium. Meanwhile, the curl equations can be decoupled by dividing (3.16) by $\varepsilon(\mathbf{r})$, taking the curl, then substituting in (3.15) for $\nabla \times \mathbf{E}(\mathbf{r})$. Making note of the identity for the speed of light in vacuum ($c = 1/\sqrt{\varepsilon_o \mu_o}$), the result is

$$\nabla \times \left(\frac{1}{\varepsilon(\mathbf{r})} \nabla \times \mathbf{H}(\mathbf{r}) \right) = \left(\frac{\omega}{c} \right)^2 \mathbf{H}(\mathbf{r}) \tag{3.17}$$

Given a structure $\varepsilon(\mathbf{r})$, this equation can be solved to find the modes $\mathbf{H}(\mathbf{r})$ and their corresponding frequencies, subject to transversality requirement (3.14). Then, (3.16) can be used to recover $\mathbf{E}(\mathbf{r})$:

$$\mathbf{E}(\mathbf{r}) = \frac{i}{\omega\varepsilon_o\varepsilon(\mathbf{r})}\nabla\times\mathbf{H}(\mathbf{r}) \tag{3.18}$$

While it might not be apparent at first glance, (3.17) is actually an **eigenvalue problem**. Namely, a series of operations are performed on $\mathbf{H}(\mathbf{r})$, and if $\mathbf{H}(\mathbf{r})$ is actually an allowable mode in the system, then the result is a constant times the original function $\mathbf{H}(\mathbf{r})$. (3.17) can be rewritten to make this property clear:

$$\hat{\Theta}\mathbf{H}(\mathbf{r}) = \left(\frac{\omega}{c}\right)^2\mathbf{H}(\mathbf{r}) \tag{3.19}$$

where $\hat{\Theta}$ is defined as the differential operator:

$$\hat{\Theta}\mathbf{H}(\mathbf{r}) = \nabla\times\left(\frac{1}{\varepsilon(\mathbf{r})}\nabla\times\mathbf{H}(\mathbf{r})\right) \tag{3.20}$$

In this case, $\mathbf{H}(\mathbf{r})$ may be referred to as an **eigenfunction**.

The operator $\hat{\Theta}$ happens to be a special type of operator known as a **Hermitian operator**. To explain what it means for an operator to be Hermitian, the concept of an inner product must be introduced. An inner product of two vector fields $\mathbf{F}(\mathbf{r})$ and $\mathbf{G}(\mathbf{r})$ is given by

$$(\mathbf{F},\mathbf{G}) = \int\mathbf{F}^*(\mathbf{r})\cdot\mathbf{G}(\mathbf{r})\,\mathrm{d}^3\mathbf{r} \tag{3.21}$$

As a simple consequence, $(\mathbf{F},\mathbf{G}) = (\mathbf{G},\mathbf{F})^*$ for any $\mathbf{F}$ and $\mathbf{G}$. Next, an operator Ξ is said to be Hermitian if $(\mathbf{F},\Xi\mathbf{G}) = (\Xi\mathbf{F},\mathbf{G})$ for any vector fields $\mathbf{F}(\mathbf{r})$ and $\mathbf{G}(\mathbf{r})$. That is, if the operator is Hermitian, then it does not matter which vector field is operated upon before taking the inner product. Several insightful properties may be drawn about eigenfunctions of a Hermitian operator: they have real eigenvalues, they are orthogonal to one another, they can be obtained by a variational principle, and they may be grouped by their symmetry properties. These properties will be developed one by one.

First, the frequency ω can be shown to be real by taking the complex conjugate of the inner product of (3.17) with $\mathbf{H}(\mathbf{r})$:

$$\begin{aligned}
\hat{\Theta}\mathbf{H}(\mathbf{r}) &= \left(\frac{\omega^2}{c^2}\right)\mathbf{H}(\mathbf{r}) \\
(\mathbf{H}, \hat{\Theta}\mathbf{H}) &= \left(\frac{\omega^2}{c^2}\right)(\mathbf{H}, \mathbf{H}) \\
(\mathbf{H}, \hat{\Theta}\mathbf{H})^* &= \left(\frac{\omega^2}{c^2}\right)^*(\mathbf{H}, \mathbf{H})
\end{aligned}$$

From the definitions of the inner product and Hermitian operator, it is known that $(\mathbf{H}, \hat{\Theta}\mathbf{H})^* = (\hat{\Theta}\mathbf{H}, \mathbf{H}) = (\mathbf{H}, \hat{\Theta}\mathbf{H})$. This is simply the left-hand side of the equation before the complex conjugate is taken. Thus,

$$\begin{aligned}
(\mathbf{H}, \hat{\Theta}\mathbf{H}) &= \left(\frac{\omega^2}{c^2}\right)^*(\mathbf{H}, \mathbf{H}) \\
\left(\frac{\omega^2}{c^2}\right) &= \left(\frac{\omega^2}{c^2}\right)^*
\end{aligned}$$

which is only true if ω^2 is real. Next, the inner product $(\mathbf{H}, \hat{\Theta}\mathbf{H})$ is evaluated to obtain

$$\begin{aligned}
(\mathbf{H}, \hat{\Theta}\mathbf{H}) &= \int \frac{1}{\varepsilon(\mathbf{r})}|\nabla \times \mathbf{H}|^2 \, \mathrm{d}^3\mathbf{r} \\
(\mathbf{H}, \mathbf{H})\left(\frac{\omega^2}{c^2}\right) &= \int \frac{1}{\varepsilon(\mathbf{r})}|\nabla \times \mathbf{H}|^2 \, \mathrm{d}^3\mathbf{r}
\end{aligned}$$

$\varepsilon(\mathbf{r})$ is finite and positive everywhere, so the integrand is non-negative everywhere. Also, the inner product of any vector field with itself is positive and real, so $(\mathbf{H}, \mathbf{H})$ is positive and real. Considered all together, every eigenvalue ω^2 must be non-negative, so ω is real.

Second, the Hermitian operator forces any two harmonic modes $\mathbf{H}_1(\mathbf{r})$ and $\mathbf{H}_2(\mathbf{r})$

with different frequencies ω_1 and ω_2 to have an inner product of zero:

$$\begin{aligned}(\mathbf{H}_2, \hat{\Theta}\mathbf{H}_1) &= \left(\frac{\omega_1^2}{c^2}\right)(\mathbf{H}_2, \mathbf{H}_1)\\ (\hat{\Theta}\mathbf{H}_2, \mathbf{H}_1) &= \left(\frac{\omega_1^2}{c^2}\right)(\mathbf{H}_2, \mathbf{H}_1)\\ \left(\frac{\omega_2^2}{c^2}\right)(\mathbf{H}_2, \mathbf{H}_1) &= \left(\frac{\omega_1^2}{c^2}\right)(\mathbf{H}_2, \mathbf{H}_1)\\ (\omega_1^2 - \omega_2^2)(\mathbf{H}_2, \mathbf{H}_1) &= 0 \qquad (3.22)\end{aligned}$$

Since $\omega_1 \neq \omega_2$, $(\mathbf{H}_2, \mathbf{H}_1)$ must equal zero. Modes with an inner product of zero are said to be **orthogonal**. This means that the scalar product $\mathbf{H}_2 \cdot \mathbf{H}_1$ must be negative as much as it is positive over the region of interest. To achieve this effect, each mode must contain a different number of nodes – that is, locations where $\mathbf{H}(\mathbf{r})$ is zero (excluding the end points). In general, higher-frequency modes contain more nodes than lower-frequency modes.

Orthogonality implies that the frequency **spectrum** of a photonic crystal is discrete rather than continuous. The spectrum of a photonic crystal is the totality of all eigenvalues ω associated with modes that satisfy (3.17). A photonic crystal is spatially periodic and therefore represents a bounded profile over which the inner product is integrated. An intuitive explanation as to why a Hermitian eigenvalue problem necessitates the requirement for a discrete set of eigenvalues is as follows. First suppose that the range of eigenvalues is continuous rather than discrete. It is reasonable to suppose, then, that as ω is changed continuously, $\mathbf{H}(\mathbf{r})$ is changed continuously as well. Let eigenvalue ω correspond to mode $\mathbf{H}$ and eigenvalue $\omega + \delta\omega$ correspond to mode $\mathbf{H}+\delta\mathbf{H}$. If both of these modes satisfy the Hermitian eigenvalue problem (3.17), then they must be orthogonal; that is, their inner product, $(\mathbf{H}, \mathbf{H}) + (\mathbf{H}, \delta\mathbf{H})$, is zero. However, if these two modes are spatially bound, then their inner product cannot be zero, as $(\mathbf{H}, \mathbf{H})$ is positive and $(\mathbf{H}, \delta\mathbf{H})$ is arbitrarily small for integration over a finite domain. Thus, a continuous spectrum is incompatible with the orthogonality requirement, unless the modes are of infinite extent.

Third, modes of a Hermitian eigenvalue problem are constrained by a **variational theorem**. Namely, the smallest eigenvalue ω_o^2/c^2 corresponds to a field pattern that minimises the functional

$$U_f(\mathbf{H}) = \frac{(\mathbf{H}, \hat{\Theta}\mathbf{H})}{(\mathbf{H}, \mathbf{H})} \tag{3.23}$$

That is, ω_o^2/c^2 is the minimum of U_f over all conceivable field patterns $\mathbf{H}$ subject to the transversality constrant (3.14). The mode with the next-lowest ω value minimises U_f within the subspace of functions that are orthogonal to all modes with lower ω values. Rewriting U_f in terms of $\mathbf{E}$ using (3.15), (3.16), and (3.20) reveals an important heuristic rule about modes in dielectric media:

$$\begin{aligned} U_f(\mathbf{H}) &= \frac{(\nabla \times \mathbf{E}, \nabla \times \mathbf{E})}{(\mathbf{E}, \varepsilon(\mathbf{r})\mathbf{E})} \\ &= \frac{\int |\nabla \times \mathbf{E}(\mathbf{r})|^2 \mathrm{d}^3\mathbf{r}}{\int \varepsilon(\mathbf{r})|\mathbf{E}(\mathbf{r})|^2 \mathrm{d}^3\mathbf{r}} \end{aligned} \tag{3.24}$$

The heuristic rule is thus: to minimise U_f, the electric field $\mathbf{E}$ should be concentrated in regions of high dielectric constant ε to maximise the denominator and should minimise the number of spatial oscillations to minimise the numerator, all the while remaining orthogonal to lower-frequency modes. Consequently, lower-frequency modes will tend to reside in regions with a higher dielectric constant, whereas higher-frequency modes will tend to reside in regions with a lower dielectric constant. It will be seen in the section "One-Dimensional Photonic Crystals and the Physical Origin of the Photonic Band Gap" that making note of this difference is crucial to understanding the physical origin of a PBG in photonic crystals.

Since photonic crystals are periodic media, mathematical techniques used in solid-state theory can be applied to help solve (3.17). Photonic crystals have *discrete translational* symmetry; that is, they are invariant under translations of distances that are a multiple of some fixed length. The simplest of such a system is a structure illustrated in figure 3.1. This system has discrete translational symmetry in the x direction. The basic step length is described by the **primitive lattice vector** $\mathbf{a} = a\hat{\mathbf{x}}$, where a is the lattice constant. The system also has *continuous translational* symmetry in the y direction, meaning that the system is invariant under translations of *any* distance in the y direction. The dielectric unit cell, which is is repeated over and over, is highlighted in figure 3.1 with a box.

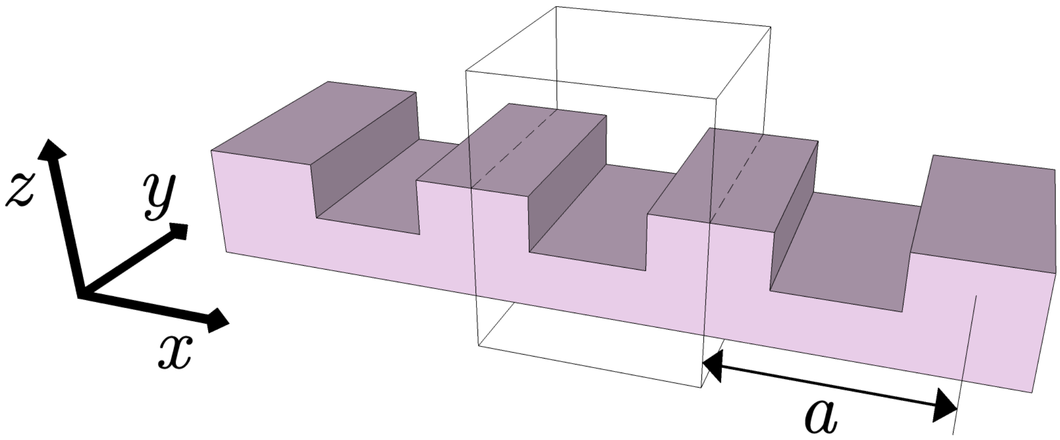


Figure 3.1: A dielectric system with discrete translational symmetry in the x direction. The unit cell of this periodic system is a yz slab with width a in the x direction. If this cell is repeated forever, then shifting the system by an integral multiple of a in the x-direction has no effect on the system's mode solutions.

Because of translational symmetry, modes of $\hat{\Theta}$ can be identified as simultaneous eigenfunctions of translation operators. Let $\hat{T}_{\mathbf{d}}$ represent an operator that translates the system by vector $\mathbf{d}$. Then, translation symmetry is described by the following eigenequation.

$$\begin{aligned} \hat{T}_{\mathbf{d}}\varepsilon(\mathbf{r}) &= \varepsilon(\mathbf{r}-\mathbf{d}) && (3.25) \\ &= \varepsilon(\mathbf{r}) && (3.26) \end{aligned}$$

In the x direction, $\varepsilon(\mathbf{r}) = \varepsilon(\mathbf{r}-\mathbf{R})$ for any $\mathbf{R} = l\mathbf{a}$, where l is an integer. In the y direction, $\varepsilon(\mathbf{r}) = \varepsilon(\mathbf{r}-\mathbf{d})$ for any distance $\mathbf{d} = d\hat{\mathbf{y}}$. The eigenfunctions of these translation operators have the functional form $e^{i\mathbf{k}\cdot\mathbf{r}}$. To wit,

$$\begin{aligned} \hat{T}_{\mathbf{R}}e^{ik_x x} &= e^{ik_x(x-la)} = (e^{-ik_x la})e^{ik_x x} && (3.27) \\ \hat{T}_{d\hat{\mathbf{y}}}e^{ik_y y} &= e^{ik_y(y-d)} = (e^{-ik_y d})e^{ik_y y} && (3.28) \end{aligned}$$

The corresponding eigenvalues are $e^{-ik_x la}$ in the x direction and $e^{-ik_y d}$ in the y direction. Modes are classified by specifying k_x and k_y. Note, however, that not every value of k_x yields a unique eigenvalue. In fact, modes whose k_x values differ by an additive constant $m(2\pi/a)$, where m is an integer, have the same $\hat{T}_{\mathbf{R}}$ eigenvalues and, thus, form a degenerate set. In other words, adding integer multiples of $b = 2\pi/a$ to k_x leaves the mode unchanged. The variable $\mathbf{b} = b\hat{\mathbf{xdis}}$ is called the primitive **reciprocal lattice vector**.

Since $\hat{\Theta}$ is a linear operator, linear combinations of the above degenerate modes can be taken to form new modes which are themselves eigenfunctions of $\hat{\Theta}$. Let eigenvector $\mathbf{H}(\mathbf{r})$ be generalised as a sum of such combinations:

$$\begin{aligned} \mathbf{H}_{k_x,k_y}(\mathbf{r}) &= e^{ik_y y} \sum_m \mathbf{c}_{k_x,m}(z) e^{i(k_x+mb)x} \\ &= e^{ik_x x} e^{ik_y y} \sum_m \mathbf{c}_{k_x,m}(z) e^{imbx} \\ &= e^{ik_x x} e^{ik_y y} \mathbf{u}_{k_x}(x,z) \end{aligned} \tag{3.29}$$

where $\mathbf{c}$'s are expansion coefficients, and $\mathbf{u}(x,z)$ is a periodic function in x. (3.29) is simply the product of a plane wave modulated by a function that is periodic in x due to the crystal lattice. This result is known as **Bloch's theorem**. Modes of the form of (3.29) are known as **Bloch states**. Bloch states with wave vectors k_x that differ in value by mb are identical, and so they share the same eigenvalues $\omega(k_x) = \omega(k_x+mb)$. Thus, from a physical point of view, one needs not consider k_x values that extend beyond a range of one reciprocal lattice vector. This range is conventionally centred about zero, giving important, non-redundant k_x values of $-\pi/a < k_x \leq \pi/a$. This range is called the **Brillouin zone**.

Although it will not be shown, an analogous argument can be made that involves discrete translational symmetry in all three directions. In this case, the system is invariant under translations through a three-dimensional vector $\mathbf{R}$. This vector is comprised of three primitive – and not necessarily cartesian – lattice vectors $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$. Every translation through $\mathbf{R} = l\mathbf{a}_1 + m\mathbf{a}_2 + n\mathbf{a}_3$, where l, m, and n are integers, leaves the system unchanged. The vectors $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ give rise to three primitive reciprocal lattice vectors $(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ following the definition $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$, where δ_{ij} is the Kronecker delta. These reciprocal lattice vectors form a three-dimensional reciprocal lattice within which a three-dimensional Brillouin zone exists. Each wave vector $\mathbf{k} = k_1\mathbf{b}_1 + k_2\mathbf{b}_2 + k_3\mathbf{b}_3$ that falls within this Brillouin zone identifies an eigenvector of $\hat{\Theta}$ with frequency $\omega(\mathbf{k})$ and form $\mathbf{H}_\mathbf{k}$ given by

$$\mathbf{H}_\mathbf{k}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}\mathbf{u}_\mathbf{k}(\mathbf{r}) \tag{3.30}$$

where $\mathbf{u}_\mathbf{k}(\mathbf{r}) = \mathbf{u}_\mathbf{k}(\mathbf{r}+\mathbf{R})$ for all $\mathbf{R}$.

Solutions to (3.17) are perhaps most descriptive when eigenvalues ω are solved in terms of wave-vector $\mathbf{k}$. The family of eigenvalues in terms of $\mathbf{k}$ is referred to as the photonic crystal's **band structure**, or, alternatively, dispersion diagram. This band structure is found by substituting (3.30) into (3.17) and solving for $\mathbf{u}_\mathbf{k}(\mathbf{r})$:

$$\begin{aligned}
\hat{\Theta}\mathbf{H}_\mathbf{k} &= \left(\frac{\omega(\mathbf{k})}{c}\right)^2 \mathbf{H}_\mathbf{k} \\
(i\mathbf{k}+\nabla)\times\frac{1}{\varepsilon(\mathbf{r})}(i\mathbf{k}+\nabla)\times\mathbf{u}_\mathbf{k}(\mathbf{r}) &= \left(\frac{\omega(\mathbf{k})}{c}\right)^2 \mathbf{u}_\mathbf{k}(\mathbf{r}) \\
\hat{\Theta}_\mathbf{k}\mathbf{u}_\mathbf{k}(\mathbf{r}) &= \left(\frac{\omega(\mathbf{k})}{c}\right)^2 \mathbf{u}_\mathbf{k}(\mathbf{r}) \qquad (3.31)
\end{aligned}$$

This is yet another eigenvalue problem with Hermitian operator $\hat{\Theta}_\mathbf{k}$ that depends on $\mathbf{k}$:

$$\hat{\Theta}_\mathbf{k} = (i\mathbf{k}+\nabla)\times\frac{1}{\varepsilon(\mathbf{r})}(i\mathbf{k}+\nabla)\times \tag{3.32}$$

The function $\mathbf{u}_\mathbf{k}(\mathbf{r})$ is an eigenfunction of this operator and is subject to the transversality requirement $(i\mathbf{k}+\nabla)\cdot\mathbf{u}_\mathbf{k} = 0$, as well as the periodicity condition $\mathbf{u}_\mathbf{k}(\mathbf{r}) = \mathbf{u}_\mathbf{k}(\mathbf{r}+\mathbf{R})$. Because of this periodicity, the eigenvalue problem is restricted to a finite volume and, thus, leads to a discrete spectrum of eigenvalues $\omega(\mathbf{k})$ at each value of $\mathbf{k}$. $\mathbf{k}$ itself, however, appears as a continuous parameter in $\hat{\Theta}_\mathbf{k}$, so each eigenvalue can be expected to vary continuously with $\mathbf{k}$. In this way does the band structure arise: it is a family of curves $\omega_n(\mathbf{k})$ indexed in order of increasing frequency by band number n. Within this band structure is most of the information required to predict the optical properties of a photonic crystal.

It has been shown that discrete translational symmetry confines the number of non-redundant $\mathbf{k}$ values to a finite region called the Brillouin zone. If the system also has *rotational* symmetry, then non-redundant $\mathbf{k}$ values are confined to an even smaller region. Suppose that rotational operator $\mathcal{R}(\hat{\mathbf{n}},\phi)$ rotates vectors by an angle

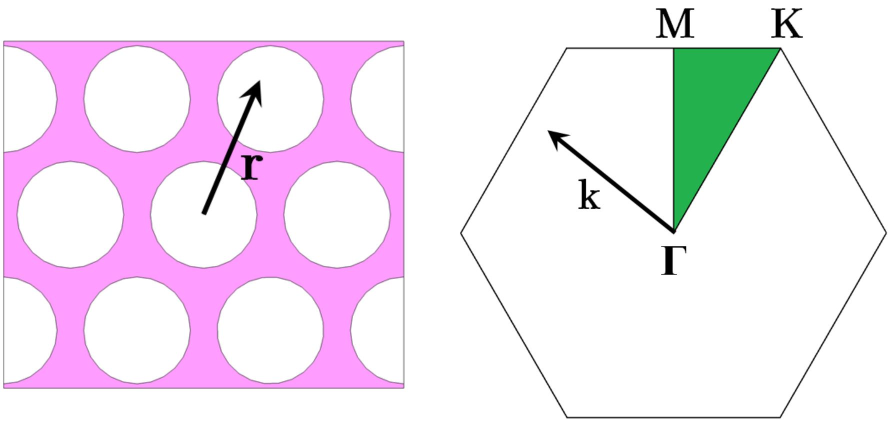


Figure 3.2: (Left) a photonic crystal whose dielectric function is a triangular lattice of holes, as depicted in real space. (Right) the photonic crystal's corresponding Brillouin zone, as depicted in wave-vector space. The area highlighted in green is the irreducible Brillouin zone. Labels M, K, and Γ are high-symmetry points demarcating the corners of the irreducible Brillouin zone. Space vector **r** and wave vector **k** are in arbitrary directions.

ϕ about the $\hat{\mathbf{n}}$ axis. Suppose further that operator $\hat{O}_{\mathcal{R}}$ rotates a vector field and its argument using $\mathcal{R}$. If vector field $\mathbf{H}_{\mathbf{k}n}$ is invariant under rotation by $\mathcal{R}$, then $\hat{O}_{\mathcal{R}}\mathbf{H}_{\mathbf{k}n}$ satisfies (3.17) with the same eigenvalue as $\mathbf{H}_{\mathbf{k}n}$. In other words, the rotated mode itself is an allowed mode in the system and has the same frequency. In fact, the rotated mode corresponds to a Bloch state with wave vector $\mathcal{R}\mathbf{k}$. It follows that $\omega(\mathcal{R}\mathbf{k}) = \omega(\mathbf{k})$. Thus, wave vectors within the Brillouin zone that differ by rotation $\mathcal{R}$ are redundant. The smallest region within the Brillouin zone for which the $\omega_n(\mathbf{k})$ values are not related by symmetry is called the **irreducible Brillouin zone**. For example, a photonic crystal with a triangular lattice of holes as shown in figure 3.2 has a Brillouin zone in the shape of a hexagon. The irreducible zone is a wedge with one-twelfth the area of the full Brillouin zone, corresponding to a dielectric pattern with six-fold rotational symmetry. The rest of the Brillouin zone contains redundant copies of **k** and, for this reason, is often ignored in band structure analysis.

All non-magnetic photonic crystals can be shown to exhibit one more symmetry known as *time-reversal* symmetry. If the complex conjugate of (3.17) is taken, then $\mathbf{H}_{\mathbf{k}n}^*$ is seen to satisfy the same equation as $\mathbf{H}_{\mathbf{k}n}$, with the same eigenvalue ω_n (since ω_n is real). From (3.30), $\mathbf{H}_{\mathbf{k}n}^*$ is simply a Bloch state with $\mathbf{k} \rightarrow -\mathbf{k}$. It follows that

$\omega(\mathbf{k}) = \omega(-\mathbf{k})$. Consequently, a photonic crystal's band structure is mirrored about $\mathbf{k} = 0$. This symmetry is called time-reversal symmetry because taking the complex conjugate of $\mathbf{H}_{\mathbf{k}n}$ is equivalent to reversing the sign of t in the Maxwell equations.

Before closing this section, it must be noted that at no point in the foregoing discussion was a fundamental length scale introduced. For the macroscopic Maxwell equations, there is no fundamental constant with the dimension of length; the equations are *scale-invariant.* This property leads to some simple scaling relationships. For example, let mode $\mathbf{H}(\mathbf{r})$ with frequency ω in a dielectric configuration $\varepsilon(\mathbf{r})$ satisfy (3.17). First, suppose that the dielectric configuration is expanded by some scale constant s such that $\varepsilon'(\mathbf{r}) = \varepsilon(\mathbf{r}/s)$. Making a change of variables $\mathbf{r}' = s\mathbf{r}$ and $\nabla' = \nabla/s$, (3.17) becomes

$$s\nabla' \times \left(\frac{1}{\varepsilon(\mathbf{r}'/s)} s\nabla' \times \mathbf{H}(\mathbf{r}'/s)\right) = \left(\frac{\omega}{c}\right)^2 \mathbf{H}(\mathbf{r}'/s) \tag{3.33}$$

$$\nabla' \times \left(\frac{1}{\varepsilon'(\mathbf{r}')} \nabla' \times \mathbf{H}(\mathbf{r}'/s)\right) = \left(\frac{\omega}{sc}\right)^2 \mathbf{H}(\mathbf{r}'/s) \tag{3.34}$$

This is simply (3.17) with mode profile $\mathbf{H}'(\mathbf{r}') = \mathbf{H}(\mathbf{r}'/s)$ and frequency $\omega' = \omega/s$. Thus, scaling up the dielectric pattern by a multiple s scales up the mode profile and scales down the frequency by that same multiple. Second, suppose that the dielectric constant is scaled down everywhere by a constant s^2 such that $\varepsilon'(\mathbf{r}) = \varepsilon(\mathbf{r})/s^2$. Substituting $s^2\varepsilon'(\mathbf{r})$ into (3.17) for $\varepsilon(\mathbf{r})$ yields

$$\nabla \times \left(\frac{1}{\varepsilon'(\mathbf{r})} \nabla \times \mathbf{H}(\mathbf{r})\right) = \left(\frac{s\omega}{c}\right)^2 \mathbf{H}(\mathbf{r}) \tag{3.35}$$

The mode $\mathbf{H}(\mathbf{r})$ is unchanged, but the frequencies are scaled up by a factor of s. These two scaling relationships are powerful analytical tools: if a photonic crystal is solved at one scale, then it is solved at *every* scale (provided that the assumption that the system is macroscopic is upheld). This is why the band structures of photonic crystals are commonly displayed using normalised axes – $\omega a/2\pi c$ for the ordinate and $ka/2\pi$ for the abscissa – as opposed to absolute axes ω and k. With normalised axes, a single band structure applies to a photonic crystal of any length scale; all that is needed is a substitution for lattice constant a.

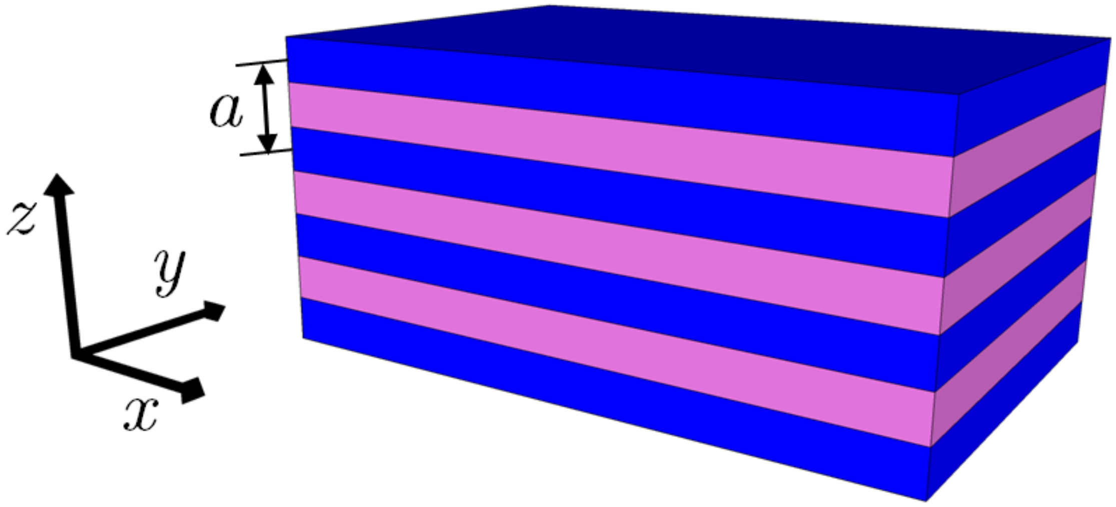


Figure 3.3: A one-dimensional photonic crystal, called thus because its dielectric function varies along only one direction (here z). Each of the two colours, blue and pink, represents a material with a unique dielectric constant. The system's unit cell is an xy slab with length a in the z direction.

3.2 One-Dimensional Photonic Crystals and the Physical Origin of the Photonic Band Gap

For the sake of instruction, the first photonic crystal that is introduced is the simplest of kinds: a one-dimensional photonic crystal. A one-dimensional photonic crystal is identical to the multilayer film presented in classical optics. The multilayer film is classically analysed by considering the sum of reflections and transmissions at dielectric boundaries. Here, however, it is analysed using the photonic-bands theory established in the previous section, which is applicable even in the more complicated two- and three-dimensional cases. In spite of its simplicity, the one-dimensional photonic crystal is shown to exhibit important features in general, such as PBGs and modes that are localised around defects. These two features will be discussed in turn.

A one-dimensional crystal, shown in figure 3.3, consists of alternating layers of material with different dielectric constants. Here, the system's dielectric function is periodic in the z direction and continuous in the xy plane. This allows the system's

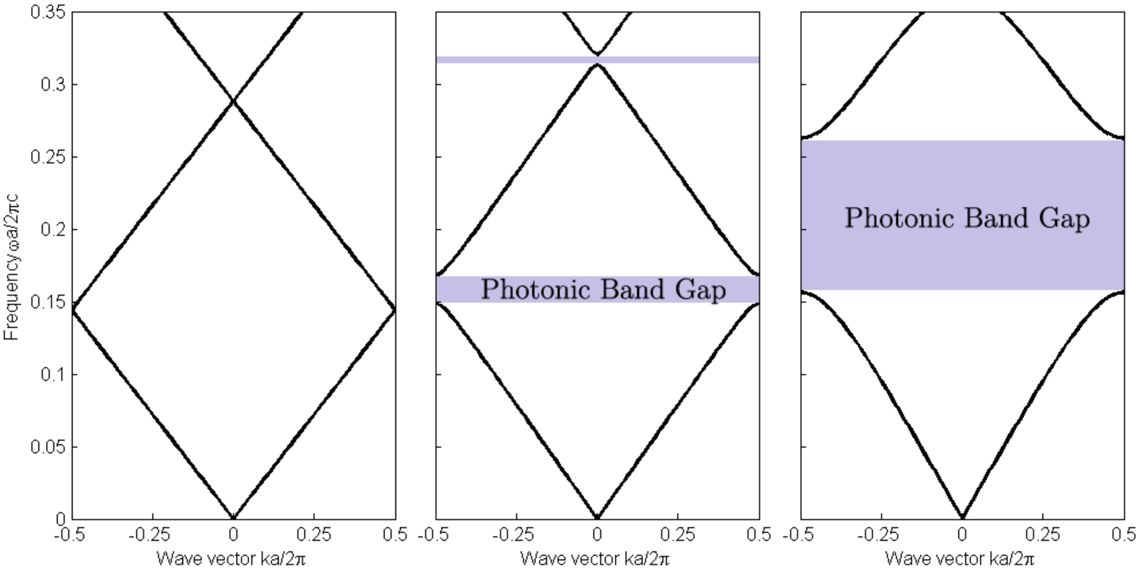


Figure 3.4: Band structures of one-dimensional photonic crystals for on-axis propagation. (Left) every layer has the same dielectric constant $\varepsilon = 12$; (centre) layers alternate between ε of 12 and 8; (right) layers alternate between ε of 12 and 1. In all three cases, each layer has a width of $0.5a$. A dielectric constant of 12 roughly corresponds to that of crystalline silicon at infrared wavelengths.

modes to be classified using two wave vectors: $\mathbf{k}_{||}$ and $\mathbf{k}_z$, where $\mathbf{k}_{||}$ points in the xy plane and $\mathbf{k}_z$ points in the z direction. The system's modes can be written in Bloch form:

$$\mathbf{H}_{n,\mathbf{k}_{||},\mathbf{k}_z} = e^{i\mathbf{k}_{||}\cdot\rho} e^{i\mathbf{k}_z\mathbf{z}} \mathbf{u}_{n,\mathbf{k}_{||},\mathbf{k}_z}(z) \tag{3.36}$$

n is the band number, ρ is a space vector in the xy plane, and $\mathbf{u}(z)$ is a periodic function obeying $\mathbf{u}(z) = \mathbf{u} + \mathbf{R}$, where R is an integral multiple of spatial period a. Since the dielectric function has continuous translational symmetry in the xy plane (ie. it is homogeneous along x and y), $\mathbf{k}_{||}$ may take on any value. $\mathbf{k}_z$, however, is restricted to a finite interval because the dielectric function has discrete translational symmetry in the z direction. This restriction is represented as a Brillouin zone with extent $-\pi/a < k_z < \pi/a$ in reciprocal-lattice space.

The photonic band structure is found by numerically solving (3.31) for combinations ω and $\mathbf{k}$. PBGs manifest only along directions with periodicity, so, for simplicity, only solutions for which $\mathbf{k}_{||} = 0$ are considered. Wave-vector $\mathbf{k}_z$ can thus be relabelled k without confusion. In figure 3.4, the band structure is plotted for photonic crystals with three different dielectric contrasts. The leftmost plot is for a crystal in which every layer has the same dielectric constant of 12, the centre plot is for a crystal with layers of alternating dielectric constants 12 and 8, and the rightmost plot is for a crystal with layers of alternating dielectric constants 12 and 1.

The leftmost plot is for a homogeneous dielectric medium for which an arbitrary period of a has been assigned. Accordingly, this system has a single photonic band given by the dispersion relation for a simple plane wave:

$$\omega(k) = \frac{ck}{\sqrt{\varepsilon}} \tag{3.37}$$

Since it has been insisted that k repeats itself outside the Brillouin zone, a multiple of $2\pi/a$ is added (subtracted) to the solution when it reaches the negative (positive) edge of the Brillouin zone. This is known as *zone-folding*. The centre plot is for a system whose dielectric medium is nearly homogeneous. The centre plot resembles the leftmost one, but has one important difference: the bands at either end of the Brillouin zone are separated. No solutions to (3.31) exist for the range of frequencies between the two bands. This lack of solutions is precisely the PBG. Modes within the PBG are evanescent; that is, their wave-vectors are purely imaginary. If light of a frequency within the PBG strikes the photonic crystal at normal incidence, it will be perfectly reflected. The rightmost plot shows that this PBG widens as the dielectric contrast is increased.

The physical origin of the PBG is explained by considering the electric field profiles immediately above and below the PBG. At the edge of the Brillouin zone, where $|k| = \pi/a$, modes have a wavelength of $2a$, or twice the crystal's lattice constant. Figure 3.5 shows two ways modes of such a wavelength can be oriented in the crystal: one, the nodes of the mode's electric field are centred in each low-ε layer, and, two, the nodes are centred in each high-ε layer. Any other arrangement violates the symmetry of the unit cell about its centre. From the discussion about the variational theorem, it is known that low-frequency modes concentrate their energy (which is proportional to $|\mathbf{E}|^2$) in high-ε regions, while higher-frequency modes, in observance

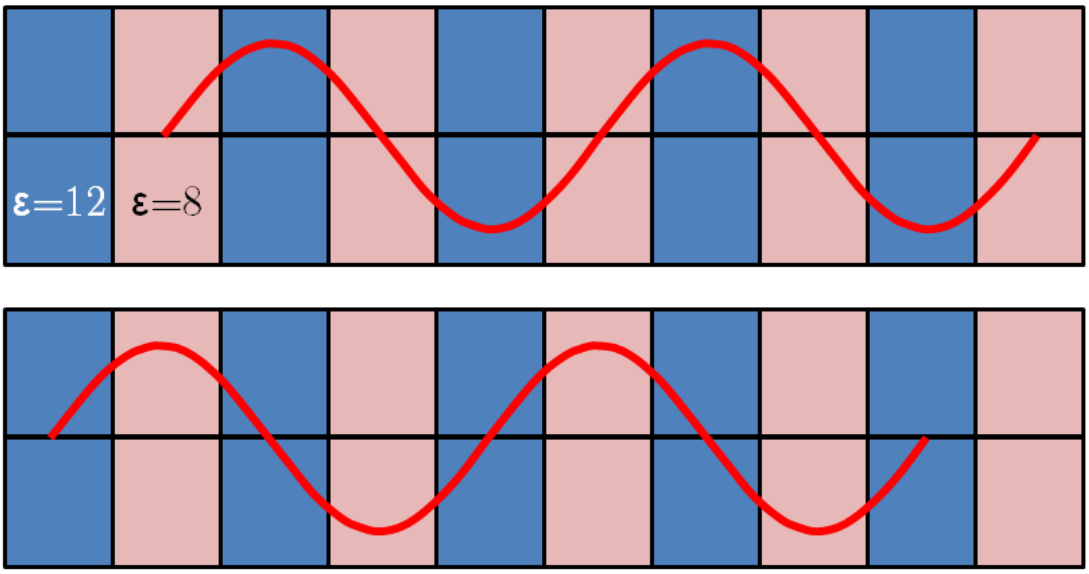


Figure 3.5: The approximate electric field profiles of modes at the edge of the Brillouin zone of a one-dimensional photonic crystal. (Top) the electric field of the mode in the lower band; (bottom) the electric field of the mode in the upper band. The mode in the lower band has a higher electric-field density in the regions with a higher dielectric constant than the mode in the upper band.

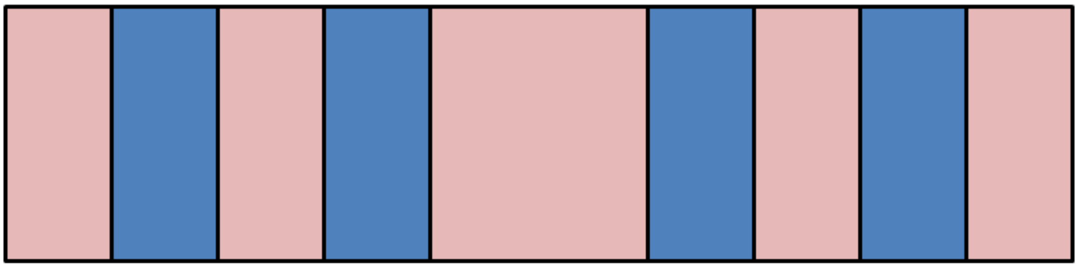


Figure 3.6: A defect in a one-dimensional photonic crystal, formed by increasing the thickness of a single layer.

of the orthogonality condition, are forced to concentrate a larger portion of their energy in low-ε regions. With this in mind, it is evident why there is a difference in frequency between the two cases. The mode under the gap has more energy concentrated in the high-ε region, giving it a lower frequency, while the mode above the gap has more energy concentrated in the low-ε region, giving it a higher frequency. As the dielectric contrast between the two regions is increased, the denominator of the variational theorem scales more steeply with electric-field energy, making this difference in frequency more pronounced.

While modes within a PBG cannot be extended in general, it is still possible to support modes within a PBG that are *localised* around a defect. A defect may be formed by increasing the thickness of a single layer, as shown in figure 3.6. Doing so breaks discrete translational symmetry along the z direction and, thus, prevents modes from being described using k_z. Nevertheless, one may intuitively conclude that, since periodicity is maintained everywhere else, modes still cannot be extended on either side of the defect. If a mode within the defect has a frequency in the PBG, then it must exponentially decay when it crosses into either side. Both sides act as perfectly reflecting mirrors, and the defect behaves as a resonant cavity. If off-axis wave vectors are included, then modes remain localised in the z direction, but propagate in the other directions. Thus, the defect can act as a planar waveguide.

Localising light around a defect entails that the localised modes become quantised in frequency. As the thickness of the defect layer is increased, the frequencies of the localised modes *decrease* because the modes have more space to oscillate; equivalently, the numerator of the variational theorem decreases. In effect, this *pulls down* modes from the upper band into the PBG. On the other hand, if the thickness is kept constant, but the dielectric constant is decreased, then the frequencies of the localised modes *increase* due to a decrease in the denominator of the variational theorem. This effectively *pushes up* modes from the lower band into the PBG. As will be shown in the next section, pushing or pulling modes into the PBG is the key to achieving waveguiding in photonic crystals.

3.3 Two-Dimensional Photonic Crystals and Polarization Dependence of Band Gaps

A two-dimensional photonic crystal is periodic in two dimensions and continuous along the third. One such crystal is illustrated in figure 3.7. This crystal is assumed to be infinitely extended along the z direction. (The case of finite extension along the z direction is considered in the next section.) Before this crystal's dielectric function is even substituted into the Maxwell equations, it is known *a priori* from symmetry that this crystal's modes take the form of Bloch states:

$$\mathbf{H}_{n,\mathbf{k}_{||},\mathbf{k}_z} = e^{i\mathbf{k}_{||}\cdot\rho} e^{i\mathbf{k}_z\cdot\mathbf{z}} \mathbf{u}_{n,\mathbf{k}_{||}\mathbf{k}_z}(\rho) \tag{3.38}$$

This form is similar to that for a one-dimensional crystal, only that it is now $\mathbf{k}_{||}$ that is restricted to a Brillouin zone, while $\mathbf{k}_z$ can take on any value. Also, this Brillouin zone is now two-dimensional. ρ is a space vector in the plane of periodicity – that is, the xy plane – and $\mathbf{u}(\rho)$ is a periodic function obeying $\mathbf{u}(\rho) = \mathbf{u}(\rho + \mathbf{R})$, where $\mathbf{R}$ is any linear combination of primitive lattice vectors $a\hat{\mathbf{x}}$ and $a\hat{\mathbf{y}}$.

A two-dimensional photonic crystal is able to prohibit the propagation of light in any direction within its plane of periodicity. Modes propagating strictly within this plane can be categorised into two distinct polarisations: transverse-electric (TE) and transverse-magnetic (TM). TE modes have their $\mathbf{H}$ components perpendicular to the plane, or $\mathbf{H} = H(\rho)\hat{\mathbf{z}}$ and $\mathbf{E} \cdot \hat{\mathbf{z}} = 0$. TM modes are just the opposite: their

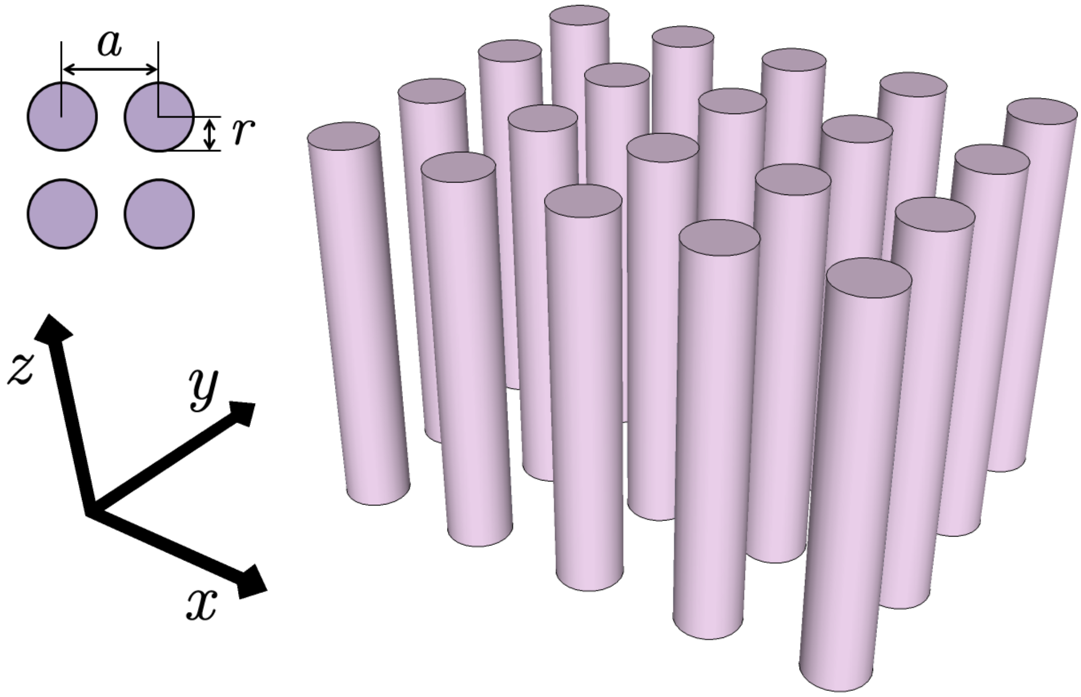


Figure 3.7: A two-dimensional photonic crystal consisting of a square lattice of dielectric columns with radius r. The columns are periodic along the x and y directions with lattice constant a. The left inset shows the photonic crystal from above.

$\mathbf{E}$ components are perpendicular to the plane, or $\mathbf{E} = E(\rho)\hat{\mathbf{z}}$ and $\mathbf{H} \cdot \hat{\mathbf{z}} = 0$. In short, *transverse* describes the field component that lies in the plane of periodicity. The band structures of these two polarisations are generally different. For example, a photonic crystal may have a PBG for one polarisation but not the other. The polarisation for which a crystal has a PBG depends on crystal's dielectric function, of which there are many varieties. Three of these varieties – a square lattice of dielectric columns, a square lattice of dielectric veins, and a triangular lattice of holes – are investigated below.

Consider a square lattice of dielectric columns, such as the one depicted in figure 3.7, with lattice constant a. The rods are made of silicon ($\varepsilon_{Si} = 12.0$), have a radius of $r = 0.2a$, and are suspended in air ($\varepsilon_{air} = 1.0$). The TE and TM band structures of this crystal are plotted in figure 3.8. The abscissa of this plot shows the value of $\mathbf{k}_{||}$ as it moves along the edges of the irreducible Brillouin zone. The reason why only the edges is considered is because the maxima and minima of the band structures almost always appear at these edges. As shown in the left inset, the square lattice has a square Brillouin zone, of which a triangular wedge of one-eighth its area is irreducible. The remaining seven-eighths of its area can be related to this wedge by rotational symmetry. The corners of this wedge are labelled with *high-symmetry* points Γ, X, and M. In cartesian components, these points correspond to wave vectors $\mathbf{k}_{||} = 0$, $\mathbf{k}_{||} = \pi/a\hat{\mathbf{x}}$, and $\mathbf{k}_{||} = \pi/a\hat{\mathbf{x}} + \pi/a\hat{\mathbf{y}}$, respectively.

This photonic crystal is seen to exhibit a TM bandgap, but no TE bandgap. This fact can be explained by examining the field patterns at the high-symmetry points. Figure 3.9 depicts the magnitude of the scalar field component – E_z for TM modes and H_z for TE modes – at the X high-symmetry point. For modes at this point, the scalar field oscillates in sign between adjacent unit cells in the x direction. For TM modes, the mode's E_z field in the lower band is maximum at the centre of the dielectric rods (left figure) while the mode's E_z field in the upper band is maximum between the rods (right figure). Put another way, a nodal plane runs between the rods for the lower band, and across the rods for the upper band. As a result, a significant fraction of electric-field energy of the mode in the lower band is concentrated in regions with a high dielectric constant, whereas a much smaller fraction is concentrated in

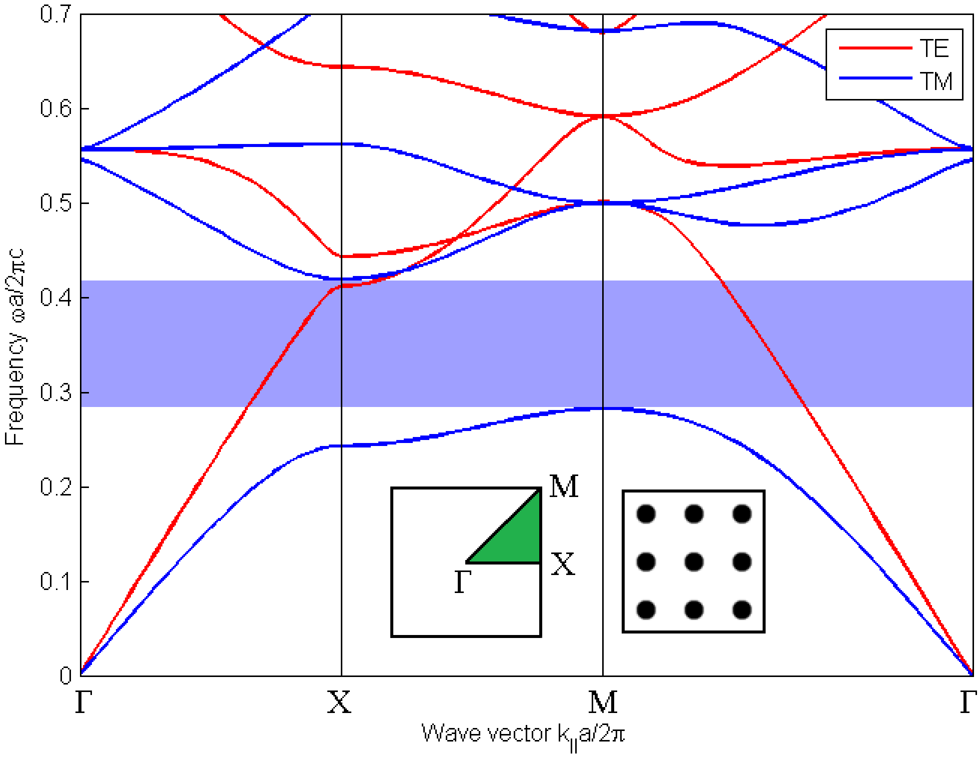


Figure 3.8: The photonic band structure of a two-dimensional photonic crystal consisting of a square lattice of dielectric rods with radius $r = 0.2a$ and dielectric constant $\varepsilon = 12.0$ in air. The red and blue bands represent TE and TM modes, respectively. The blue-shaded region identifies a TM band gap between normalised frequencies 0.28 and 0.42. There is no TE band gap. The left inset shows the lattice's Brillouin zone, with the irreducible region shaded in green. Labels Γ, X, and M are high-symmetry points identifying the corners of the irreducible region. The right inset shows the lattice's dielectric function, with dark-shaded regions denoting areas with dielectric constant ε.

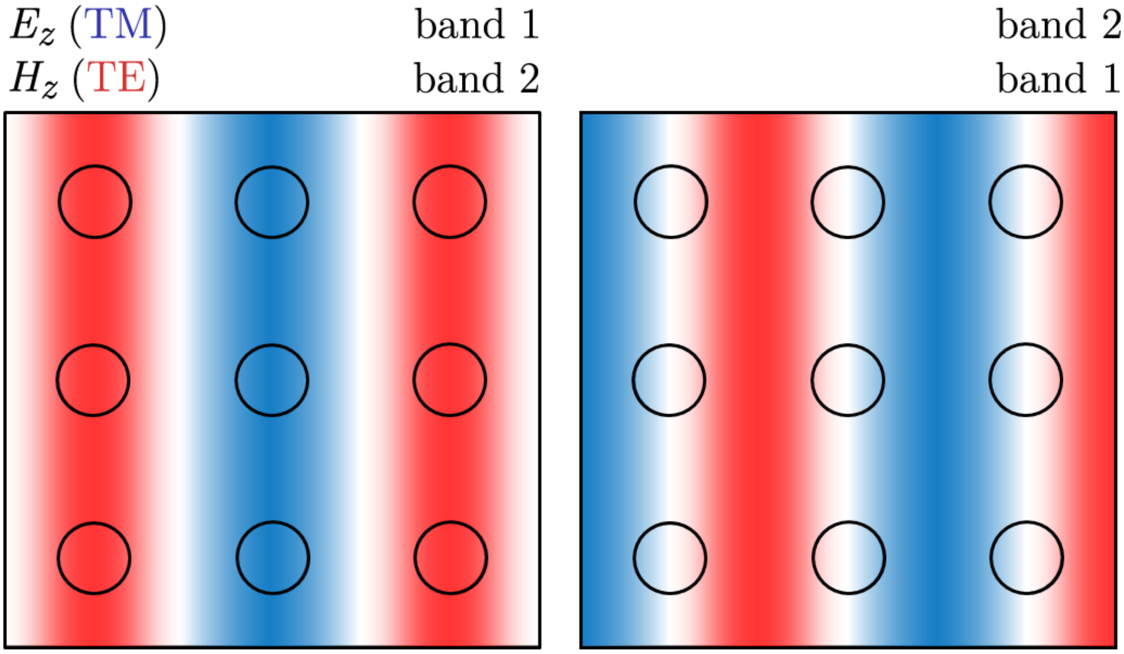


Figure 3.9: Depictions of the scalar field patterns of the first two bands at the X-point of a square lattice of dielectric rods at a particular instant in time. For TM and TE modes, the scalar fields are E_z and H_z, respectively. Colour represents the magnitude of the scalar field, with red and blue being positive and negative, respectively. The black edges indicate the rods. In the context of the discussion in the text, bands 1 and 2 are the lower and upper bands, respectively.

these regions for the mode in the upper band. As was established in the previous section, this discrepancy in electric-field energy concentration leads to a splitting of frequencies between bands. Thus, the two TM bands are appreciably separated in frequency at the X point. Although it will not be shown, field patterns at the M high-symmetry point exhibit a similar discrepancy, so a complete bandgap appears for TM modes.

For TE modes, the discrepancy in electric-field energy concentrations is much smaller. In this case, it is the modes' H_z components that are scalar. The H_z component is maximum between the rods for the mode in the lower band (right figure) and in the centre of the rods for the mode in the upper band (left figure). From (3.18), it is known that the electric field tends to be maximum along the nodal planes of the magnetic field $\mathbf{H}$. Evidently, this implies that the electric-field energy of the mode in the upper band is primarily concentrated in the air regions, as its nodal plane lies entirely in these regions. In contrast, the nodal plane of the mode in the lower band cuts across the dielectric rods, but because the electric field must be continuous everywhere, the electric field is forced to extend into the air regions, as well. In other words, the electric field, in spite of having its maximum in the vicinity of the rods, cannot be contained entirely within the rods as it is for TM modes. Thus, the electric-field energy is concentrated primarily in the air regions for both bands, leading to TE modes that are higher in frequency than their TM equivalents and the lack of a TE bandgap.

Another two-dimensional photonic crystal is a square lattice of dielectric veins. Consider one such crystal whose veins have a thickness of $0.165a$ and a dielectric constant of $\varepsilon = 12.0$. The regions between the veins are filled with air ($\varepsilon = 1.0$). The TE and TM band structures of this crystal are plotted in figure 3.10. Here, the crystal has a TE bandgap, but no TM bandgap. Again, this fact can be explained by investigating the field patterns at high-symmetry points. The scalar field components – E_z for TM modes and H_z for TE modes – for two possible arrangements at point X are depicted in figure 3.11. For TM modes, the E_z field is concentrated along the vertical veins for the lower band (left figure) and across the horizontal veins for the upper band (right figure). Modes in both bands are able to concentrate a significant

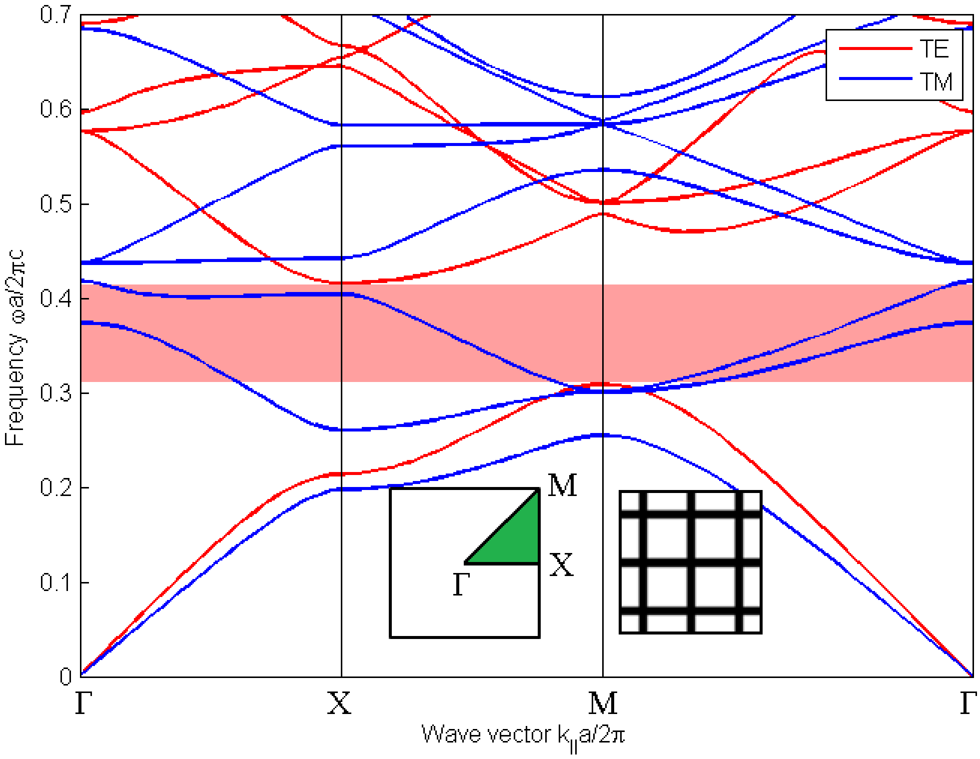


Figure 3.10: The photonic band structure of a two-dimensional photonic crystal consisting of a square lattice of dielectric veins with thickness $0.165a$ and dielectric constant $\varepsilon = 12.0$ in air. The red and blue bands represent TE and TM modes, respectively. The red-shaded region identifies a TE band gap between normalised frequencies 0.31 and 0.42. There is no TM band gap. The left inset shows the lattice's Brillouin zone, with the irreducible region shaded in green. The right inset shows the lattice's dielectric function, with dark-shaded regions denoting areas with dielectric constant ε.

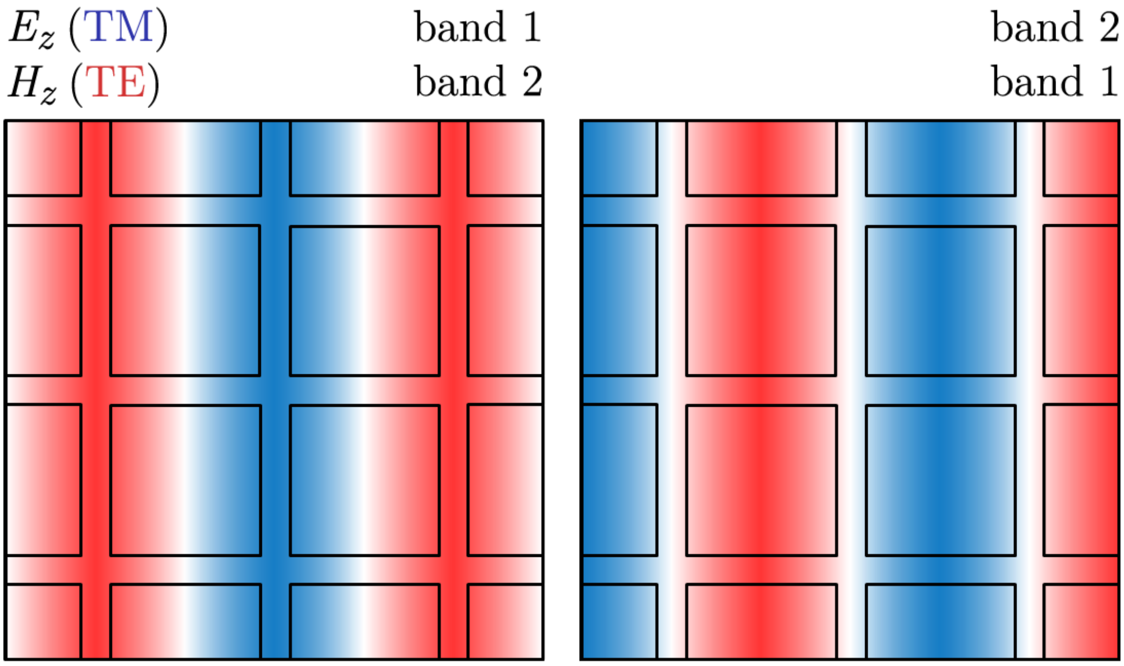


Figure 3.11: Depictions of the scalar field patterns of the first two bands at the X-point of a square lattice of dielectric veins at a particular instant in time. For TM and TE modes, the scalar fields are E_z and H_z, respectively. Colour represents the magnitude of the scalar field, with red and blue being positive and negative, respectively. The black edges indicate the veins. In the context of the discussion in the text, bands 1 and 2 are the lower and upper bands, respectively.

fraction of their electric-field energy in the high-ε veins, so there is no large splitting of frequency between bands. Thus, a TM band gap does not appear. For TE modes, the H_z field is concentrated across the horizontal veins for the lower band (right figure) and along the vertical veins for the upper band (left figure). Consequently, a nodal plane runs along a dielectric vein for the lower band, and cuts across dielectric veins for the upper band. This implies that electric field's maxima are contained entirely within the high-ε region for the lower band, and cut across high-ε regions and into the surrounding low-ε air regions for the upper band. Hence, the lower band is able to concentrate a large fraction of its energy in the high-ε regions, whereas the upper band cannot. This discrepancy in concentration leads to a splitting of frequencies between bands and the production of a TE band gap.

Investigations of a lattice of rods and a lattice of veins lead to two important observations: TM bandgaps tend to appear in a lattice of isolated high-ε regions, and TE bandgaps tend to appear in a connected lattice. By combining these observations, a photonic crystal can be designed to have bandgaps for *both* polarisations. If these two bandgaps are made to overlap, then the crystal has a complete bandgap.

For a photonic crystal to have a complete bandgap, it must have isolated high-ε regions to produce a TM gap, yet it must remain connected to produce a TE gap. A triangular lattice of air columns in a dielectric substrate, shown in figure 3.12, marries these two seemingly disparate qualities. If the air columns have a sufficiently wide radius ($r > 0.4a$), then the dielectric regions between three neighbouring columns appear isolated, providing the isolation necessary for a TM gap. On the other hand, if the air columns are not too wide ($r < 0.5a$), then thin dielectric veins between every two adjacent columns are linked together, providing the connectivity necessary for a TE gap. The band structure of such a photonic crystal, with column radius $r = 0.45a$ in a substrate with dielectric constant $\varepsilon = 12.0$, is shown in figure 3.13. For this arrangement, bandgaps for both TE and TM polarizations exist and overlap in frequency. Within this frequency overlap, the in-plane propagation of any mode is completely prohibited.

As with one-dimensional photonic crystals, the lattice of two-dimensional crystals may be perturbed to support defect states. As depicted in figure 3.14, a single rod can

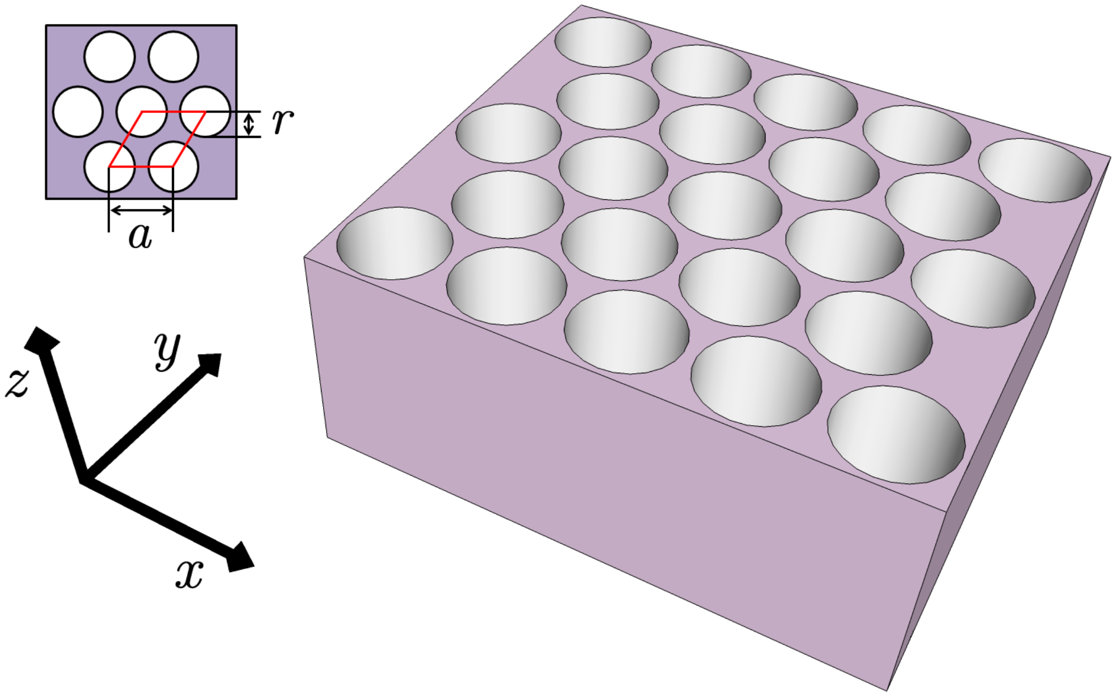


Figure 3.12: A two-dimensional photonic crystal consisting of a triangular lattice of air columns with radius r in a dielectric substrate. The left inset shows the photonic crystal from above, with the unit cell outlined in red. This unit cell has a lattice constant of a.

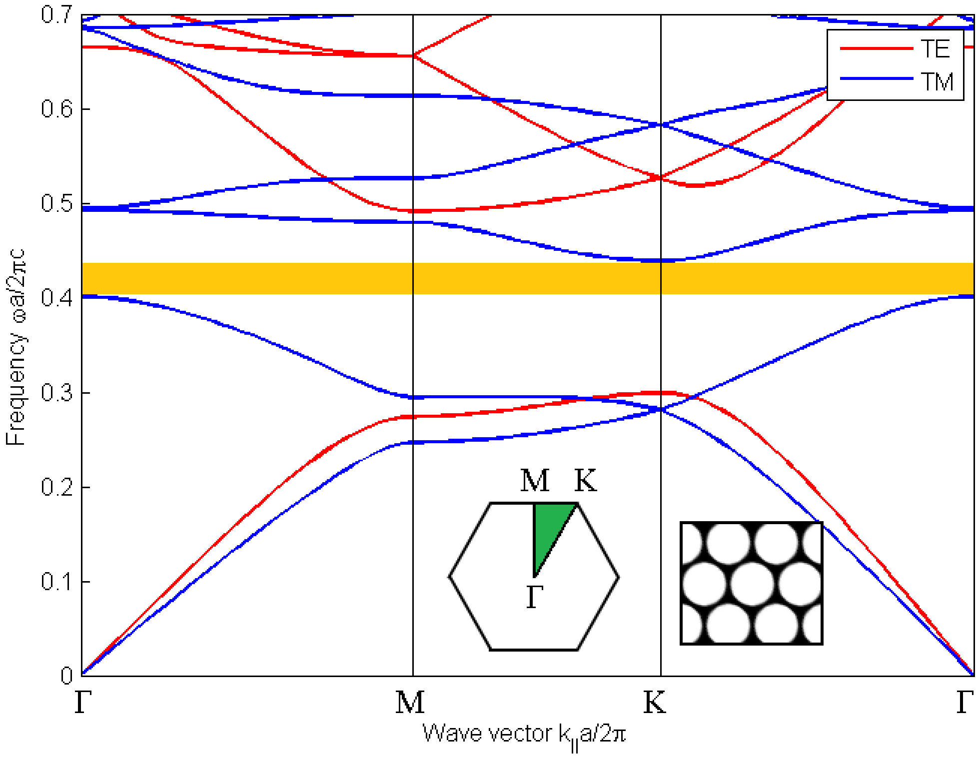


Figure 3.13: The photonic band structure of a two-dimensional photonic crystal consisting of a triangular lattice of air columns with radius $r = 0.45a$ in a substrate with dielectric constant $\varepsilon = 12.0$. The red and blue bands represent TE and TM modes, respectively. The gold-shaded region identifies a complete band gap between normalised frequencies 0.40 and 0.44. The left inset shows the lattice's Brillouin zone, with the irreducible region shaded in green. Labels Γ, M, and K are high-symmetry points identifying the corners of the irreducible region. The right inset shows the lattice's dielectric function, with dark-shaded regions denoting areas with dielectric constant ε.

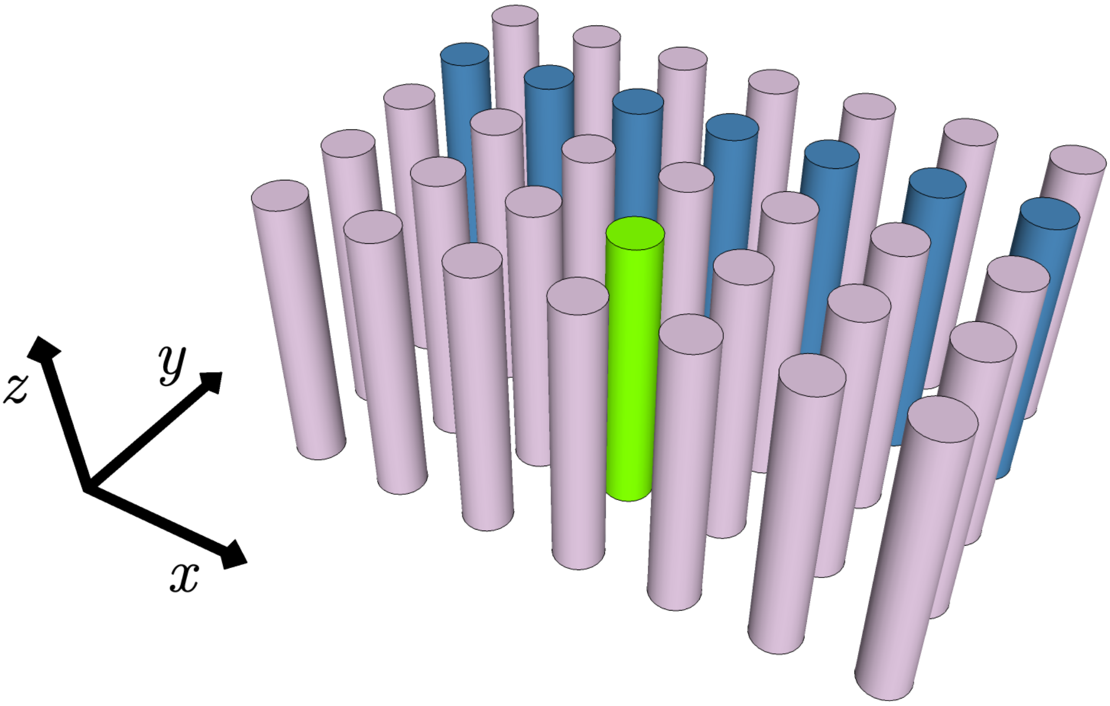


Figure 3.14: Illustration of possible defects in a two-dimensional photonic crystal. Modifying one column (green) might permit a defect state to be localised in both x and y directions. Modifying a row of columns (blue) might permit a state to be localised in one direction (here x).

be perturbed to form a point defect, or a series of rods can be perturbed to form a line of defects (or a *line defect*). Rods can be perturbed by changing their shape, radius, or dielectric constant. Rods can even be removed entirely, which is equivalent to reducing their radii to zero or, alternately, setting their dielectric constants to match the background. For certain perturbations, the defect states can have frequencies that lie within the crystal's PBG. States of such frequencies are confined in both x and y directions around a point defect, or are guided along the axis of a line defect.

As mentioned before, reducing the average dielectric constant may effectively push up modes into the PBG, whereas increasing the average dielectric constant may pull down modes into the PBG. This distinction is important for point defects, as the mode that is localised around the defect has the same field pattern – or the same number of nodal planes – as the band from which it originates. In general, modes that are pushed upward have fewer nodal planes than modes that are pulled downward.

Special attention must be drawn to line defects. By removing a row of columns, the photonic crystal loses discrete translational symmetry in one direction. Only in the direction of the line defect's axis is discrete translational symmetry preserved. Here, the line defect's axis coincides with the x direction. Consequently, k_x remains a conserved quantity, while k_y does not. This has an important ramification on the crystal's band structure: modes may now freely couple into any other mode for which k_x is the same, but k_y is different. Thus, when calculating whether a mode with coordinates (k_x, ω) lies within the PBG, one must check whether there is *any* value of k_y that places the mode on a band. If there is some k_y such that $\omega_n = \omega(k_x, k_y)$, then the mode lies on a band and is extended in the crystal. This process of examining every value of k_y for a given value of k_x is called **projecting the band structure**. As an example, the projected band structure of triangular lattice of holes with a linear defect is shown in figure 3.15. The green curves are guided modes. The red-shaded region is the projection of all in-plane ($k_z = 0$) states onto their (k_x, ω) coordinates. Modes lying within this region do not propagate along the line defect. Rather, they leak into the crystal.

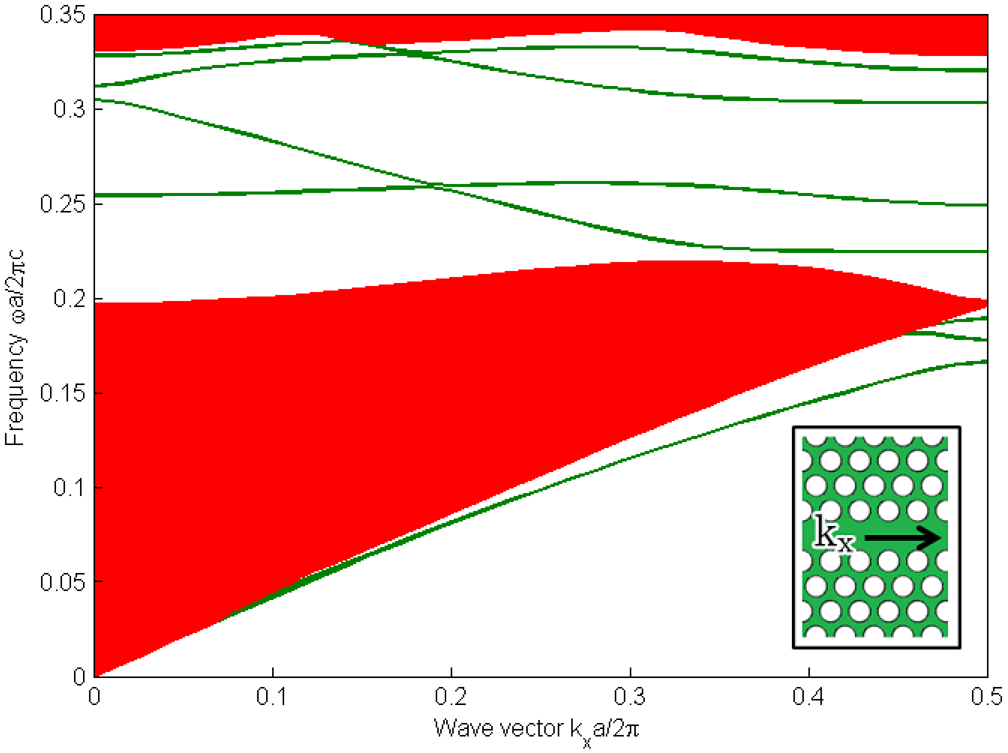


Figure 3.15: The projected band structure of a line defect formed by removing a row of holes in a triangular lattice of holes, as indicated in the inset. The holes have a dielectric constant of 1.0 and a radius of $0.35a$. The background has a dielectric constant of 12.0. Only TE modes are shown. Extended modes fill a continuous region, which is shaded in red. Green curves indicate modes that are guided along the line defect. The green curves lying below the extended modes are confined by total internal reflection as opposed to the PBG.

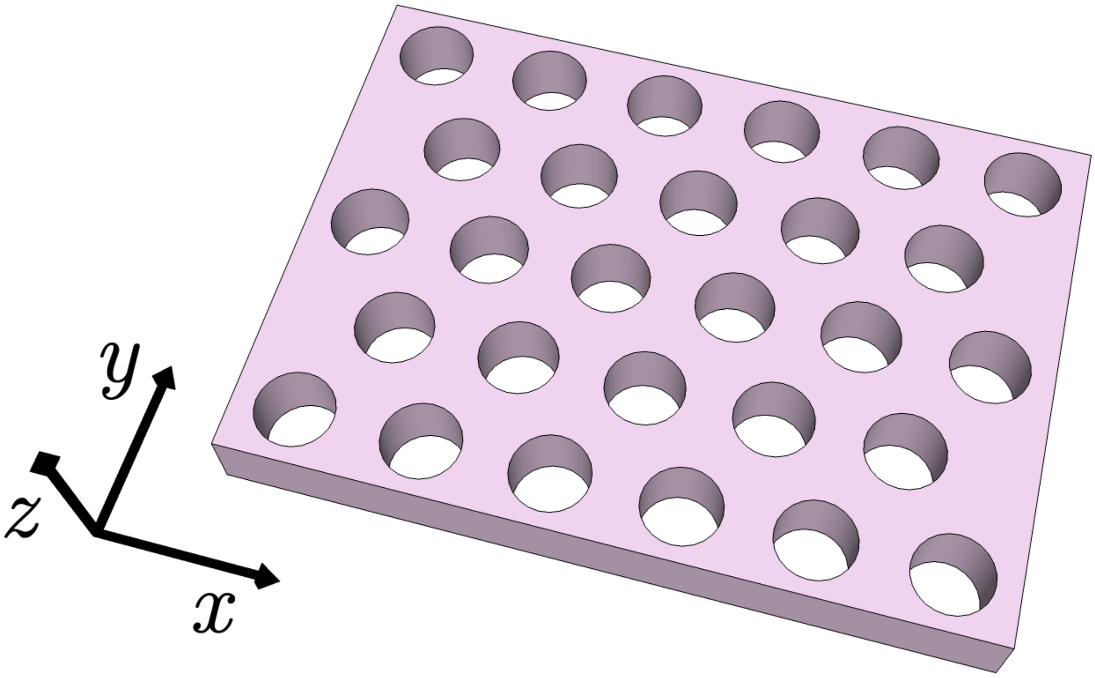


Figure 3.16: Example of a photonic crystal slab, which combines two-dimensional periodicity in the x and y directions and index-guiding in the z direction.

3.4 Photonic Crystal Slabs

The discussion of photonic crystals concludes with a brief treatment of dielectric structures with two-dimensional periodicity and finite thickness. These structures, such as the one shown in figure 3.16, are known as **photonic crystal slabs**. By combining PBGs in the x and y directions and index-guiding in the z direction, these slabs can be used to confine light in three dimensions. A line of defects in the slabs can be used to form waveguides with two-dimensional confinement. Above all, as three-dimensional objects, slabs are experimentally realisable. While they resemble two-dimensional photonic crystals, their finite thickness introduces a *light cone* to their band structure, which may cause light to leak vertically out of the slab. This light cone and other attendant effects are described in this section.

Before discussing the light cone, it pays to introduce the concept of **total internal reflection**. Classically, this phenomenon is explained using ray optics, in which a ray

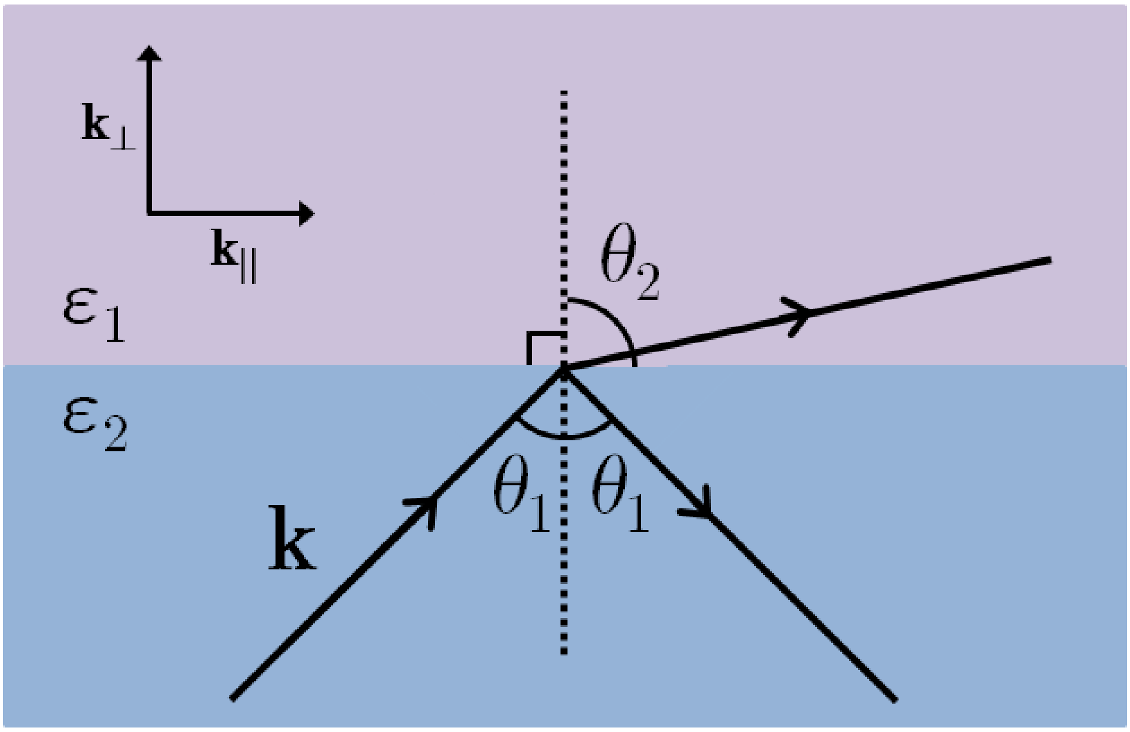


Figure 3.17: A ray-optics representation of light that approaches an interface between two dielectrics ε_1 and ε_2 at angle θ_1 from within dielectric ε_1. In general, light is partially reflected back into dielectric ε_1 at angle θ_1 and partially refracted into dielectric ε_2 at angle θ_2. Given θ_1, θ_2 can be obtained using Snell's law. However, when $\varepsilon_1 > \varepsilon_2$, there are values of θ_1 for which there are no solutions for θ_2. This fact is a consequence of translational symmetry, which dictates that wave-vector component $k_{||}$ must be conserved.

strikes an interface between two media of different dielectric constants at too steep an angle and is completely reflected. If the angle is not too steep, the ray is partially reflected into the same medium and partially refracted into the other medium. The refraction of a ray of light at an interface between dielectrics ε_1 and ε_2 is illustrated in figure 3.17. The ray of light approaches the interface at an angle of θ_1, is partially reflected at this same angle θ_1 back into dielectric ε_1, and is refracted at angle θ_2 into dielectric ε_2. These angles are described by **Snell's law**, $n_1 \sin\theta_1 = n_2 \sin\theta$, where $n_1 = \sqrt{\varepsilon_1}$ and $n_2 = \sqrt{\varepsilon_2}$. If $\varepsilon_1 > \varepsilon_2$, then a value of $\theta_1 > \sin^{-1}(n_2/n_1)$ necessitates that $\sin\theta_2 > 1$, for which there is no real solution. The value $\theta_c = \sin^{-1}(n_2/n_1)$ is referred to as the **critical angle**. The classical interpretation of this circumstance is that the approaching ray is completely reflected from the interface.

Snell's law is actually a combination of two conservation laws that follow from symmetry: the conservation of frequency ω from time-inversion symmetry of Maxwell's equations and the conservation of wave vector $\mathbf{k}_{||}$ from translational symmetry parallel to the interface. These two values are given by $k_{||} = |\mathbf{k}|\sin\theta$ and $|\mathbf{k}| = n\omega/c$. Since $\mathbf{k}_{||}$ is conserved, it must retain the same value on either side of the interface. In other words,

$$\begin{aligned} \left(k_{||}\right)_1 &= \left(k_{||}\right)_2 \\ \frac{n_1\omega}{c}\sin\theta_1 &= \frac{n_2\omega}{c}\sin\theta_2 \\ n_1\sin\theta_1 &= n_2\sin\theta_2 \end{aligned}$$

which is precisely Snell's law. If $\theta_1 > \theta_c$, then $k_{||}$ exceeds the maximum possible value that it can assume in dielectric ε_2, which is $k_{||} = n_2\omega/c$. Since the conservation of $k_{||}$ must be upheld, light is consequently forbidden from propagating in dielectric ε_2. The advantage of thinking about total internal reflection in this way is that ray optics, which is valid only on length scales much larger than the wavelength of light, is not invoked.

The concept of total internal reflection may now be applied to photonic crystal slabs. A slab has a higher dielectric constant than its surrounding medium, has in-plane translatonal symmetry, and is solved using Maxwell's equations, so the same conservation laws from which Snell's law is obtained apply here. To begin, consider modes that are not confined to the slab. These modes cross the dielectric interface into the surrounding medium and extend vertically to infinity. Far from the interface, these modes must resemble plane waves with frequency $\omega = c|\mathbf{k}| = c\sqrt{k_{||}^2 + k_{\perp}^2}$ for some real-valued perpendicular wave vector component $k_{\perp}$. In the context of figure 3.16, $k_{\perp} = k_z$. This quantity is not conserved because the slab, being of finite thickness, does not have translational symmetry in the z direction. Thus, $k_{\perp}$ can take on any value. Consequently, for a given value of $k_{||}$, there are an infinite number of modes for which $\omega > ck_{||}$. The value $\omega = ck_{||}$ is the **light line**, and the region of the band structure above it ($\omega > ck_{||}$) is the **light cone**. Modes in the light cone are solutions to Snell's law with incident angles smaller than the critical angle. Conversely, modes outside of the light cone, or below the light line, are solutions with incident angles greater than the critical angle. These modes are totally internally reflected and are, thus, confined to the slab, or **index-guided**. In this work, only these index-guided modes are pertinent to the analysis of photonic-crystal-slab band structures.

The band structure of a photonic crystal slab consisting of a triangular lattice of holes is shown in figure 3.18. The slab region and surrounding region have dielectric constants of 12.0 and 2.1, respectively. Since the slab modes partially extend into the surrounding region, the effective in-plane dielectric constants experienced by the modes is a weighted average of these two values. The purple-shaded region in the band structure is the light cone for which $\omega > c\mathbf{k}_{||}$. Below the light cone are discrete guided bands that decay exponentially away from the slab. It must be noted that modes in a photonic crystal slab are only truly TE- or TM-polarised in xy-planes for which the system has mirror symmetry. Here, the system has mirror symmetry only in the middle of the slab (ie. where $z = 0$). Away from this plane, both E_z and H_z components are generally finite. Nevertheless, if the slab is sufficiently thin, modes may be *mostly* TE- or TM- polarised. This is noted by referring to bands as "TE-like" and "TM-like," rather than "TE" and "TM" as before. There is a TE-like PBG between the first and second TE-like bands, although, in this case, this PBG applies only to guided modes beneath the light cone. For this reason, this PBG is called an **incomplete bandgap**.

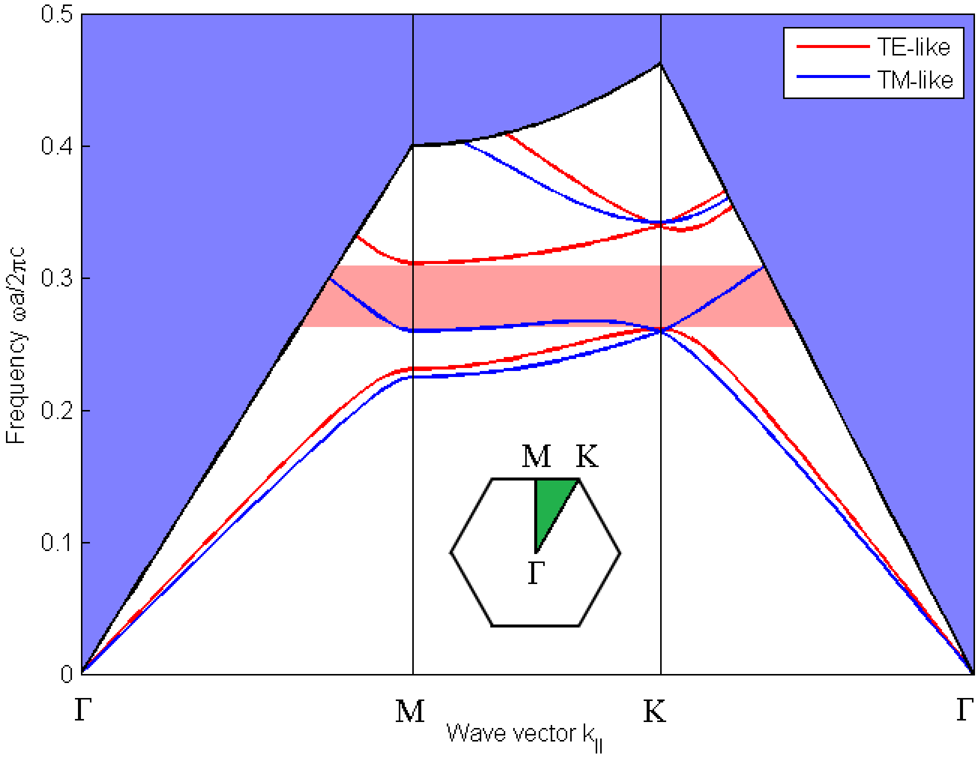


Figure 3.18: The photonic band structure of a photonic crystal slab consisting of a triangular lattice of holes with radius 0.35a. The slab and surrounding medium have dielectric constants of 12.0 and 2.1, respectively. An incomplete PBG for TE-like modes exists between normalised frequencies 0.26 and 0.31. The purple-shaded region is a light cone in which modes are extended into the surrounding medium.

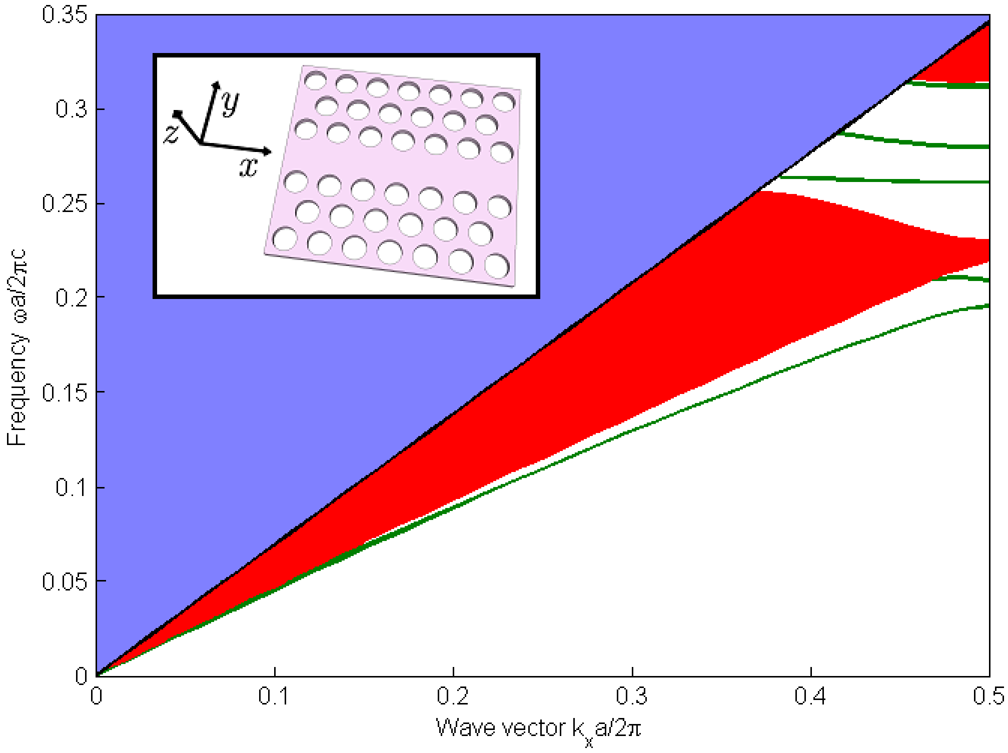


Figure 3.19: The projected band structure of TE-like modes for a line defect formed by removing a row of holes in a triangular lattice of holes, as indicated in the inset. The holes have a radius of $0.35a$. The slab has a dielectric constant of 12.0, a thickness of $0.22a$, and is surrounded by a medium with a dielectric constant of 2.1. Red-shaded regions indicate modes that are extended in the crystal, and the purple-shaded region indicates modes that are extended out-of-plane. Green curves indicate modes that are guided along the line defect. The green curves lying below the extended modes are confined by total internal reflection in the y direction as opposed to the PBG.

As before, a row of unit cells can be perturbed in a photonic-crystal slab to pull or push guided modes into the PBG. By perturbing the periodicity so, PBG confinement in the y direction may be combined with total internal reflection in the z direction to create truly one-dimensional waveguiding. A slab that achieves a coincidence of these two confinement techniques is an example of a PCW. With neither the k_y or k_z components of the wave vector conserved, this system's band structure exhibits both projection into the y direction and a light cone. This is illustrated in figure 3.19, which is the projected band structure of a hole slab in which one row of holes has been filled. Compare figure 3.19 with figure 3.15, which is the projected band structure for a two-dimensional photonic crystal. In figure 3.15, at least one guided mode exists for nearly every frequency in the PBG. Here, however, the guided modes span over a much smaller frequency range due to the presence of the light cone. This raises an important consideration when designing PCWs: for maximum bandwidth – here meaning the range of frequencies over which a mode is guided – one should clad the PCW with a material with as low of a dielectric constant as possible. By doing so, the light cone is lifted to higher frequencies, which, thereby, makes more of the PBG usable for waveguiding. Air is a popular choice of cladding for this very reason, even though an air-bridge structure is not as structurally rigid as its alternatives.

3.5 Slow Light in Photonic Crystals

Group velocity and group-velocity dispersion of modes can be obtained from the band structure. In general, the group velocity, $\mathbf{v}_n$, of band n at frequency ω_n is given by

$$\begin{aligned} \mathbf{v}_n(\mathbf{k}) &= \nabla_{\mathbf{k}}\omega_n && (3.39) \\ &= \frac{\partial\omega_n}{\partial k_x}\hat{\mathbf{x}} + \frac{\partial\omega_n}{\partial k_y}\hat{\mathbf{y}}\frac{\partial\omega_n}{\partial k_z}\hat{\mathbf{z}} && (3.40) \end{aligned}$$

Group velocity describes the rate of energy transport through the medium. Group-velocity dispersion, D_w, also known as waveguide dispersion, is proportional to the partial derivative of group velocity with respect to frequency:

$$D_w \frac{\partial v_g}{\partial \omega} \tag{3.41}$$

Group-velocity dispersion describes the rate at which a pulse's energy spreads out while propagating. As can be inferred from figure 3.19, the guided modes near the Brillouin zone edge in a hole slab have group velocities which are small in value, propagate backward, and are highly dispersive. At the edge, group velocity is zero, implying that the modes do not propagate. Modes with low group velocity, as seen in the PCWs presented here, are known as **slow light**. These kinds of modes can be understood to coherently scatter backward nearly as much as they propagate forward.

Slow light can be used to enhance material effects in the overlapping dielectric. A wave front that experiences a reduction in group velocity upon entering a PCW has its electromagnetic energy compressed into a smaller region of space. The rate of certain light-matter interactions may be expected to scale with the degree of this compression. This *slow-light enhancement* is quantified by the **slowdown factor**, S, which is given by [40]

$$S = \frac{v_\phi}{v_g} \tag{3.42}$$

where, respectively, v_ϕ and v_g are the phase and group velocities of light. It can be shown that the slowdown factor and intensity are proportional to each other. Assuming that no light is lost to scattering when entering the PCW and that the mode shape does not change, the same amount of energy is concentrated in a smaller volume. Let a pulse be described by a Gaussian curve:

$$I(x) = I_o e^{(-Sx)^2} \tag{3.43}$$

I is intensity, I_o is peak intensity, and x is a spatial coordinate normalised to the size of the pulse. The full-width half-maximum of this pulse is inversely proportional to S; a larger slowdown factor results in a shorter pulse. Since the amount of energy in the pulse does not change and energy must be conserved, the integral of (3.43) over the spatial domain is constant. Assuming that the pulse's magnitude is normalised to a value of unity,

$$\int_{-\infty}^{\infty} I_o e^{(-Sx)^2}\,\mathrm{d}x = 1 \tag{3.44}$$

$$I_o \frac{\sqrt{\pi}}{S} = 1 \tag{3.45}$$

Thus, peak intensity scales linearly with the slowdown factor. Light-matter interactions whose rates are proportional to intensity, including electron-hole pair generation in semiconductors, are expected to scale linearly with the slowdown factor, in turn.

Naturally, enhancing light-matter interactions by slowing light has its limitations. Ideally, one wishes to slow light of all frequencies equally well. In general, this is not possible for a PCW. Modes with low group velocities tend to be highly dispersive in a photonic crystal, especially near the Brillouin zone edge, where group velocity approaches zero. In dispersive regions of the crystal's band structure, each frequency component of the pulse experiences a unique group velocity, so the pulse broadens. This broadening can outweigh any benefit of slow-light compression. It is therefore important to design a PCW that exhibits slow light in segments of the band structure with low second- and higher-order dispersion. This may be done, for example, by changing the radius of the rows of holes nearest the line defect [37], or by shifting the position of these holes laterally to the line defect's axis [39]. Doing so may flatten and straighten the guided-mode dispersion curve around what is known as an *anti-crossing point* – which lies at a distance from the Brillouin zone edge – resulting in low-group-velocity, low-dispersion (LVLD) light. The key to controlling dispersion, therefore, is to move the operating point of the PCW away from the Brillouin zone edge.

Nevertheless, the dispersion curve can be straightened for only a finite range of frequencies, so bandwidth is invariably limited. Symmetry constrains the number of non-redundant wave vectors to the Brillouin zone. Consequently, the maximum allowable bandwidth is inversely proportional to the slowdown factor. This is shown as follows. First, the slope of a mode in the band structure is given by

$$\begin{aligned}\left(\frac{\Delta\omega a}{2\pi c}\right)\left(\frac{2\pi}{\Delta ka}\right) &= \left(\frac{\Delta\omega}{\Delta k}\right)\frac{1}{c} \\ &= \frac{v_g}{c} = \frac{1}{n_g}\end{aligned} \tag{3.46}$$

where n_g $(= c/v_g)$ is the group refractive index. If this index is constant throughout the entire Brillouin zone, then the mode spans a maximum range of frequencies given by

$$\begin{aligned}\frac{\Delta\omega a}{2\pi c} &= \frac{1}{n_g}\left(\frac{1}{2}\right) \\ \Delta\omega &= \frac{2\pi c}{2n_g a} \\ \Delta\nu &= \frac{c}{2n_g a}\end{aligned} \tag{3.47}$$

where $\omega = 2\pi\nu$. Provided that the phase index of refraction is constant across this bandwidth, n_g scales linearly with the slowdown factor, so maximum bandwidth is inversely proportional to the slowdown factor, as stated. In practise, usable bandwidth is only a fraction of this maximum due to dispersion and the presence of a light cone.

A second limitation with slowing light arises from the assumption of an unchanged mode shape. In a PCW, this assumption generally does not hold. At low group velocities, a mode tends to spread out in-plane and, thereby, distribute its energy over a wider area. This spreading reduces peak intensity, leading to the conclusion that slow-light enhancement of linear phenomena, such as electron-hole pair generation, is proportional to S^P, where $0 < P < 1$.

Finally, slow modes in PCWs are highly susceptible to loss. To begin with, exciting a slow mode is non-trivial, as the change in the mode's shape as it is coupled into the PCW leads to scattering loss associated with modal mismatch. Then, once the slow mode is excited, the mode suffers from propagation loss due to material absorption and fabrication disorder. While absorption and disorder plague modes of every sort, slow modes are especially susceptible, as absorption scales with the inverse of group velocity and disorder-scattering scales with the inverse-square. Some degree of loss is inevitable, but loss can be minimised by, for example, making transitions into the slow mode as gradual as possible (ie. *adiabatic*) and carefully fabricating PCWs to avoid non-uniformity of the unit cells . Otherwise, any benefit of slow-light compression might be lost.

3.6 Group Velocity Measurement

Experimentally relating the slow-light effect with light-matter enhancement requires direct measurement of group refractive index of waveguide modes. Approaches to this measurement include resolving Fabry-Perot (F-P) fringes contained within the waveguide's transmission spectrum [36], time-of-flight (ToF) measurement [57], and modulation phase-shifting [38, 58]. Each approach has its own limitations. For example, F-P resolution requires the waveguide under investigation to have highly reflective facets from which fringe spacing can be extracted, and extraction can be done only for as many points as there are fringes. As another example, ToF measurement is limited by heavy spectral dispersion characteristic of PCWs, requiring the pulse's centre of mass to be estimated when calculating time-of-flight. The error associated with this estimation is magnified if spurious reflections from reflective boundaries are present, and they often are due to the problem of modal-mismatch near the cutoff wavelength – that is, the wavelength of the mode at the Brillouin zone edge). Here, Fourier transform spectral interferometry, as described in [59], is used. This approach simultaneously allows for continuous mapping of group index with wavelength, accurate measurements in highly dispersive regions of the mode's spectrum, and, above all, robust, rapid acquisition of data.

At the heart of Fourier transform spectral interferometry is a Mach-Zehnder interferometer (MZI). In one arm of the MZI is the sample under investigation; the other is vacant. Let the former be denoted as the *sample* arm and the latter as the *reference* arm. The frequency-dependent wavefunctions associated with these two arms are

$$\psi_s(\omega) = A(\omega)e^{i(\omega t+\phi_s(\omega))} \tag{3.48}$$

$$\psi_r(\omega) = B(\omega)e^{i(\omega(t+\tau)+\phi_r(\omega))} \tag{3.49}$$

where $\psi_s(\omega)$ and $\psi_r(\omega)$ are the wavefunctions of the sample arm and reference arm, respectively. The inclusion of time constant τ in (3.49) accounts for a discrepancy in optical path length between the two arms. The reference arm is made to be of adjustable length, allowing for arbitrary control of this constant's value. In this sense, the reference arm can be thought of as a delay line. $\phi_s(\omega)$ and $\phi_r(\omega)$ are phase terms of the sample arm and reference arm, respectively. While the numeric values of these phase terms are inconsequential, their difference is not, as their difference accounts for any dissimilar shifting of phase between the two arms.

Coupling light into each arm and combining each arm's output yields interference fringes from which the sample's group index can be derived. Assuming that the light is coherent, the interference pattern, $I(\omega)$, of (3.48) and (3.49) is

$$\begin{aligned} I(\omega) &= |\psi_s(\omega) + \psi_r(\omega)|^2 \\ &= \left(\psi_s^*(\omega) + \psi_r^*(\omega)\right)\left(\psi_s(\omega) + \psi_r(\omega)\right) \\ &= \psi_s^*(\omega)\psi_s(\omega) + \psi_r^*(\omega)\psi_r(\omega) + \psi_s^*(\omega)\psi_r(\omega) + \psi_r^*(\omega)\psi_s(\omega) \end{aligned}$$

Substituting for wavefunction,

$$\begin{aligned} I(\omega) = {} & A^*(\omega)A(\omega) + B^*(\omega)B(\omega) + A^*(\omega)B(\omega)e^{i(\omega\tau - (\phi_s(\omega) - \phi_r(\omega)))} \\ & + A(\omega)B^*(\omega)e^{-i(\omega\tau - (\phi_s(\omega) - \phi_r(\omega))} \end{aligned} \tag{3.50}$$

The first two terms are the spectral densities of the sample and reference arms, respectively. The remaining two terms are the interference component and its complex conjugate. For brevity, let

$$S(\omega) = A^*(\omega)A(\omega) \tag{3.51}$$

$$R(\omega) = B^*(\omega)B(\omega) \tag{3.52}$$

$$\Phi(\omega) = \phi_s(\omega) - \phi_r(\omega) \tag{3.53}$$

In addition, assume that $A(\omega)$ and $B(\omega)$ are real over the frequency range of interest. Then, (3.50) may be condensed to read as

$$I(\omega) = S(\omega) + R(\omega) + \sqrt{S(\omega)R(\omega)}e^{i(\Phi(\omega) - \omega\tau)} + \text{c.c.} \tag{3.54}$$

where c.c. denotes complex conjugate.

Information about the sample's group index is contained entirely within the interference component of $I(\omega)$. This term describes interference fringes whose period is inversely proportional to the optical path difference between the two arms. If τ is chosen such that the reference arm is of shorter length than the sample arm for all frequencies, then the slowing of light within the sample will lead to a monotonic *increase* of optical path difference and *decrease* of fringe spacing. One may then immediately calculate group index from these fringes. However, such calculation carries a large amount of uncertainty from intrinsic Fabry-Perot modulation and can only be done for a limited number of points. If, instead, a Fourier transform of $I(\omega)$ is taken, the phase term $\Phi(\omega) - \omega\tau$ can be extracted, allowing for analysis of group index across all frequencies.

Analysis begins with measurement of $S(\omega)$, $R(\omega)$, and $I(\omega)$ using a spectrometer, then subtraction of $S(\omega)$ and $R(\omega)$ from $I(\omega)$, leaving

$$I'(\omega) = \sqrt{S(\omega)R(\omega)}e^{i(\Phi(\omega)-\omega\tau)} + \text{c.c.} \tag{3.55}$$

Taking the Fourier transform of (3.55) yields two peaks at $t = \pm\tau$. Explicitly,

$$\mathcal{F}\left(I(\omega)\right) = \mathrm{F}(t-\tau) + \mathrm{F}(t+\tau) \tag{3.56}$$

where

$$\mathrm{F}(t\pm\tau) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \sqrt{S(\omega)R(\omega)}e^{i(\Phi(\omega)-\omega(t\pm\tau))}\,\mathrm{d}\omega \tag{3.57}$$

If τ is large enough, then one of these peaks may be numerically filtered and transformed back into the frequency domain. Had $S(\omega)$ and $R(\omega)$ not been subtracted from $I(\omega)$, then a third peak centred at $t = 0$ would be present, as well. Its location at $t = 0$ is indicative of its non-interfering nature. Thus, by subtracting $S(\omega)$ and $R(\omega)$, numerical filtering of one of the interference peaks is optimized.

The filtering and inverse transform of (3.56) yields (3.55) without its complex conjugate, or, mathematically,

$$I''(\omega) = \sqrt{S(\omega)R(\omega)}e^{i(\Phi(\omega)-\omega\tau)} \tag{3.58}$$

From here, the difference in group delay between the two arms of the MZI, τ, may be obtained as follows. First, extract $\Phi(\omega) - \omega\tau$ by taking the imaginary part of the natural logarithm of (3.58). $\Phi(\omega) - \omega\tau$ oscillates between $\pm\pi$, so it needs to be unwrapped. Once it is, differentiate it by ω to solve for τ provided that $\mathrm{d}\Phi/\mathrm{d}\omega$ is approximately zero. It should be noted that background F-P oscillations from the sample's end facets exist in the phase. This unwanted contribution is minimised by performing a running average on the phase with an averaging window of at least twice the F-P resonance period before performing differentiation. Also, note that τ is obtained for $I(\omega)$'s entire range of frequencies and that its value is subject to change with frequency in accordance with the sample's dispersive properties.

Calculating the sample's corresponding group index is done by solving for τ again, but this time for a separate calibration waveguide with a known group index $n_g^{(c)}$. Let the group delays of the sample and calibration waveguides be $\tau_g^{(s)}$ and $\tau_g^{(c)}$, respectively. If the sample and calibration waveguides are of identical length L, then the sample waveguide's group index, $n_g^{(s)}$, is

$$n_g^{(s)} = \frac{c(\tau_g^{(s)} - \tau_g^{(c)})}{L} + n_g^{(c)} \tag{3.59}$$

Expressed thus, the contribution of every optical element in the MZI except for the sample of length L is subtracted. Hence, the sample's group index is accurately singled out. This subtraction has the added benefit of eliminating noise that is common to both sample and calibration measurements, such as reflections within the reference arm.

Applying Fourier analysis avoids the need for determining the position of fringe extrema or nonlinear filtering. Furthermore, if a broadband source is used, characterisation can be performed over a large spectral range in a single shot, eliminating the need for time-consuming wavelength scans. However, this approach is not without its limitations. First and foremost, the spectrometer's resolution places an upper limit on the maximum difference in group delay that can be measured. As a consequence, the range of group indices that can be resolved at once is limited. This range is inversely proportional to the waveguide's length. Second, random phase fluctuations along the optical path of both arms introduce experimental uncertainty in group delay. Restricting each arm of exposure to mechanical stress or environmental turbulence helps reduce this error. The phase-smoothing step helps in reducing this error, as well. Third, calculating group index in lossy samples proves difficult in practise, because $S(\omega)$ is too small to create an interference pattern with discernible fringes. Nevertheless, the speed and repeatability with which this approach measures group index makes it useful for quantifying the slow-light effect in PCWs.

Chapter 4

Design and Fabrication

Two types of devices are considered for investigation of slow light's enhancement of light-matter interaction. The first is a waveguide attenuator, and the second is a sub-bandgap photodetector. This chapter documents the design, parameter space that was explored for the purpose of characterisation, and fabrication of these devices.

4.1 Waveguide Attenuators

The waveguide attenuators are a series of photonic crystal waveguides (PCWs) whose purpose is to correlate with group refractive index the portion of transmission loss associated with the presence of carrier-generation centres, or defects, within their guiding region. The waveguides are designed to exhibit a different range of indices within the C band, allowing for correlation to be performed at specific wavelengths. This is done to eliminate from comparisons between waveguides any possible material dependence – of the generation centres or otherwise – on wavelength. Silicon divacancies, which are to serve as the defects, are introduced into the guiding region via silicon ion implantation. Section 4.1.1 describes the design of these waveguides, including the technique used to subtly adjust their guided-mode properties, and section 4.1.2 describes their fabrication.

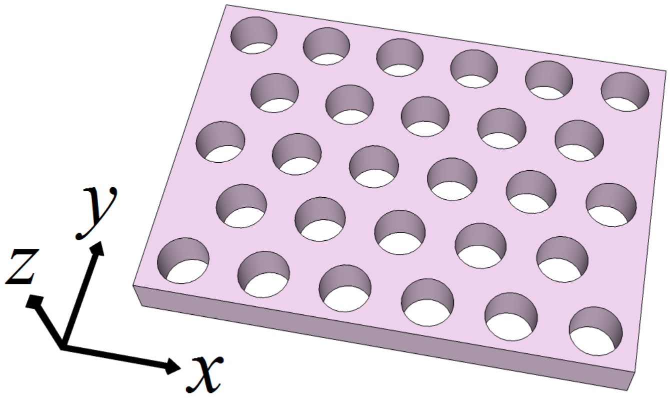


Figure 4.1: A photonic crystal slab, which combines two-dimensional periodicity in the xy-directions and index-guiding in the z-direction. The oxide substrate and superstrate have been removed for visual purposes.

4.1.1 Design

Design begins with a featureless silicon-on-insulator (SOI) photonic crystal slab, such as the one depicted in figure 4.1. The slab consists of a periodic triangular lattice of holes in silicon of 220nm thickness. The thickness is chosen thus to ensure single-mode propagation in the waveguide that is to follow and, thereby, prevent modal dispersion from confounding comparative analysis. To demonstrate CMOS compatibility, the slab is clad in silica, consisting of a buried-oxide (BOx) layer below and a deposited oxide above and within the holes. (Note that neither silica layer is shown in figure 4.1.) The slab's dielectric function, namely its hole radius r and pitch a, is adjusted so that the slab exhibits a photonic band gap (PBG) that includes C band frequencies. This is done by first determining the effective index of the slab using the *effective index method*. With the assistance of Lumerical™ FDTD Solutions, the effective index within the plane of the silicon slab for TE modes with 1550 nm wavelength is determined to be 2.85. This value is then used to calculate the band diagram of two-dimensional photonic crystals with variable hole-radius/pitch ratios using the MIT Photonic-Bands simulation package.

Two dielectric functions with appropriate PBGs were settled upon, one with $r = 0.312a$ and the other with $r = 0.269a$. Their band diagrams are depicted in figure 4.2.

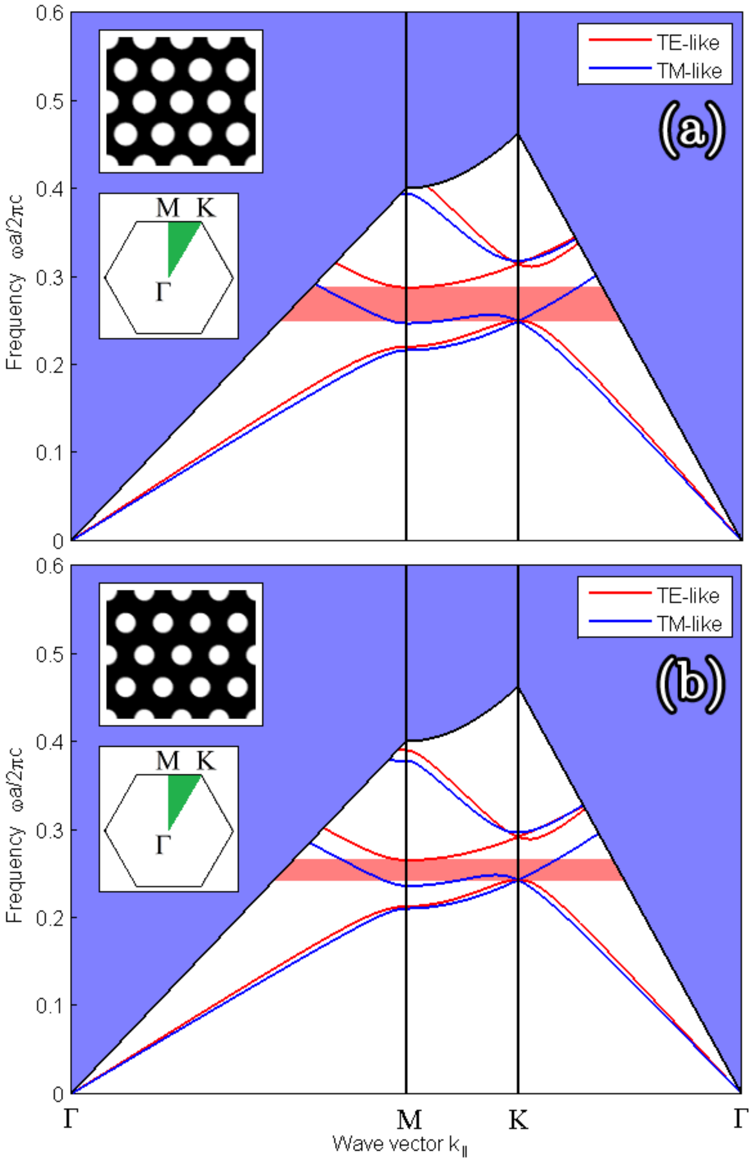


Figure 4.2: Band diagrams of photonic-crystal slabs with (a) $r = 0.312a$ and (b) $r = 0.269a$. r is hole radius, a is hole pitch, ω is angular frequency, and $k_{||}$ is the in-plane k-vector. A gap for TE modes (shaded in red) exists in both slabs. The blue shaded area is a light cone owing to the presence of oxide cladding ($n = 1.44$). The top inset provides a cross-sectional view of each crystal's dielectric function. The bottom inset shows the high-symmetry points Γ, M, and K at the corners of the irreducible Brillouin zone.

The regions shaded in red denote PBGs for TE modes. The first function has a 13.8% gap between normalised frequencies 0.250 and 0.287. The second has a 8.7% gap between frequencies 0.243 and 0.265. In conformity with scale invariance of Maxwell's equations, a is freely chosen to be 385 nm for the first function, giving a hole radius of 120 nm and a gap between 1341 nm and 1540 nm. Likewise, a is chosen to be 390 nm for the second function, giving a hole radius of 105 nm and a gap between 1471 nm and 1605 nm. Note that neither function has a prominent TM gap, as typically expected for a lattice of holes.

Of special note is the light cone, which is shaded in purple in both diagrams. Modes which lie within this cone extend outside of the photonic crystal slab. If modes of these k-vector and frequency combinations are excited, they will radiate out-of-plane, as they are not confined by total internal reflection. Waveguiding cannot be achieved for such combinations. Similarly, waveguiding cannot be achieved in directions along which the k-vector is not conserved – that is, directions along which discrete translational symmetry is broken – since modes will couple into this light cone. The light cone's slope is inversely proportional to the cladding's refractive index. Consideration of this fact reveals that the cladding is a limiting factor in the number of modes available for waveguiding. This limitation is in full agreement with the fundamentals of total internal reflection.

Of course, it is the modes within the gap, and not within the photonic crystal itself, that are to be considered for waveguiding. Modes can be made to appear within the gap by removing a row of holes along the x-direction, or introducing a *line-defect* as it is otherwise known. The increase in effective index brought upon by this removal serves to *pull down* high-frequency modes into the gap. Such modes may be confined in the y-direction by the gap. The portion of the mode that falls below the light line is also confined in the z-direction via total internal reflection, resulting in two-dimensional confinement as required for low-loss waveguiding. It should be noted that by removing a row of holes in this manner, discrete translational symmetry in the y-direction is broken. This means that the y-component of the k-vector, k_y, is no longer conserved, resulting in unrestricted coupling of modes along this direction. Thus, when calculating the band diagram, for a given x-component of the k-vector, k_x, which is still conserved, and frequency ω, modes for all k_y must now be considered. To wit, if there is some k_y for a given (k_x, ω) that puts the guided mode on a band, that mode will *not* be confined by the gap and will leak into the crystal. The process of finding extended modes in the crystal for all k_y is called *projecting the band structure*. Projected band structures for two waveguides are shown in figure 4.3.

The slope of the guided modes pulled down into the gap, and thus these modes' group refractive index, can be controlled by perturbing the dielectric function around the line defect. For low-group-velocity low-dispersion (LVLD) propagation, this slope should be made both flat (for low group velocity) and straight (for low modal dispersion). There are several ways to achieve this quality, such as by decreasing the radius of the nearby holes [61] or by laterally shifting the rows of holes nearest and next-nearest to the line defect [39]. The latter approach is taken here for its reproducibility and relative ease of fabrication. In general, the first-row shift, s_1, influences the slope of the mode far from the Brillouin zone edge, where the group velocity is relatively large, while the second-row shift, s_2, influences the slope near the zone edge, where the group velocity approaches zero. This can be deduced from the field patterns shown in figures 4.4a and 4.4b. Far from the zone edge, the mode is contained mostly within the line defect, so the mode is largely under the influence of s_1 alone. However, as the mode's group velocity slows down near the band edge, the mode penetrates deeper into the crystal, making s_2 significant as well. This practise of adjusting s_1 and s_2 to tailor the guided mode's dispersive properties, in what may be called *dispersion-curve engineering*, is important for achieving LVLD propagation necessary to enhance light-matter interaction. In the present work, four combinations of row shifts are chosen for the first dielectric function and one for the second, for a total of five unique waveguides. These waveguides are summarised in table 4.1, and their band diagrams are shown in figure 4.5.

Three features of these band diagrams are worth mentioning. First, judging from the slope of the guided modes' dispersion curves, the group velocity of the guided mode within each waveguide is expected to gradually decrease with decreasing frequency, starting at the light line, until it reaches the Brillouin zone edge, at which point velocity is theoretically zero. Propagation is expected to be lossy at frequencies above

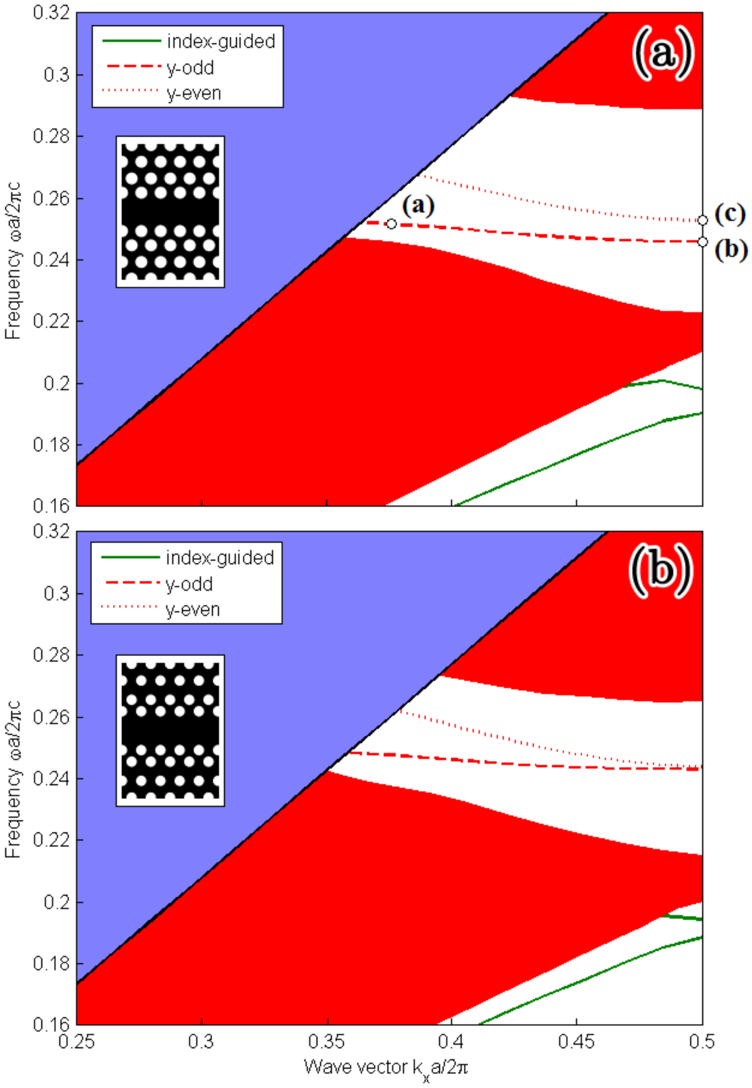
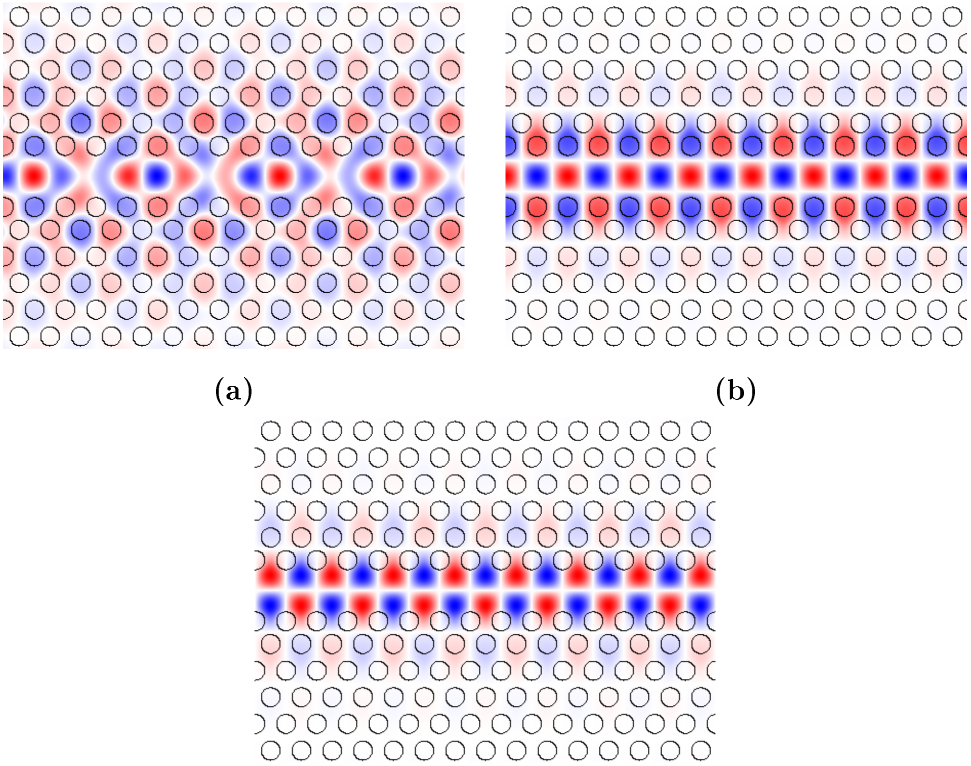


Figure 4.3: Projected band structure of TE-like modes for photonic-crystal waveguides with (a) r = 120 nm and a = 385 nm, and (b) r = 105 nm and a = 390 nm. In (a), the row of holes nearest the line-defect have been laterally shifted away by 50 nm. Similarly, in (b), the nearest row has been shifted away by 55 nm and the next-nearest row has been shifted toward by 25 nm. Gap-guided modes are classified as y-odd (dashed line) or y-even (dotted line). Red-shaded regions denote extended TE modes in the crystal. The blue-shaded region is the light cone. Also present are index-guided modes below the extended region. Field patterns corresponding to labelled points (a-c) are shown in figure 4.4.



(a)

(b)

(c)

Figure 4.4: H_z field patterns for photonic-crystal waveguides at the $z = 0$ mirror plane, with (a-c) corresponding to the labelled points in figure 4.3a. Because H_z is a pseudovector, modes that appear y-odd are actually y-even, and vice versa.

Waveguide	Radius r (nm)	Pitch a (nm)	1st Row Shift s_1 (nm)	2nd Row Shift s_2 (nm)
1	120	385	50	0
2			40	0
3			30	0
4			40	-10
5	105	390	55	-25

Table 4.1: Summary of the photonic-crystal waveguides featured in this thesis for displaying slow-light propagation. Positive and negative shifts indicate lateral displacement away from and toward the line-defect, respectively.

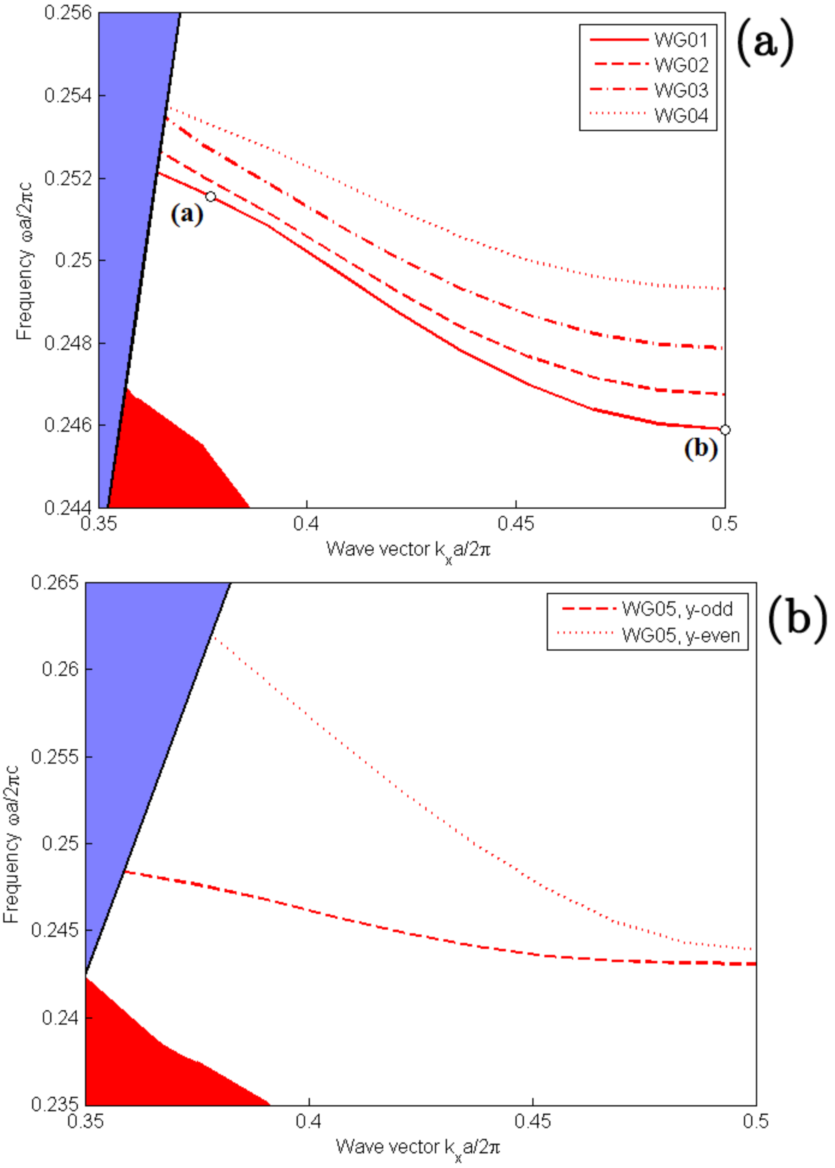


Figure 4.5: Band structure for the photonic-crystal waveguides listed in table 4.1. The modes in (a) are entirely y-odd. y-even modes are omitted because they lie approximately 50 nm below the spectral region of interest. In (b), both y-even and y-odd modes are present. Field patterns corresponding to labelled points (a-b) are shown in figure 4.4. The blue-shaded region is the light cone.

the light-line and completely prohibited at frequencies below the guided mode's end. The frequency range between these two points is the guided mode's bandwidth. While each bandwidth spans across a different range, there is a common overlap between roughly 1527 nm and 1546 nm wavelengths. Within this overlap, for waveguides 1-4, the guided mode's group velocity is expected to change dramatically. For waveguide 5, on the other hand, group velocity is expected to be high and constant. This is due to the fact that its guided mode's slope is comparatively larger and straighter than the others. It should be noted that all waveguides have both y-even and y-odd guided modes within their respective gaps. For waveguide 5, these two modes are closely spaced in frequency, whereas for the other waveguides the y-odd mode lies outside the spectral region of interest and is thus omitted. Nevertheless, it is waveguide 5's y-odd mode that shares a bandwidth overlap with the other waveguides, so it is this mode alone that is of value for comparative analysis.

Second, the family of curves in figure 4.5a demonstrates how increasing the group index of a guided mode invariably incurs a cost in the mode's bandwidth. In other words, waveguide design is constrained by a finite index-bandwidth product. Furthermore, the light-line subtracts from this product, since it places an upper limit on the frequencies available for lossless guiding. In light of this fact, it is prudent to use cladding with as low a refractive index as possible. Evidently, the silica cladding used here limits the usable bandwidth.

Third, it must be reaffirmed that these band diagrams apply exclusively to TE modes. There are no guided modes for TM-polarised light because there is no TM gap. Should the cladding above and below the photonic-crystal slab have different refractive indices, $z = 0$ mirror symmetry will be broken, leading to the mixing of TE and TM polarisations [62]. In effect, this mixing will cause even modes which are originally TE-polarised to radiate in-plane. This is a large part of why the waveguides here are designed to have symmetric oxide cladding.

The waveguide attenuators consist of the as-discussed PCWs implanted with Silicon ions and coupled on either end to wire-waveguides leading to fiber-grating couplers. The grating couplers have a pitch of 630nm and are shallow-etched 70nm into the silicon layer with a 50% duty cycle. These dimensions allow for efficient coupling of 1550nm TE-polarised light into the wire-waveguides from external fiber-optic cables held above the grating at 12 °from normal incidence. The wire-waveguides adiabatically taper this grating-coupled light into the PCWs, from 10μm width at the coupler to 0.5μm width at the PCW, across approximately 1800μm in length. In order to avoid prohibitively large modal-mismatch coupling loss between the fast-moving wire-waveguide mode and slow-moving crystal mode, the pitch at either end of the crystal is elongated along the x-direction in four stages of 10nm increments, with each stage being two periods in length. The progressive increase in the guided mode's group velocity through these elongated sections serves to further adiabatically couple light into the PCW. In order to robustly characterise waveguide response, photonic crystal length l is varied from 80μm to 320μm for all waveguides. Silicon ions are implanted along the entire length of the photonic crystal guiding-regions for all lengths to allow defect-mediated absorption to be quantified per unit length. Also, following the discussion in section 3.6, featureless wire-waveguides of 0.5μm width and matching lengths are included to enable measurement of group velocity. Using RSoft™ FDTD simulation software, the effective group refractive index of the fundamental TE-polarised mode in these waveguides is determined to be 2.3 for a wavelength of 1550nm. A schematic of the device layout is illustrated in figure 4.6.

4.1.2 Fabrication

IMEC's ePIXfab was commissioned to fabricate the devices using their in-house SiPhotonics technology. The devices were fabricated from a SOI wafer consisting of a 220nm-thick silicon device layer and a 2000nm-thick buried silica layer atop a silicon substrate. Device patterns as shown in figure 4.6 were prepared using Design Workshop Technologies™ dw-2000 layout software. The layout consisted of 18 waveguides in total: five PCWs as shown in table 4.1 plus one calibration waveguide for 80μm, 160μm, and 320μm lengths. Waveguides were spaced 250μm apart. The wire-waveguide tapers were outlined with 3.5 μm-wide trenches. This layout was transferred onto a photomask using 0.13nm electron-beam lithography. This mask was then used to define the pattern onto the silicon layer using a combination of 193nm

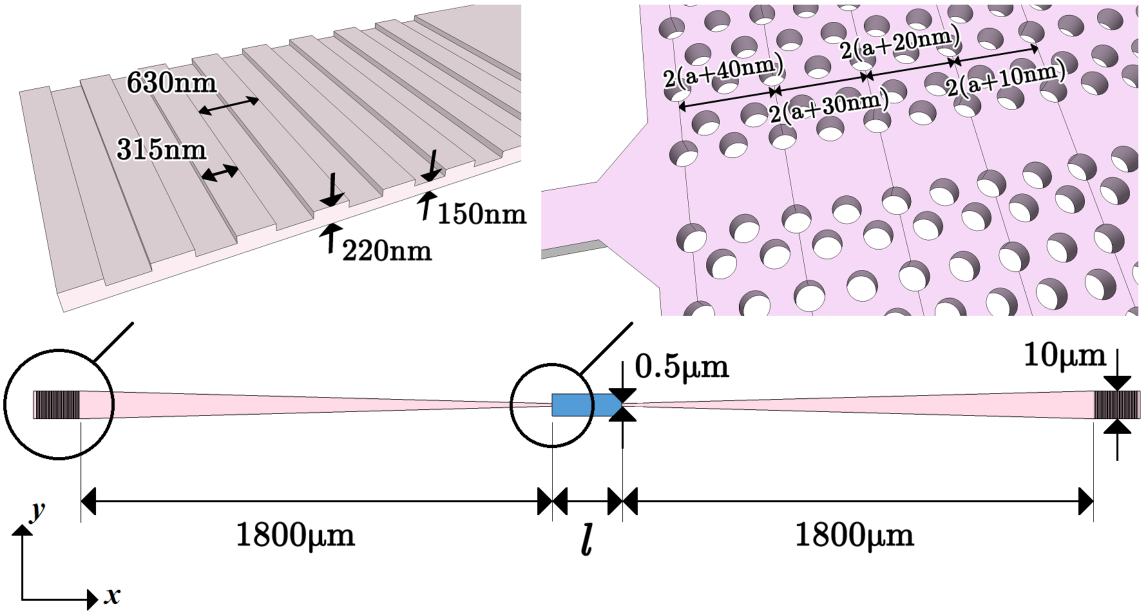


Figure 4.6: Schematic of the waveguide attenuator. *Top left*: The fiber-grating coupler used to couple TE-polarised light from an external fiber-optic cable into the wire-waveguide. *Top right*: Interface between the wire- and photonic-crystal waveguides. The crystal's pitch is elongated along the x-direction in four stages in order to adiabatically taper light into the slow-moving crystal mode. *Bottom*: The PCW (shaded in blue) of length l together with wire-waveguide tapers on either end. l is varied from 80 μm to 320 μm for each PCW listed in table 4.1.

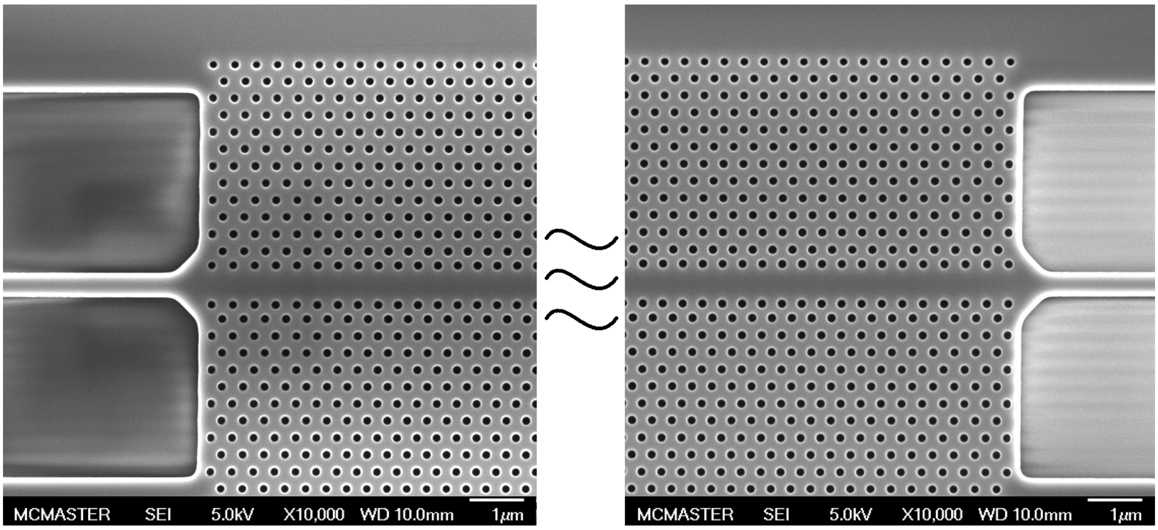


Figure 4.7: SEM images of both ends of a PCW with a hole radius and pitch of 120nm and 385nm, respectively. Row-shifts $s1$ and $s2$ are 50nm and 0nm, respectively.

deep-UV lithography and inductively-coupled-plasma reactive-ion etching (ICR-RIE). Photolithography was done in 13 positions across the length of the wafer, resulting in 13 copies of the pattern, or 13 *die*, being transferred onto the wafer. Furthermore, the exposure dose between die was varied between approximately 22mJ/cm^2 at one end of the wafer and 35mJ/cm^2 at the other. Hole radius is proportional to exposure dose, whereas hole pitch is not. This means that the photonic crystals on each die were given unique hole-radius/pitch ratios and would thus exhibit dramatically different spectral responses. This variation of exposure dose between die permitted investigation of multiple dielectric functions in a single fabrication run. Following photolithography and a subsequent etch, silicon oxide of 1.25μm thickness was deposited onto the wafer using an in-house recipe and planarised for efficient fiber-coupling, then the wafer was diced into die. SEM images of one waveguide without a top oxide are shown in figure 4.7.

After processing with IMEC, a photomask to be used for Silicon ion implantation was designed using dw-2000. The mask design was sent to University of Alberta's nanoFab where a 5-by-5-inch chrome-on-quartz mask was written via electron-beam lithography. This mask was then sent to the University of Western Ontario to be trimmed to 4 by 4 inches in size for compatibility with photolithography equipment at McMaster University. The photomask consisted of 15 $2\,\mu$m-wide implantation windows, one for each PCW, of lengths matching each PCW. Additional patterns were included to assist alignment during photolithograpy. Three select die were prepared for ion implantation by first spinning Microposit$^{\text{TM}}$ SC$^{\text{TM}}$ 1827 positive photoresist onto the die at 3000 RPM, covering the surface of the die with approximately $2.7\,\mu$m of photoresist. At this thickness, the resist was able to be used as a soft mask for ion implantation. Then, windows were opened over the PCWs by using the photomask to selectively expose the resist to UV light, then etching away the exposed resist by submerging the die in a 1:5 solution of Microposit$^{\text{TM}}$ 351-Developer:H_2O for 35 seconds. The die were then hard-baked at 130 °C for two minutes to harden the photoresist for masking. An approximate alignment error of 0.5 μm accompanied this window definition.

Ion implantation was carried out at the University of Western Ontario. The implant energy and dose were selected to be 1.8MeV and 6.0×10^{12}cm^{-2}, respectively. At 1.8MeV energy, the Silicon ions were able to penetrate the top oxide and cause damage within the silicon lattice necessary to form divacancy-type generation centres. Implantation was performed at normal incidence. Following implantation, the photoresist was stripped from the dies by soaking them in acetone overnight, rinsing them with methanol, then submerging them in a 1:4 solution of H_2O_2:H_2SO_4 at 120 °C for 10 minutes.

4.2 Sub-Bandgap Photodetectors

The sub-bandgap photodetectors are Silicon-based PCWs designed to investigate slow-light enhancement of measurable current generated via Shockley-Read-Hall carrier generation. Like the waveguide attenuators discussed in the previous section, these devices are made sensitive to photons of C-band frequencies, which lie below silicon's electronic bandgap, via the introduction of silicon-divacancy carrier-generation centres within their guiding regions. However, unlike the waveguide attenuators, they provide a first-hand account of the enhancement of useful, photo-generated current, rather than the total number of extra photons absorbed, not all of which may be assumed to yield an extractable electron-hole pair. Electron-hole pairs are extracted by means of p^+-v-n^+ diode structures integrated along the length of the PCWs. Design and fabrication was done in co-operation with the Microphotonics and Photonic Crystals Group at the University of St. Andrews. Section 4.2.1 describes the design process of these devices, including the parameters varied, and section 4.2.2 provides a detailed description of each fabrication step.

4.2.1 Design

The photodetectors consist of single-mode SOI PCWs running down the middle of p^+-v-n^+ diode structures. PCWs of two different dielectric functions are considered. Both functions consist of a triangular lattice of holes. The hole radius and pitch of the first are 110nm and 390nm, respectively, and the hole radius and pitch of the second are 115nm and 400nm, respectively. In the same style as the waveguide attenuators, these PCWs are engineered to support slow-light propagation by laterally displacing the rows of holes nearest and next-nearest to the line-defect 55nm away from the line-defect ($s_1 = 55$nm) and 25nm toward the line-defect ($s_2 = -25$nm), respectively. A top oxide is deposited atop the silicon layer. The band diagrams of these PCWs are shown in figure 4.8. Defects are introduced to the PCWs via Silicon-ion bombardment. Electron-hole pairs generated at these defects are separated by means of the intrinsic electric field between the p^+ and n^+ dopant regions. Metal contacts are deposited on top of these dopant regions through vias within the top oxide to both allow the separation to be detected as current and permit the enhancement of photo-response with reverse-bias operation. Recognizing that generation centres double as recombination centres, bombardment is limited to an area coincident with the guided-mode region (approximately two rows of holes to either side of the line-defect according to FDTD simulations) via selective masking to minimize the recombination of photo-generated carriers. Also, dopant regions and metal contacts are kept several micrometres from the guided-mode region to minimise photo-absorption caused by the presence of free carriers and MOS-capacitance effects, respectively. The device's

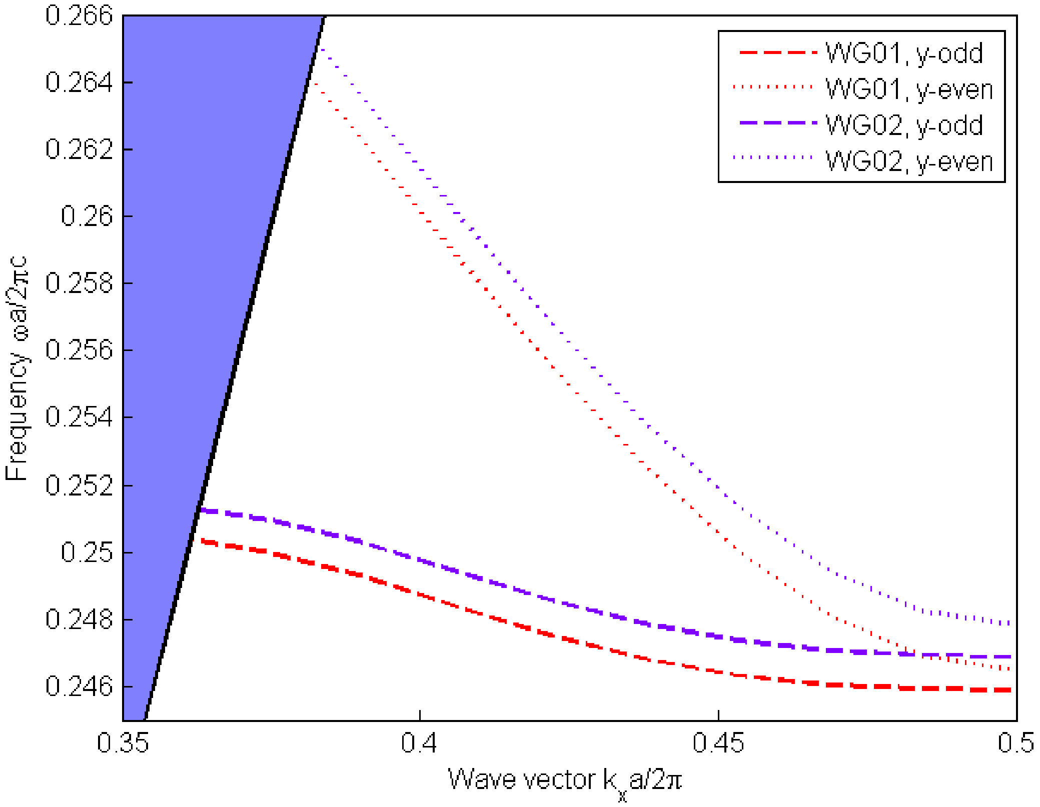


Figure 4.8: Band structure of PCWs incorporated in sub-bandgap photodetectors. WG01 corresponds to a PCW with hole radius and pitch of 110nm and 390nm, respectively. WG02 corresponds to a PCW with hole radius and pitch of 115nm and 400nm, respectively. The purple-shaded region is the light cone.

structure is partially illustrated in figure 4.9.

As with the waveguide attenuators, a photomask is fabricated with the help of University of Alberta's nanoFab and the University of Western Ontario to assist device fabrication. The photomask consists of six 2cm × 1.5cm tiles, with each corresponding to a unique photolithography step. Parameter space is explored by varying the separation of the dopant regions, x, between 2μm and 8μm and device length, l, between 20μm and 480μm for each of the two PCWs considered. In total, each tile of the photomask accommodates 24 waveguides, giving 24 unique parameter combinations. These waveguides are spaced 200μm apart. This photomask is used

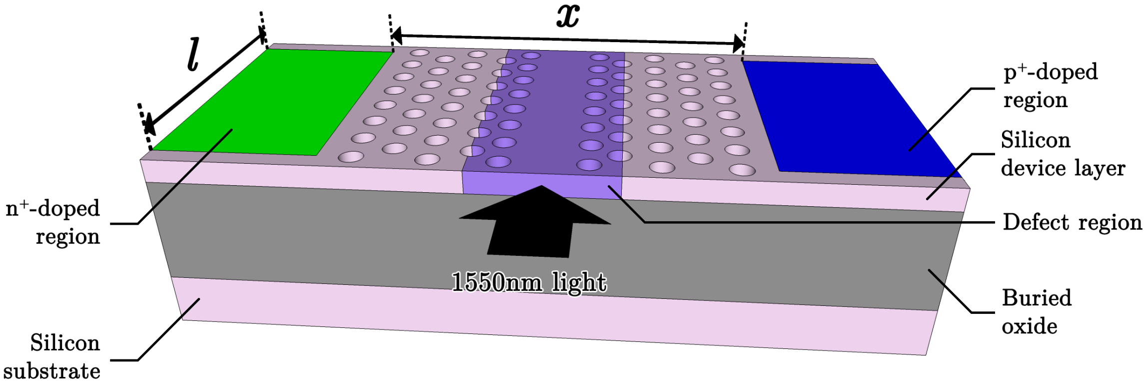


Figure 4.9: Illustration of CMOS-compatible PCW photodetector sensitive to light of 1550nm wavelength. Junction separation x and length l are varied between 2-8μm and 20-480μm, respectively. The top oxide and metal contacts are omitted for visual purposes.

in every processing step except for the definition of the PCWs, which is done via electron-beam (e-beam) lithography at the University of St. Andrews. A detailed description of each of these processing steps is provided in the next section.

4.2.2 Fabrication

In order to ensure that the photodetectors are cost-effective and readily integrable with silicon microelectronics, every fabrication step is compatible with CMOS fabrication techniques. Accordingly, every piece of equipment used in this work can be found in a modern microelectronics IC fabrication facility. A summary of these fabrication steps is shown in figure 4.10.

The detectors were fabricated on 2.5cm × 2.5cm SOI chips cleaved from a SOITEC wafer consisting of a 220nm-thick silicon device layer and a 2000nm-thick buried silica layer atop a silicon substrate (a). Alignment marks were defined on the chips via photolithography and transferred to the device layer using RIE with SF_6 gas at 120W power and 30mTorr operating pressure. These alignment marks were used to assist the alignment of features during subsequent photolithography.

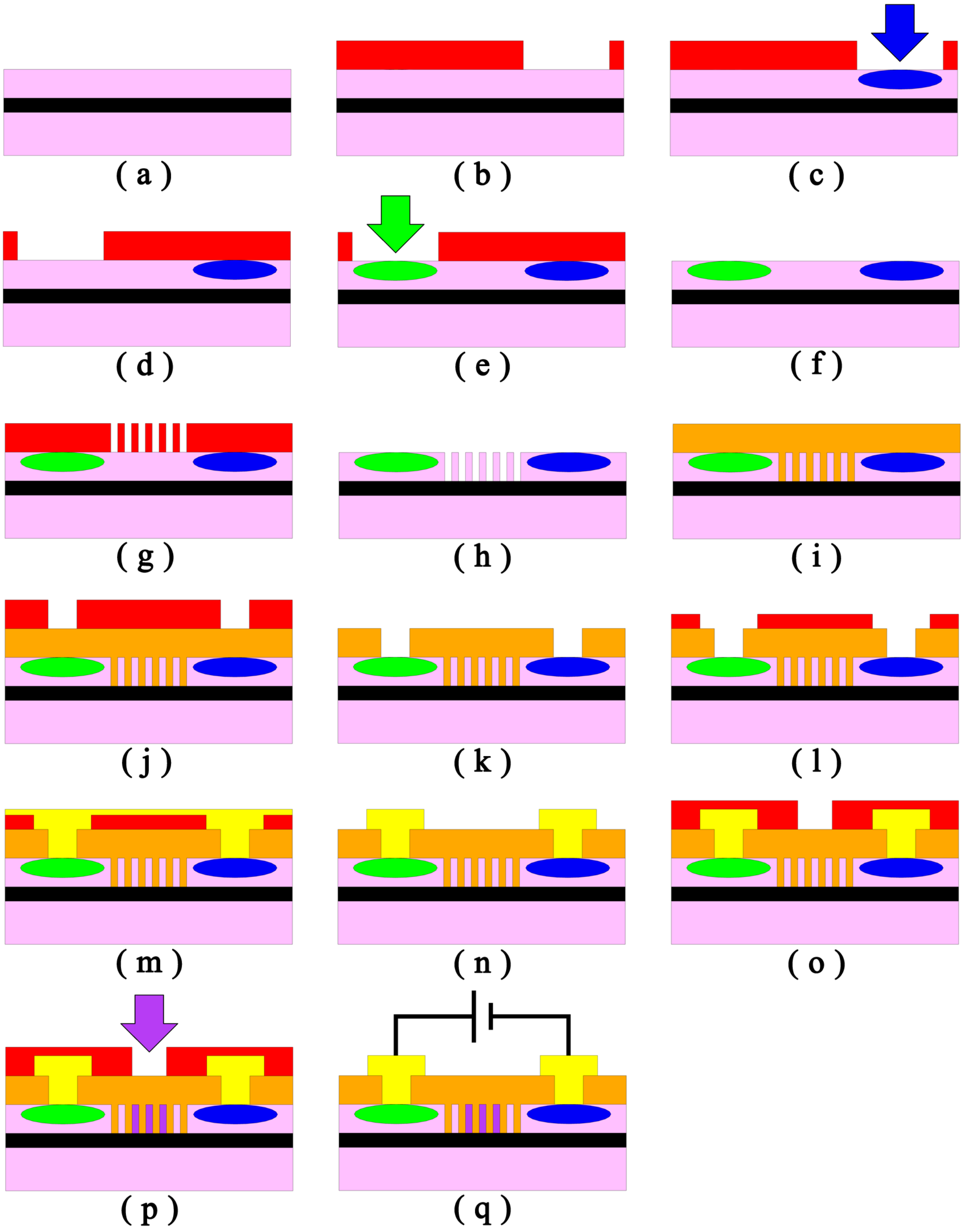


Figure 4.10: CMOS-compatible process flow for sub-bandgap PCW photodetectors. Steps (a-q) are described in the text.

P^+ and N^+ Doping

Fabrication proper began with masked phosphorus and boron ion implantation to form n-type and p-type dopant regions, respectively. Prior to implantation, Microposit™ S1808 photoresist was spun onto the chips, covering the surface in approximately 0.8μm of photoresist. Implantation windows of 30μm width and variable length l were defined in this photoresist via photolithography followed by removal of exposed resist using 351-Developer solution (b). The photoresist was then hardened by baking the chips at 130 °C for 2 minutes to ensure that ions did not penetrate through the photoresist that was not exposed. Boron was implanted into the silicon exposed by the windows at an energy and dose of 15keV and 1×10^{15} ions/cm^2, respectively (c). Implantation was done at an angle of 7 °from normal incidence to prevent channelling between crystallographic planes. These procedures were repeated for phosphorus implantation (d-e). However, windows for phosphorus were defined at a variable lateral separation x from those for boron, and phosphorus was implanted at an energy of 25 keV to account for the fact that it has a larger atomic mass and thus penetrates shorter into silicon at a given energy. According to TRIM simulation software, boron and phosphorus are expected to penetrate roughly 60nm and 38nm into the silicon layer, respectively, with minimal lateral straggle under their respective conditions. Photoresist was removed following each implantation using a sequence of acetone and methanol at room temperature followed by a 1:4 solution of H_2O_2:H_2SO_4 at 120 °C. A 1-2μm offset in x was observed due to alignment error during photolithography.

Following implantation, the chips were thermally annealed to reform the crystal lattice damaged by implantation and to incorporate as a species the dopant into the lattice (f). Annealing was done at 950 °C for 5 minutes in flowing nitrogen using a rapid thermal annealer. Annealing was done in nitrogen to minimise dopant diffusion and thus maintain separation x.

Waveguide Definition

Low-loss PCWs and adiabatic wire-waveguide tapers were fabricated using e-beam lithography at the University of St. Andrews. The waveguide pattern was defined in ZEP-520A e-beam resist using a hybrid Zeiss GEMINI 1550/Raith ELPHY Plus e-beam writer with a 200μm write field (g). The resist was used as a mask to transfer the pattern into the device layer using RIE with a 1:1 mixture of SF_6 and CHF_3 gases at 20W power (h). This etching exhibited low selectivity between silicon and resist, which necessitated the use of a relatively thick (350nm) resist layer. On the other hand, this low selectivity also minimised side-wall roughness that conventionally accompanies micro-masking, which allowed for the fabrication of waveguides with low scattering loss.

The wire-waveguide tapers extended from 0.8μm width at the PCW to 3μm across 200μm of length. Beyond each taper was a 3μm-wide wire-waveguide of a length on the order of millimeters. At this width, tapered fiber was able to be used to couple light into the waveguide. All wire-waveguides were surrounded by 3μm trenches.

Of special note is the fact that the maximum PCW length considered (480μm) exceeded the e-beam writer's write-field size (200μm). PCWs of lengths exceeding the write-field size had to be written across more than one field, which introduced stitching errors at the boundaries between fields. Light experiences scattering loss at these boundaries. Given that such losses scale approximately with the square of refractive index [63], stitching error is potentially severe for slow-light PCWs. In an attempt to minimise this deleterious effect, the photonic crystal's hole pitch was increased at either end of the boundaries in the same fashion as the increase at the PCW-taper interface to bring the guided mode into the fast-light regime, where scattering loss is smaller. Nevertheless, a small penalty in transmission is expected to accompany PCWs of lengths greater than 200μm.

Following the etch, Dow Corning$^{\text{TM}}$ Flowable Oxide (FOx) was spun on the chips to serve as a top oxide layer (i). FOx is an inorganic polymer consisting of Si-H and Si-O bonds. The chips were then cured to disassociate Si-H bonds to form Si-O ones and, thereby, turn the polymer into an SiO_2-like oxide. The chips were cured at an initial temperature of 100 °C to drive out solvent in the FOx. This temperature was then gradually increased to 400 °C, where it remained for 2 hours to disassociate Si-H bonds into Si-O bonds and, thereby, turn the polymer into an SiO_2-like oxide. The resulting oxide thickness was determined to be approximately 450nm by selectively etching test samples prepared under identical conditions, then measuring etch depth with a mechanical profiler. This oxide layer both restored z-plane mirror symmetry necessary to avoid polarisation-mixing and prepared the chips for the metallisation that was to follow.

Metal Contacts

Vias 10μm wide and running along 80% of the length of the PCWs were defined in S1808 resist via photolithography (j). The vias were positioned concentrically atop both dopant regions, 10μm from either either edge. Pattern transfer into the top oxide was done using RIE at Carleton University (k). The vias were transferred using a 5:1 mixture of CHF_3 and O_2 gases at 350W power and 150mTorr operating pressure. This mixture exhibited low selectivity between the top oxide and silicon device layers, so etching was stopped with around 20nm of the oxide remaining, then wet-etched the rest of the way with a 1:10 solution of 20%-HF:H_2O. Inspection under a microscope revealed that undercutting of the resist, which here is an undesirable side-effect of wet etching, was minimal.

Aluminium contacts were deposited atop the vias using the lift-off technique. Windows for these contacts were defined in S1808 resist (l). These windows were positioned concentrically with the vias and were 3μm wider on all sides. The windows included $100\mu\text{m}\times100\mu\text{m}$ areas (not shown in figure 4.10) to be used as contact pads for probes during measurement. To increase the chance of successful lift-off, the resist was soaked in toluene for 6 minutes before being placed in developer. The toluene hardened the surface of the resist such that, when placed in developer, the resist would exhibit undercutting. Furthermore, the resist was baked at a lower temperature than usual (90 °C as opposed to 110 °C). These precautions would facilitate the infiltration of solvent into the resist. Prior to aluminium deposition, the samples were flash-etched in a 1:50 solution of 20%-HF:H_2O to remove the build-up of native oxide on the device layer exposed through the vias and, thereby, ensure the formation of low-resistance contacts. A 300nm-thick aluminium layer was deposited onto the resist via sputtering (m). Then, lift-off was performed by soaking the samples in acetone, which removed aluminium everywhere except in regions defined by the windows (n). In order to improve electrical contact, one sample was sintered at a temperature of 350 °C for 60 seconds. Subsequent current-voltage measurements revealed that ohmic contact was achieved.

Silicon Ion Implantation

The final photolithography step defined windows centred on and running the length of the PCW line-defects (o). The windows extended two rows of holes to either side, for a total width of approximately 1.35μm and 1.39μm for the 390nm- and 400nm-PCWs, respectively. A resist thickness of 1.8μm was applied to prevent the penetration of Silicon ions to regions outside the windows during implantation.

Silicon ions were implanted at an energy and dose of 800keV and 3×10^{12} ions/cm^2, respectively, at normal incidence at the University of Western Ontario (p). According to TRIM, mean implantation depth at this energy is approximately 1μm. Thus, the Silicon ions should have penetrated through the top oxide into the device layer and caused the damage necessary to form mid-bandgap defects.

Facets

Fabrication concluded with the preparation of optical-quality waveguide facets via dicing and polishing. Brand Laser Optics was commissioned to carry out this procedure. The samples were polished down to the 3μm-wide waveguides at both ends to allow light to be coupled into the photodetectors using a tapered fiber. A completed photodetector is shown in figures 4.10q and 4.11.

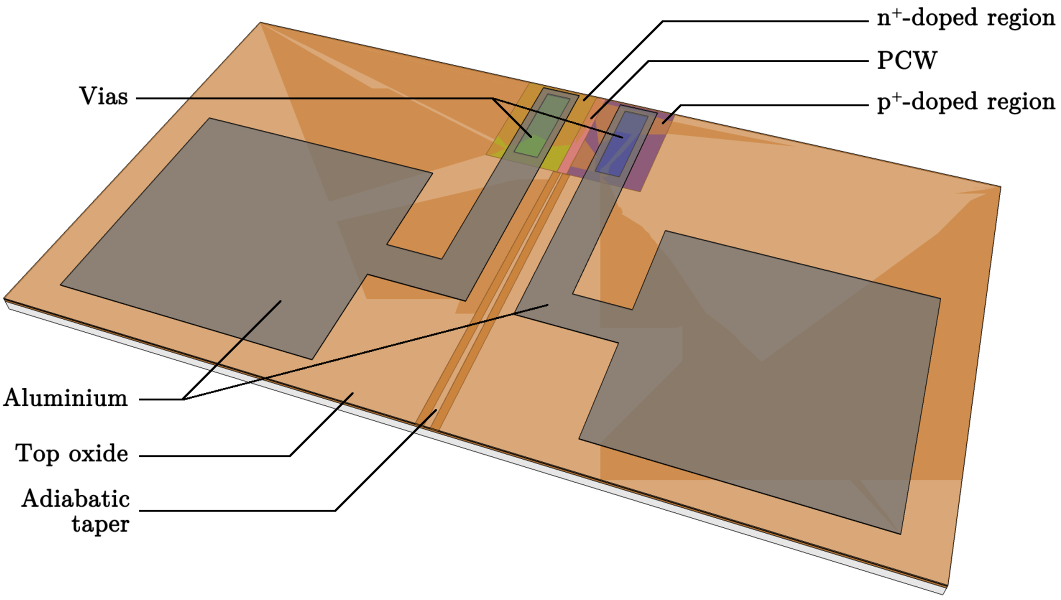


Figure 4.11: Proportionally accurate representation of a SOI sub-bandgap PCW photodetector.

Chapter 5

Experiment and Analysis

The optical response of the photonic-crystal waveguide attenuator and the electrical and optical responses of the photonic-crystal waveguide sub-bandgap photodetector have been characterised. The attenuator exhibits up to $1200\mathrm{dB}\cdot\mathrm{cm}^{-1}$ of optical absorption attributed to its implanted defects. The photodetector exhibits a responsivity of up to $34\mathrm{mA}\cdot\mathrm{W}^{-1}$ for a detector length of $80\mu\mathrm{m}$. This chapter presents the experimental setup used and analysis performed to obtain these results.

Before beginning, it must be noted that the hole radius of a photonic crystal is strongly dependent on fabrication conditions such as photoresist thickness, exposure dose, and beam-spot size (where electron-beam lithography is used). Therefore, in order to permit fair comparison between PCWs of different dielectric functions, all data are taken from PCWs belonging to the same chip unless otherwise stated. Also, while an undesired hole size might cause the characteristics of PCWs to differ from design on a quantitative basis, they are expected on first order to agree on a qualitative basis.

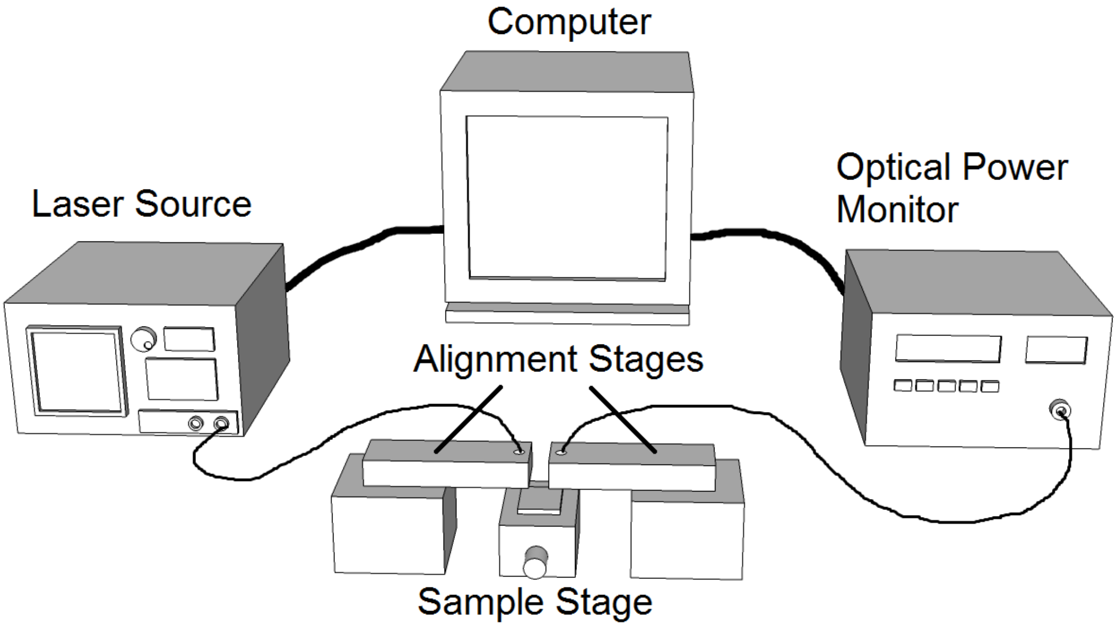


Figure 5.1: Experimental setup for measuring the optical transmission of PCW attenuators.

5.1 Waveguide Attenuators

5.1.1 Experimental Setup

Two separate experimental setups were used to characterise the waveguide attenuators. The first, illustrated in figure 5.1, was used to measure optical transmission through the attenuators. Low-loss, single-mode optical fiber was used to couple light from an Agilent 81640A laser diode (tunable across C-band frequencies) to the sample and from the sample to an ILX Lightwave FPM-8200 fiber-optic power monitor. The fiber was flat-cleaved at one end, allowing for light to be coupled to (from) the sample's on-chip fiber-grating couplers directly from (to) the fiber. z-translatable arms attached to alignment stages firmly held the fiber above these grating couplers. The alignment stages were capable of piezo-electrical xy motion, offering precision alignment for maximal coupling. The use of polarisation-maintaining fiber between the source and sample ensured that TE-polarised light was incident on the grating couplers, again for maximal coupling. Interfacing with the laser source and power monitor via GPIB cable was a personal computer, from which National Instrument's

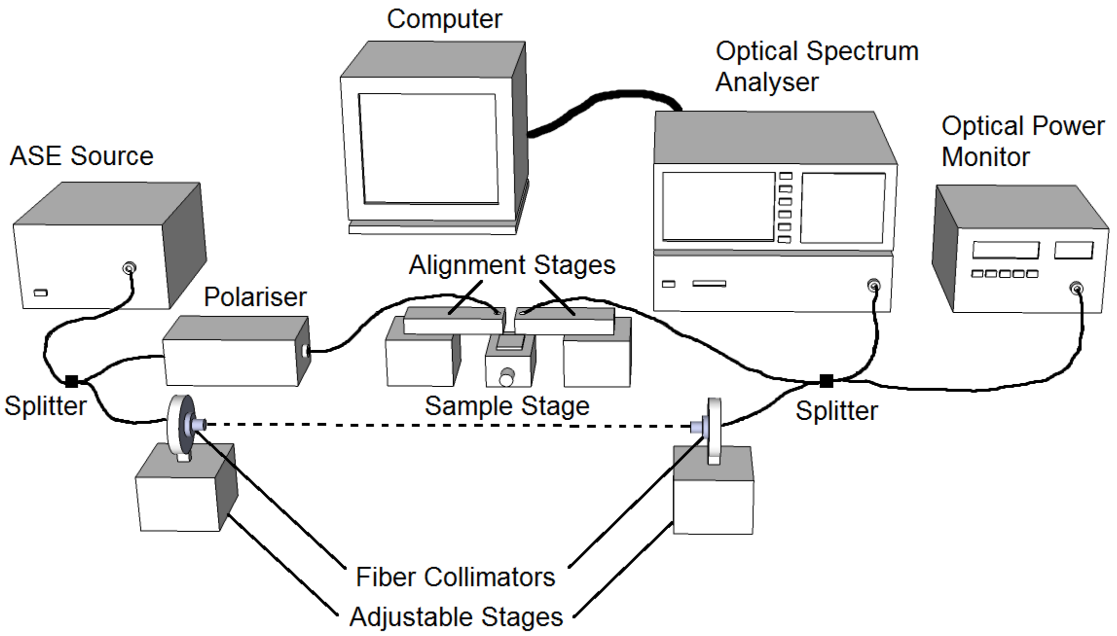


Figure 5.2: Experimental setup for measuring the group refractive index of PCW attenuators.

LabVIEW™ offered remote control of the two modules.

The second experimental setup, illustrated in figure 5.2, was a Mach-Zehnder interferometer used to measure the PCWs' group refractive index as a function of wavelength. The setup consisted of two arms: a sample arm within which the grating-coupler arrangement from the first setup was reused, and a reference arm consisting of two connectorised freespace collimators. The collimators rested on stages capable of xz motion to enable alignment between collimators and y motion to enable the matching of optical path lengths between reference and sample arms. Broadband, continuous-wave light (1530nm-1610nm) from an amplified-spontaneous-emission (ASE) source was distributed between these two arms at one end and converged at the other using fiber-optic splitters. Unlike in the first setup, no component was polarisation-maintaining; a polariser was included in the sample arm to account for this. The trailing splitter distributed the converged light between an Advantest Q8347 optical spectrum analyser (OSA) for data capture and the previously-used power monitor for assistance in fiber alignment. A personal computer interfaced with the OSA for remote, programmable data acquisition.

Following the discussion of section 3.6, group velocity was measured by analysing the interference pattern captured by the OSA. An interference pattern was obtained when the time delay between the sample and reference arms was brought below a value defined by the OSA's resolution (here a 140ps delay, corresponding to a 30pm resolution in freespace). The reference arm was deliberately made shorter than the sample arm so that the slowing of light in the PCW monotonically *increased* the optical path difference and, thus, *decreased* the interference fringe spacing. The reference arm's optical path was initially set using a control waveguide such that the fringe spacing was small enough for efficient filtering, but large enough so that the OSA's resolution was not exceeded even near the PCWs' cutoff wavelength. Once set, interference patterns were acquired for each PCW, with a reduction in fringe spacing with respect to the control waveguide corresponding to a calculable increase in group refractive index. The transmission spectrum of each arm was also acquired for subtractive filtering.

5.1.2 Optical Response

Using the setup depicted in figure 5.1, the amount of power transmitted through the waveguide attenuators was measured. Figure 5.3 shows transmitted power through PCWs of two select dielectric functions and variable length before implantation with defects. The PCW of figure 5.3a had a radius of 120nm, a pitch of 385nm, and a first-row shift of +40nm. The PCW of figure 5.3b had a radius of 115nm, a pitch of 390nm, a first-row shift of +55nm, and a second-row shift of –25nm. (These dielectric functions correspond to PCWs 2 and 5 in table 4.1.) Transmitted power is normalised with respect to that of wire-waveguides of 500nm width and matching lengths. The first PCW (figure 5.3a) exhibits low-loss waveguiding from approximately 1528nm to 1564nm. A constant drop in power of around 5-10dB within this region is attributed to reflection loss at both ends of the PCW. This reflection loss gradually increases near the cutoff wavelength of 1564nm, where the slowing of light exacerbates modal mismatch between the wire-waveguide tapers and the PCW. Beyond cutoff, transmission is minimal, as the PCW no longer supports a guided mode. The gradual increase in power beyond cutoff is attributed to the magnification of noise during normalisation, since the amount of power the grating-couplers coupled into the chip at these lengths fell to values near the noise limit. At wavelengths shorter than 1528nm, transmission loss becomes strongly dependent on length, suggesting that the guided mode crosses the light line and now radiates out-of-plane. A usable bandwidth of 36nm is demarcated by these light line and cutoff regions.

The second PCW (figure 5.3b) exhibits no such low-loss waveguiding. Its transmitted power is a function of length at all observed wavelengths. However, in spite of this, transmitted power remains comparable to that of the first PCW. Following the simulations of section 4.1.1, this is believed to be due to the excitation of two guided modes: y-even and y-odd. The length-dependent loss is attributed to out-of-plane radiation of the y-odd mode, which lies above the light line, whereas the appreciable transmission of power is attributed to the y-even mode, which lies below the light line. In addition, this y-even mode, as it will be shown, has a low group refractive index across all observed wavelengths, so reflection loss at the ends of the PCW is small. This small reflection loss is what enables the 80μm-long PCW to transmit comparable power in spite of being blighted by out-of-plane radiation.

It must be noted that the guided-mode bandwidth of the 320μm-long PCW in figure 5.3a and the 160μm- and 320μm-long PCWs in figure 5.3b are red-shifted by around 3nm. Etch-loading effects are suspected to be the cause of this discrepancy. The chips on which these devices were fabricated were shared by several parties. Due to layout constraints, some PCWs needed to be separated from the others. In doing this, these PCWs were inadvertently placed around 180μm away from an unrelated component designed to have a relatively low silicon density. This component required a larger number of reactive ions for etching, leaving fewer ions available for the PCWs in close proximity. As a consequence, the etch rate of these PCWs was lower than that of the others, so these PCWs ended up with smaller hole radii. This effectively increased the PCWs' average dielectric constant, which is known from theory to shift modes to lower frequencies. Thus, these PCWs experienced a red-shift in their optical response. The red-shift serves as an example of a PCW's high sensitivity to fabrication conditions.

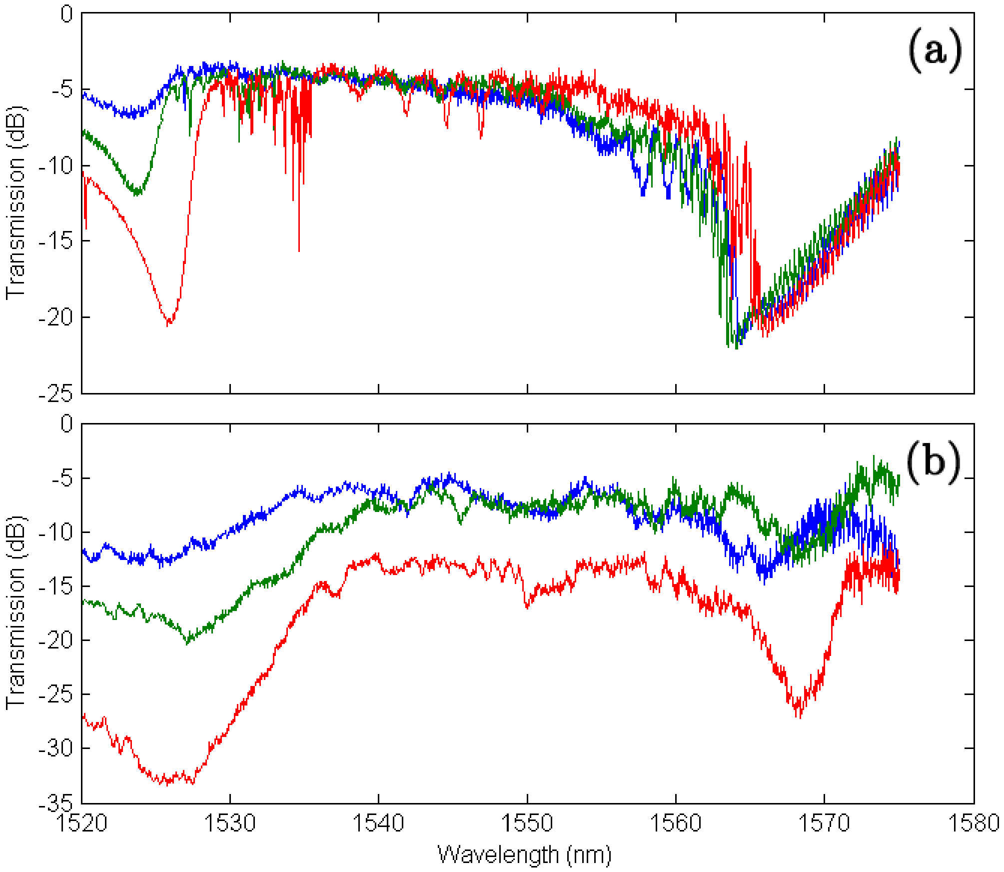


Figure 5.3: Transmitted power as a function of wavelength prior to defect implantation for PCWs (a) 2 and (b) 5 as given in table 4.1 for lengths 80μm (blue), 160μm (green), and 320μm (red). Transmission is normalised with respect to that of 500nm-wide wire-waveguides of matching lengths. The PCW in (a) exhibits total out-of-plane confinement between approximately 1528nm and 1565nm, whereas the PCW in (b) exhibits only partial confinement. Etch-loading effects are believed to be responsible for the 3nm red-shift seen in the 320μm-long PCWs' response.

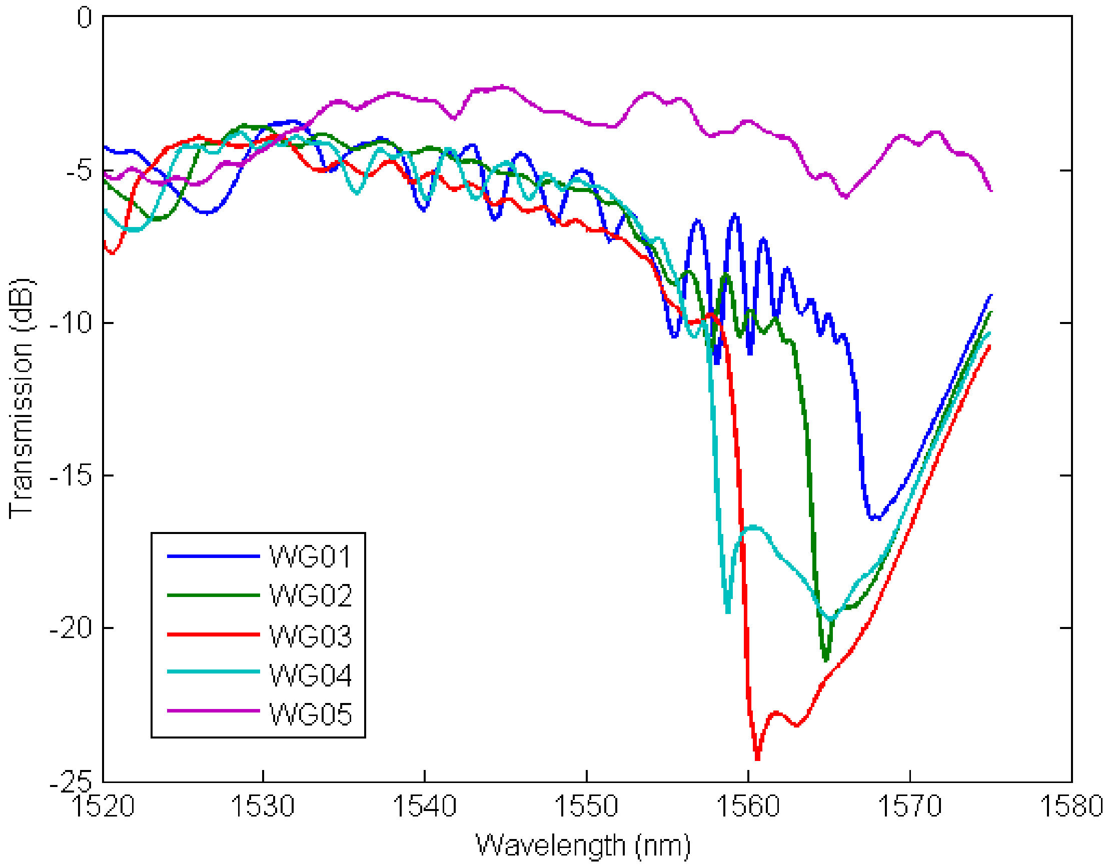


Figure 5.4: Transmitted power as a function of wavelength prior to defect implantation for all five types of PCWs considered. PCW numbers correspond to dielectric functions described in table 4.1. All PCWs are 80μm in length. Transmission is normalised with respect to that of a 500nm-long wire-waveguide of identical length. Data is smoothed using local regression with a 5% averaging window.

Transmitted power for PCWs of all five types of dielectric functions considered are shown in figure 5.4. Each PCW was 80μm in length. Cutoff wavelengths differ in value but qualitatively agree with simulation, with PCWs 1 to 4 exhibiting a cutoff wavelength at successively shorter values, from 1566μm for PCW 1 to 1558μm for PCW 4. Also in agreement is PCW 5's lack of a cutoff wavelength. Significant Fabry-Perot (FP) resonance is observed in PCW 1's transmission. The reduction of oscillatory period near cutoff suggests that this resonance originates from within the PCW, most likely at the multi-stage interface tapers where reflection is predominant. Resonance exists in PCWs 2 to 4, as well, although to a smaller degree, illustrating the dependence of taper efficiency on bulk dielectric function. FP resonance aside, the row-shifting of holes appears, from the transmission curves alone, to have tailored the guided-mode dispersion curve properties in the predicted manner.

Group refractive index profiles shown in figure 5.5 offer a closer look at these dispersion curve properties. The index profiles were measured using the setup depicted in figure 5.2. In both plots, with the exception of PCW 5, all measured indices start from a common value close to 10 and increase sharply as the mode approaches cutoff, indicating that the slow-light effect is achieved. The parabolic shape of these profiles indicates that the dispersion curves' slope gently flattens while approaching cutoff. At cutoff, group index diverges as the slope suddenly flattens to zero. In contrast, PCW 5's group index matches that of the control waveguide's (2.6) for all measured wavelengths, indicating that the waveguide exhibits no such slow-light effect. These results are in qualitative agreement with simulation. Oscillations near cutoff are found to coincide with transmission oscillations attributed to coupling issues. These oscillations have also been observed in time-of-flight measurements [57], suggesting that transfer of oscillation from transmission to index profiles are not symptomatic of the measurement technique employed here. There is a general trend of increasing group index with decreasing transmission. This trend is readily apparent when the two curves are juxtaposed as in figure 5.6. Experimental uncertainty is mostly caused by random phase fluctuations along either optical path of the MZI due to mechanical turbulence. This uncertainty is minimised by the phase-smoothing step and becomes negligible when longer PCW lengths are considered.

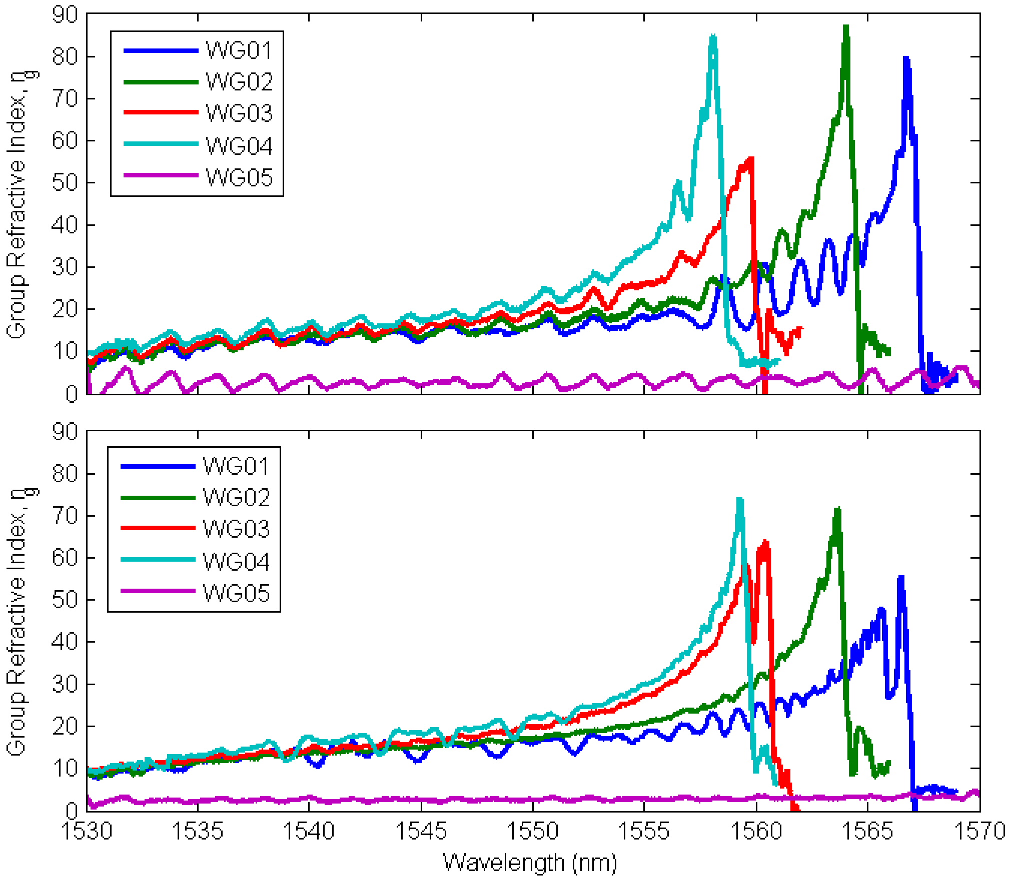


Figure 5.5: Group index profiles for PCWs of length (a) 80μm and (b) 160μm. PCW numbers correspond to dielectric functions described in table 4.1. In (b), PCW 4's response has been red-shifted by 3nm due to what is suspected to be etch-loading effects.

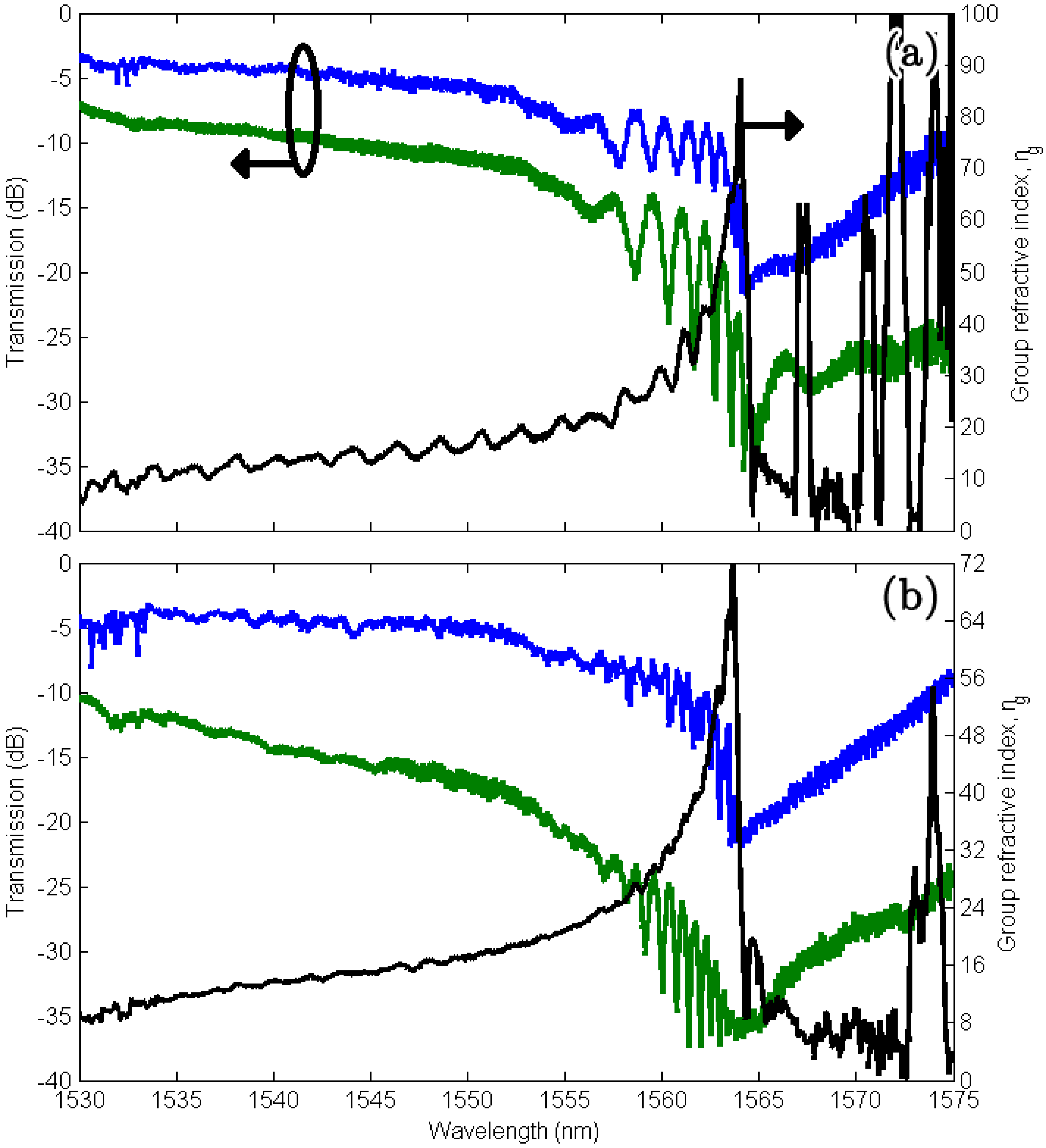


Figure 5.6: Group index and transmission profiles of PCW 2 for (a) 80μm and (b) 160μm lengths. Transmission is given before (blue) and after (green) defect implantation. Transmission is normalised with respect to that of a 500nm-long wire-waveguide of identical length. Change of group index after implantation is negligible.

The measured group index near cutoff approaches a value over 30 times as large as that of an unpatterned wire-waveguide (here 2.6). The rapid change in group index implies that a pulse of light propagating in this spectral region will be heavily distorted. For example, for waveguide 2, a pulse with 1nm linewidth will broaden by approximately 10ps$\cdot$cm^{-1} at 1540nm, or far from cutoff, but by 1400ps$\cdot$cm^{-1} at 1560nm, or near cutoff. Curiously, the maximum index values for the 160μm-long waveguides are lower, in general, than for the 80μm-long ones. This means that the maximum achievable group delay does not scale linearly with physical length. This has been observed before and is believed to be due to the superlinear relationship of optical loss with the slow-light effect near the Brillouin-zone edge [59, 63]. Furthermore, PCWs of longer lengths have more tightly-spaced interference fringes than those of smaller lengths at identical group index values. In other words, the fringe period of longer PCWs exceeds the resolution of the OSA used to measure the interference pattern at lower index values than for shorter PCWs. The deleterious effect of optical loss, together with the finite resolution of the OSA, places a limit on the maximum measurable index, even though the index theoretically diverges to infinity.

Figure 5.6 also shows transmission after defect implantation. It is readily apparent that, unlike without defects, transmission loss after implantation is a function of PCW length. Furthermore, the relative change in loss seems to be proportional to group index. This proportionality is easily seen by observing the difference in transmission far from cutoff, where group index is small, and near cutoff, where group index is large. This difference is presented as *defect loss* on a unit-length basis in figure 5.7. The difference was calculated by smoothing the transmission curves before and after implantation with a 5% averaging window in order to remove coupler-induced oscillations, then subtracting one curve from the other. By taking their difference, systematic losses, such as in coupling, are filtered out, leaving a loss value attributed solely to the presence of defects. The evolution of this value with wavelength matches that of the group index profile: monotonically increasing with wavelength and peaking at 1200dB$\cdot$cm^{-1} at cutoff. By comparison, defect loss in PCW 5 (not shown) was found to be around 200dB$\cdot$cm^{-1} for all wavelengths. The highest peak value among all PCWs measured was 1500dB$\cdot$cm^{-1} for a hole radius of

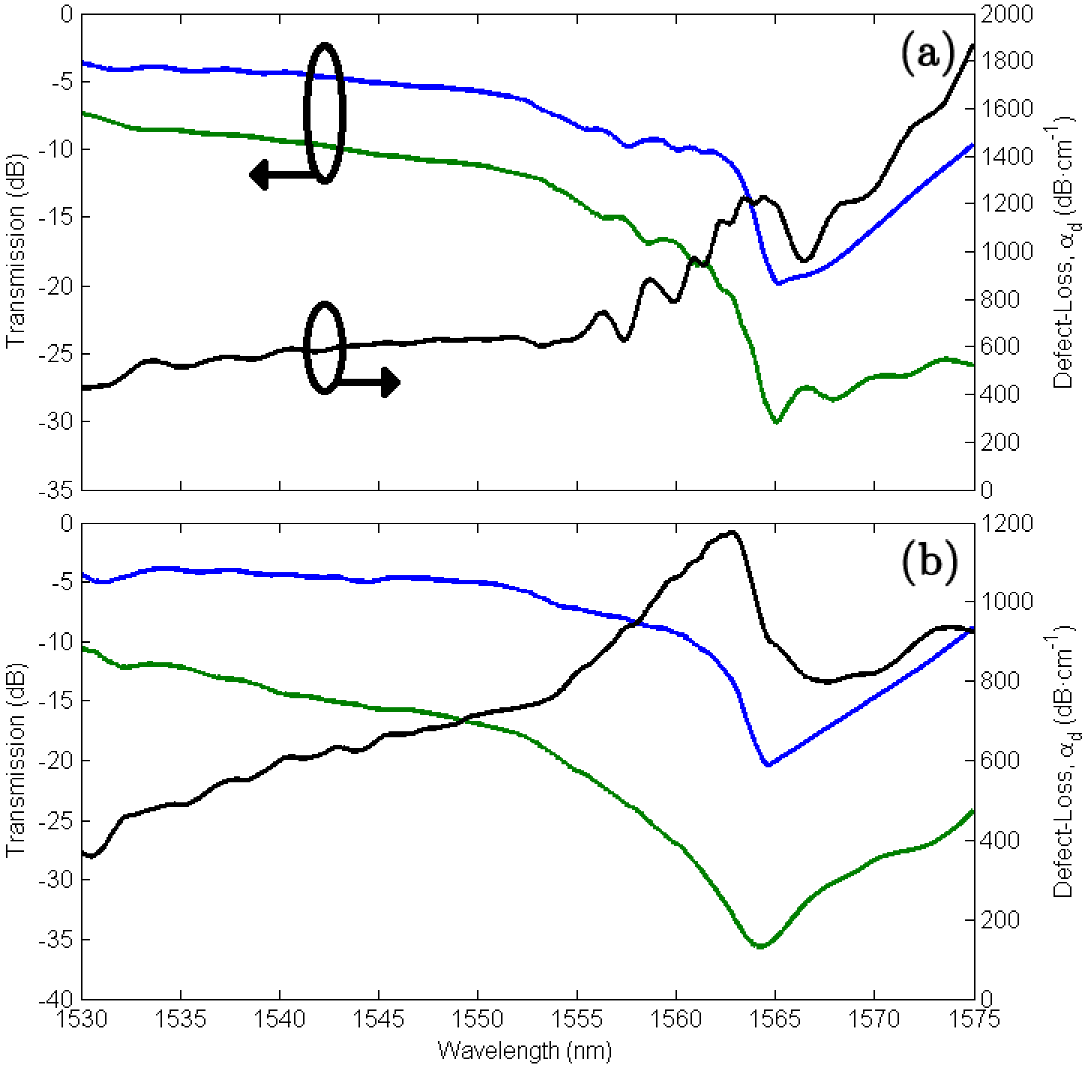


Figure 5.7: Smoothed transmission and defect-induced optical loss of PCW 2 for (a) 80μm and (b) 160μm lengths. Transmission is given before (blue) and after (green) defect implantation. Transmission is normalised with respect to that of a 500nm-long wire-waveguide of identical length.

115nm, a pitch of 385nm, a first-row hole shift of +40nm, and a second-row hole shift of –10nm (PCW 4 in table 4.1). Measurement accuracy was limited by the noise level of the detector (here on the order of 10nW, or approximately –35dB with respect to the calibration waveguide). The maximum peak value is one order of magnitude larger than defect loss previously measured in silicon rib-waveguides for a similar implantation dose [27]. The trend shared by defect loss and group index suggests that this increase is due to slow-light enhancement.

While defect loss and group index indeed share a trend, the former is yet seen to be a nonlinear function of the latter. As shown in figure 5.8, the quotient of defect loss over group index, here called *defect-loss enhancement*, decreases with increasing group index. Defect-loss enhancement is a measure of how efficiently the slowing of light enhances light-matter interaction between the PCW mode (light) and the defects (matter). For the 80μm-long PCW (figure 5.8a), enhancement drops from around 48dB·cm^{-1} at 1535nm wavelength, or a group index of 10, to 10dB·cm^{-1} at cutoff, or a group index of 88. Enhancement values do not fall as quickly for the 160μm-long PCW (figure 5.8b) due to that PCW's index profile being smaller in value for reasons discussed above. The inverse proportionality of enhancement with index is also true among different PCWs at specific wavelengths, where each PCW has a different index value. In fact, enhancement values with respect to group index were in agreement among all PCW types across their common bandwidth range of 1530nm-1557nm. This suggests that defect loss is minimally dependent on wavelength within this range. The sublinear dependence of defect-loss on group index is explained by the lateral spreading of the guided mode's volume at large index values. Such spreading diminishes the positive effect of spatial compression on peak modal intensity.

5.2 Sub-Bandgap Photodetectors

This section presents the results of experiments designed to characterise the electrical and optical properties of integrated PCW photodetectors. These properties include current-voltage (I-V) characteristics, optical loss, spectral transmission, group refractive index, and photocurrent. Spectral transmission and group refractive index

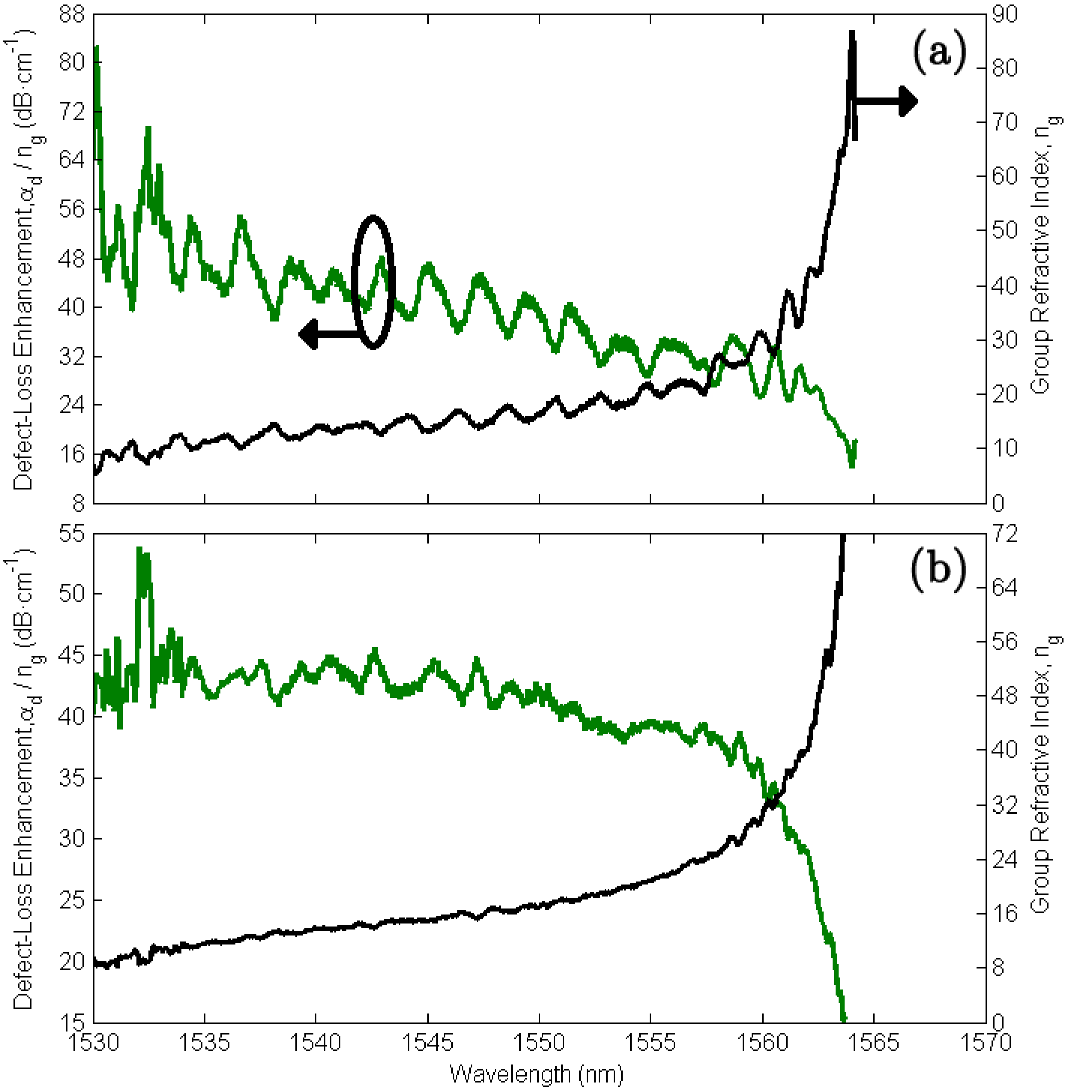


Figure 5.8: Defect-loss enhancement and group index profile of PCW 2 for (a) 80μm and (b) 160μm lengths. Defect-loss enhancement is given as the quotient of defect-loss over group index. A constant enhancement value implies that defect-loss and group index scale linearly. Dependence is sub-linear at large index values. The length-dependence of group index at large index values causes defect-loss enhancement to appear to fall more gradually for longer PCWs than for shorter ones.

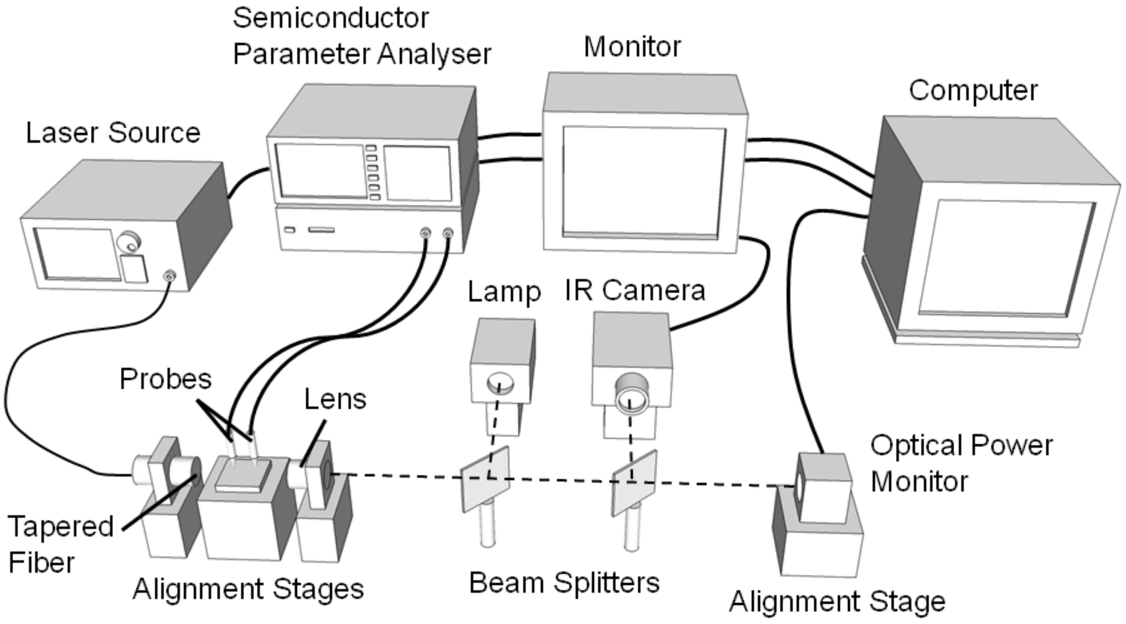


Figure 5.9: Experimental setup for characterisation of integrated PCW photodetectors.

were measured by the Microphotonics and Photonic Crystals Group at St. Andrews University; all other properties were measured at McMaster University.

5.2.1 Experimental Setup

The experimental setup illustrated in figure 5.9 was used to measure the I-V characteristics, spectral transmission, and photocurrent of the integrated PCW photodetectors. A low-loss, tapered optical fiber was used to butt-couple light from a Photonetics OSICS laser diode, tunable across C-band frequencies, into a sample waveguide. A microscope objective collimated light exiting the opposite end of the waveguide and directed it through free-space toward a Newport 818-IS-1 optical power monitor to measure transmission. The fiber, sample, and objective were rested on xyz-translatable stages to assist optical alignment. Optical alignment was a three-step process. First, the objective was focused onto the sample by directing light from a lamp through the objective using a beam splitter, then observing the clarity of the reflection off the sample using a second beam splitter and an infrared (IR) camera

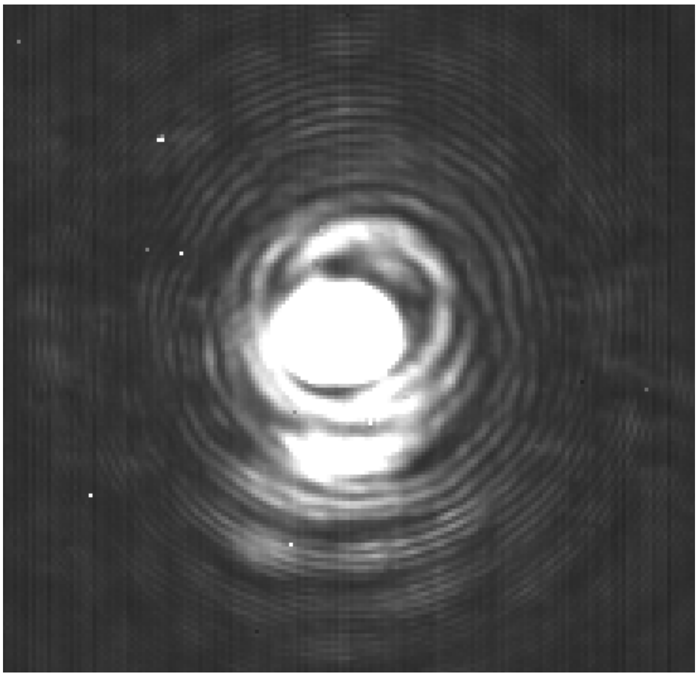


Figure 5.10: The fundamental mode of a PCW photodetector as observed by an infrared camera. The rings of light indicate imperfect collimation by the objective, and the rings' lack of concentricity with the fundamental mode indicates imperfect coincidence of the objective's focal point with the waveguide's facet.

with an attendant monitor. Second, the fiber and objective were aligned with the waveguide's facets using the laser diode and IR camera. Apart from assisting alignment, the image captured by the camera also ensured single-mode propagation. A photograph of such propagation as captured by the camera is shown in figure 5.10. Third, the power monitor was aligned with the objective so as to register maximum power. Once alignment was complete, the beam splitters were removed to eliminate undue reflective loss between the objective and the power monitor. I-V characteristics and photocurrent were measured by placing electrical probe tips in physical contact with the waveguide's metal pads, then leading the probes into a Hewlett-Packard 4145A semiconductor paramater analyser. A personal computer interfaced with the laser diode, semiconductor parameter analyser, and power monitor for remote, programmable, and coordinated data acquisition.

5.2.2 Electrical Response

I-V characteristics of each photodetector were measured before and after Si^+ ion implantation. Figure 5.11 shows clear diode-like characteristics from a photodetector with $x = 4\mu$m and $l = 180\mu$m in the absence of light. Prior to implantation, the device exhibited a turn-on voltage of around 0.7V and a forward-resistance of approximately 600Ω. The turn-on voltage softened and the forward-resistance increased to approximately 1500Ω after implantation. These changes are attributed to the introduction of defects, which here reduce the recombination lifetimes of the electrons and holes that diffuse between the p and n regions and, thus, detract from the forward current. In reverse bias ($-20\mathrm{V} < V < 0\mathrm{V}$), the current in the unimplanted devices was observed to be below the noise limit of the semiconductor parameter analyser ($<$ 1nA). After implantation, this current increased to around 3nA at 20V. This increase is attributed to SRH generation at defect sites in the depletion layer, which adds to drift current.

IV characteristics were seen to be strongly dependent on the separation distance between dopant regions in two ways. One, forward-resistance and separation distance were observed to be proportional. For example, a device with $x = 2\mu$m and $l = 80\mu$m

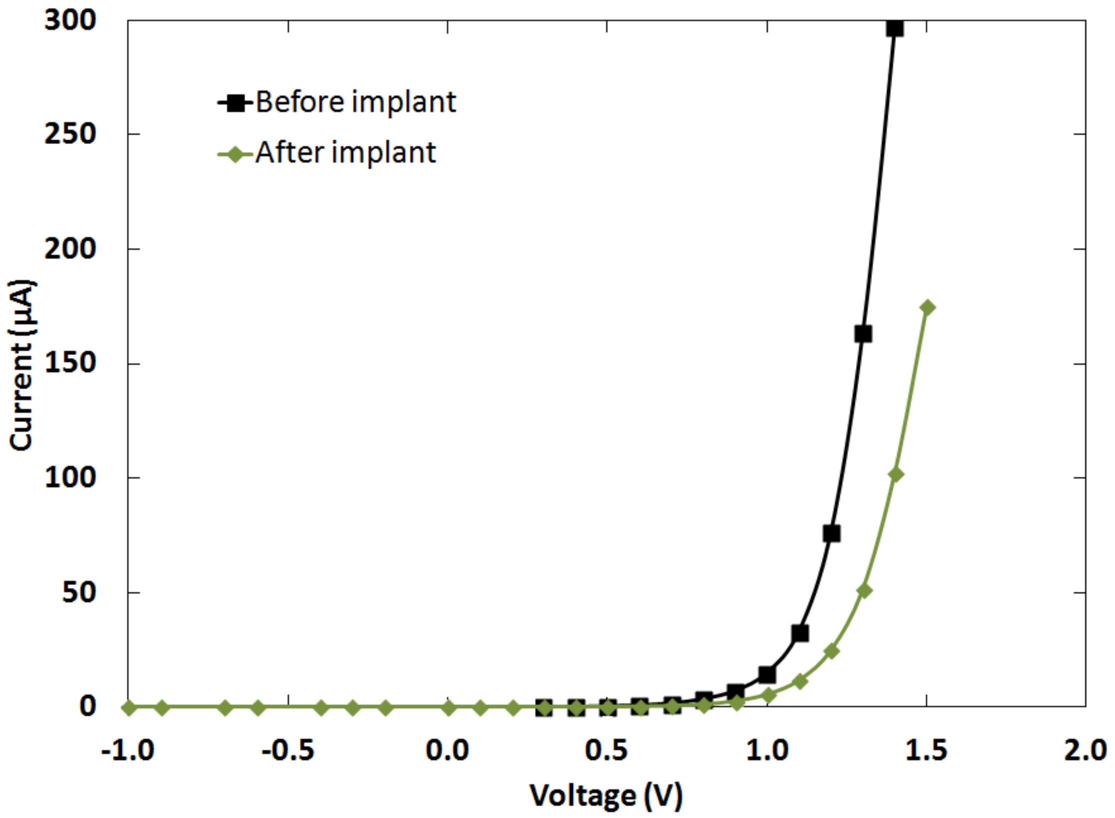


Figure 5.11: I-V characteristics before and after Si^+ implantation at a dose of $3 \times 10^{12}\mathrm{cm}^{-2}$ for a photodetector with $x = 4\mu\mathrm{m}$ and $l = 180\mu\mathrm{m}$.

had a forward-resistance of 235Ω (post-implant). Meanwhile, a device with $x = 8\mu$m – all other things equal – had a much larger forward-resistance of 105 000Ω (post-implant). Two, dark current in reverse bias and separation distance were observed to be inversely proportional. For example, at 20V reverse bias, the device with $x = 2\mu$m had a dark current of 14nA, whereas the device with $x = 8\mu$m had a smaller dark current of 2nA. These differences are attributed to recombination effects: for a device with a wider depletion region, fewer electrons and holes are able to diffuse (in forward bias) or drift (in reverse bias) to the appropriate dopant region before recombining. This is believed to be especially true for a photonic crystal, whose dielectric patterning creates additional surface defects which mediate recombination. In light of these facts, devices with shorter separations are expected to have higher responsivities, albeit with higher noise levels.

5.2.3 Optical Response

Spectral transmission and group index values were measured by the Microphotonics and Photonic Crystals Group at the University of St. Andrews. Figure 5.12 shows the spectral transmission and group index of a PCW photodetector consisting of a triangular lattice of holes with a designed hole radius and pitch of 110nm and 390nm, respectively ($r/a = 0.282$). The data was obtained after the device was implanted with Si^+ ions. Separate measurements at McMaster University incorporating a freespace polariser between chip and photodetector confirmed that transmission corresponds to TE-polarised light; no TM-polarised light is transmitted. The PCW exhibits cutoff at a wavelength of around 1565nm. Transmission is seen to gradually decrease as wavelength approaches cutoff, whereupon transmission falls to its minimum value. As with the waveguide attenuators, this wavelength-dependent loss is attributed to modal mismatch between the PCW and wire-waveguide tapers and enhanced absorption at defect sites. Conversely, group index increases with wavelength until it peaks to a value of approximately 50 at cutoff. The partial suspension in the group index's curvature at around 1550nm is in qualitative agreement with the flexing of the PCW's guided-mode band as predicted in simulation (see figure 4.8).

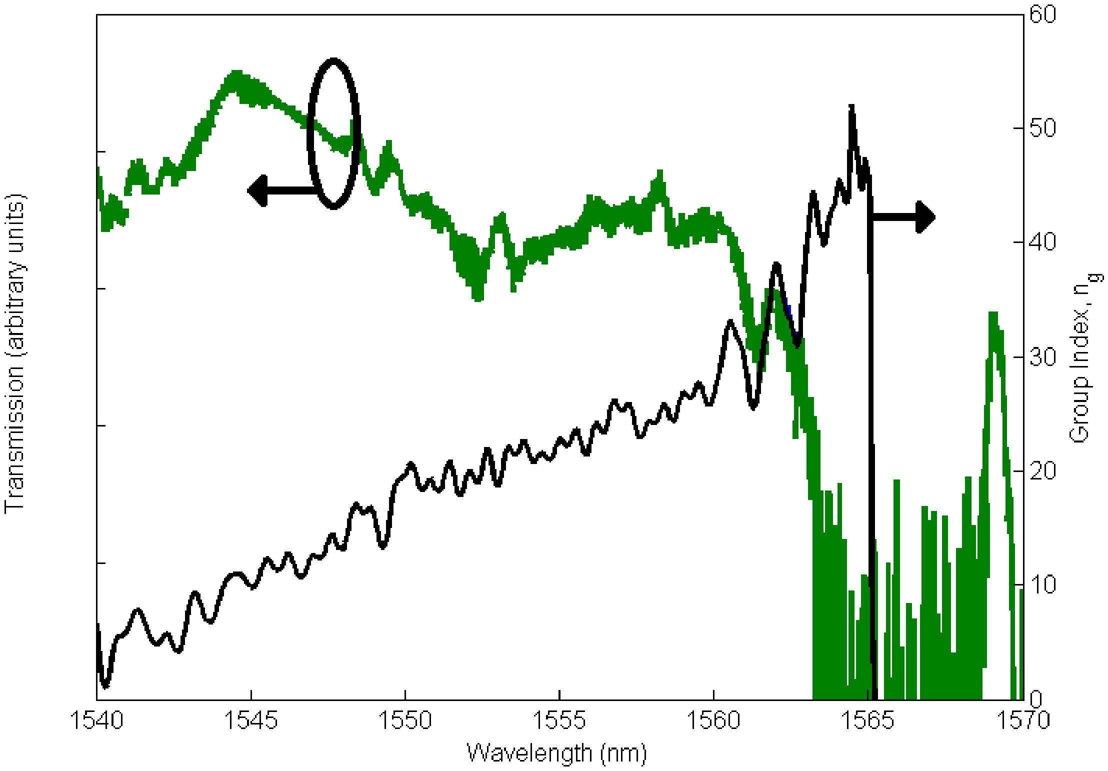


Figure 5.12: Transmission (green) and group index (black) of a PCW photodetector as a function of wavelength. The PCW's dielectric function is a triangular lattice of holes with a designed hole radius and pitch of 110nm and 390nm, respectively. (The actual hole radius and pitch are estimated to be 115nm and 390nm, respectively.) Transmitted power is given in log-scale and is not normalised. Data was taken at the University of St. Andrews.

While experiment is in qualitative agreement with simulation, it is not in quantitative agreement. First, the observed cutoff wavelength of 1565nm does not agree with the simulated cutoff wavelength for either the y-odd mode (1586nm) or y-even mode (1582nm). This disagreement is attributed to a discrepancy in hole radius. In practise, hole radius is difficult to control due to variability of the electron-beam resist thickness, electron-beam spot size, and exposure dose. Conversely, hole pitch and row-shifting are expected to agree with design, as they independent of this variability. Thus, with hole radius being the only parameter that is likely to vary, the true hole radius may be assumed to be larger than the designed size, since cutoff lies at a smaller wavelength than expected from simulation. Follow-up simulations designed to agree with experiment reveal that the true hole radius is around 115nm, giving $r/a = 0.295$.

Second, the band diagram shown in figure 4.8 predicts that the PCW supports two guided modes with cutoff wavelengths within 4nm of each other. If both modes are excited equally, one would expect to observe two peaks in the group index curve – one for each mode. In experiment, however, only one peak is observed. Follow-up simulations reveal that widening the hole radius pushes the cutoff wavelength of the y-odd (fundamental) mode to shorter wavelengths at a faster rate than that of the y-even mode and thus narrows the two peaks' spectral separation. For 115nm-radius holes, separation is nearly zero, so the two peaks cannot be distinguished. In addition, it is possible that one of the modes is not excited as efficiently as the other. If one mode is poorly excited and its transmission is weak, then it might be unable to create interference fringes of a magnitude sufficient to materialise its index curve. This interpretation, however, is inconclusive from the available data. Whatever the case might be, it may be assumed that the group index curve shown in figure 5.12 accurately describes the PCW's optical properties.

The total amount of optical power lost in devices of each length was calculated using data from transmission measurements. Figure 5.13 is a plot of measured optical power loss versus PCW length l for a fixed dopant separation of $x = 4\mu$m at a wavelength of 1560nm. The PCWs had a hole radius and pitch of 113nm and 390nm, respectively ($r/a = 0.290$). Cutoff lied at a wavelength of 1573nm. From the approx-

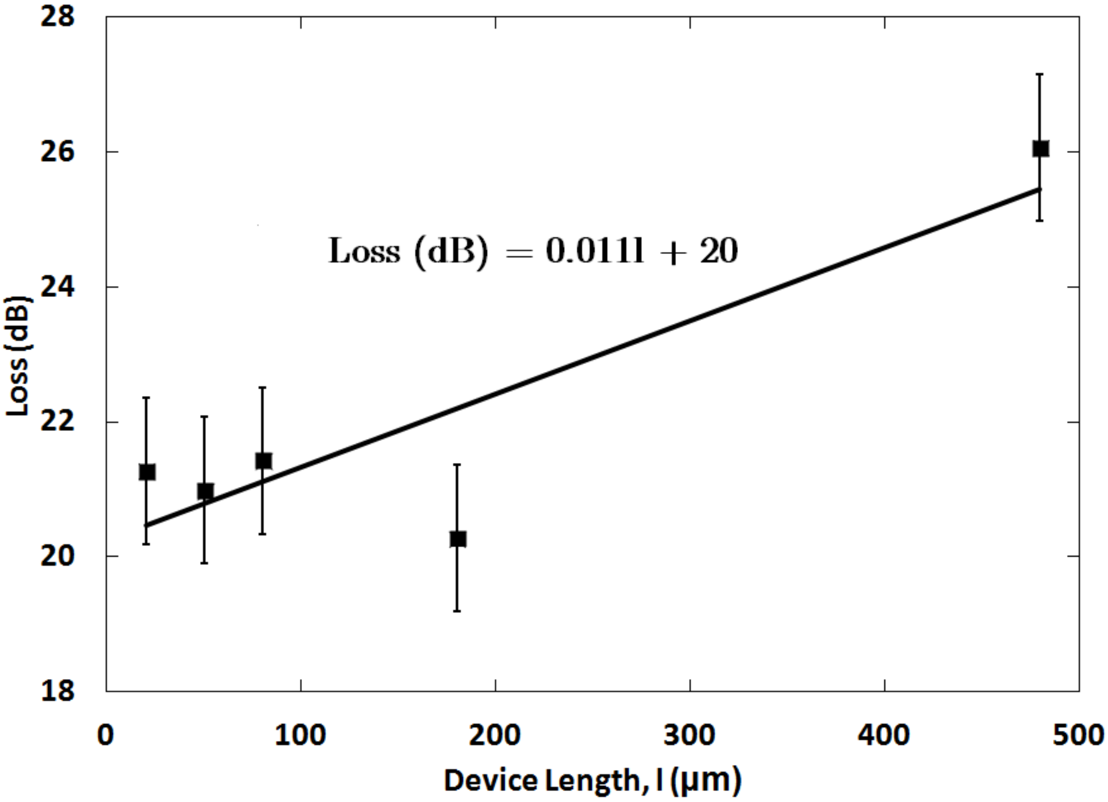


Figure 5.13: Optical power loss as a function of device length, l, including coupling loss at the input and output facets, coupling loss into and out of the PCW, and propagation loss within the wire-waveguides (external to the device) and PCW (internal to the device). Error bars denote uncertainty in incident optical power (5%) and measured transmission (20%).

imately linear relationship between loss and PCW length, propagation loss through the PCWs is estimated to be 110dB$\cdot$cm^{-1}. The majority of this loss is attributed to lattice defects introduced via Si$^+$ bombardment; smaller fractions are attributed to fabrication disorder of the hole radii and sidewall roughness. Again from figure 5.13, total coupling loss, which is a sum of coupling loss at the input and output optical facets, coupling loss into and out of the PCW, and propagation loss within the wire-waveguides, is estimated to be 20dB. The fiber-to-PCW coupling loss is half of this value, or 10dB. From the analysis of waveguide attenuators in section 5.1.2, propagation and coupling losses are expected to be larger at wavelengths closer to cutoff due to lower modal group velocities, but this investigation was inconclusive from the sum of available data.

Devices with a dopant separation of $x = 2\mu$m were seen to have 1-2dB more total loss than devices with wider separations at all measured wavelengths. In general, loss is not known to depend on dopant separation [26]. However, if a dopant region overlaps with the guided mode, then the mode experiences excess loss due to free-carrier absorption. This absorption is believed to account for the additional 1-2dB loss here. According to FDTD simulations, modal width ranges from around 0.7μm (far from cutoff) to 1.4μm (near cutoff). Positional alignment error of each dopant region during fabrication was 1-2μm, so it is possible that the two dopant regions were inadvertently placed closer than 0.7μm from each other. This would explain why the devices have both lossier transmission and smaller forward resistance than the others. Since all devices were fabricated on the same chip using the same photomask, this conclusion implies that all devices have x values that are smaller than desired.

5.2.4 Photocurrent Response

Figure 5.14 shows a plot of photocurrent for devices with length $l = 80\mu$m and $l = 480\mu$m and dopant separation $x = 4\mu$m in response to a tunable diode laser delivering approximately 180μW of power to the chip's input facet. The PCWs had a hole radius and pitch of 113nm and 390nm, respectively ($r/a = 0.290$). Photocurrent is given for three different reverse biases: 0V (dotted line), 10V (dashed line) and 20V

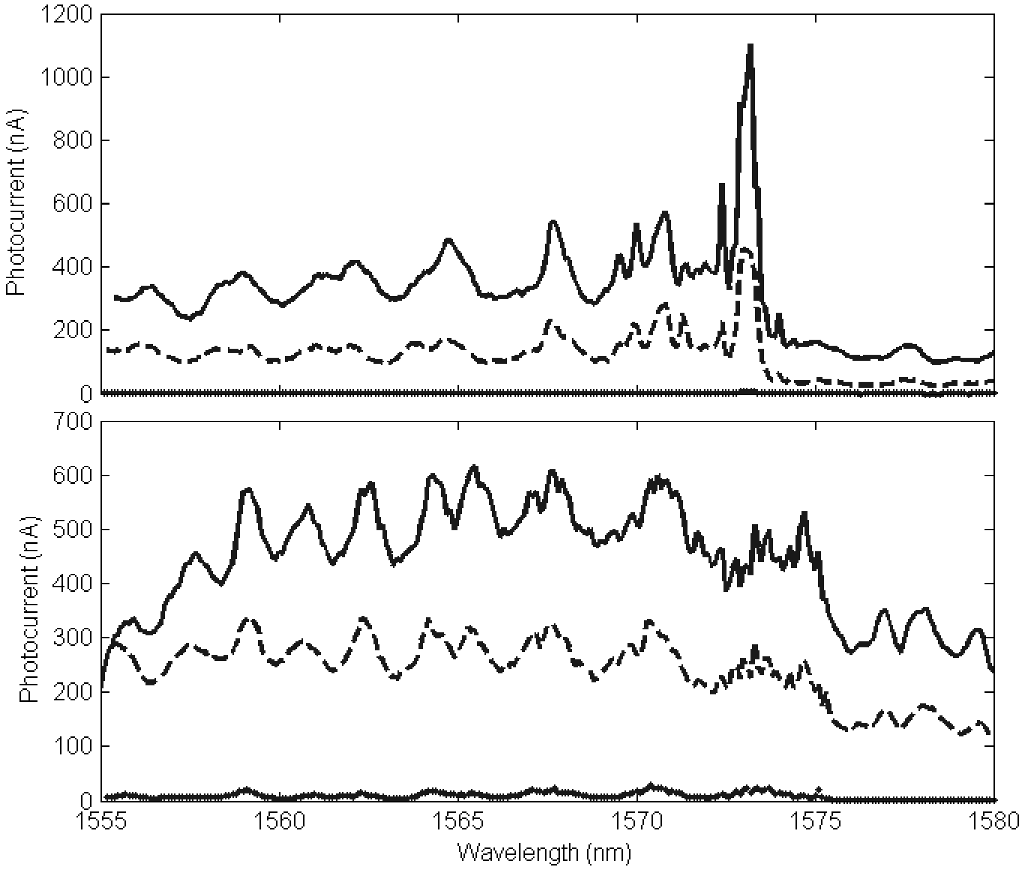


Figure 5.14: Photocurrent as a function of wavelength for devices with length $l = 80\mu$m (top) and $l = 480\mu$m (bottom) and dopant separation $x = 4\mu$m. PCW hole radius and pitch are 113nm and 390nm, respectively ($r/a = 0.290$). Curves correspond to 0V (dotted), 10V (dashed), and 20V (solid) reverse bias.

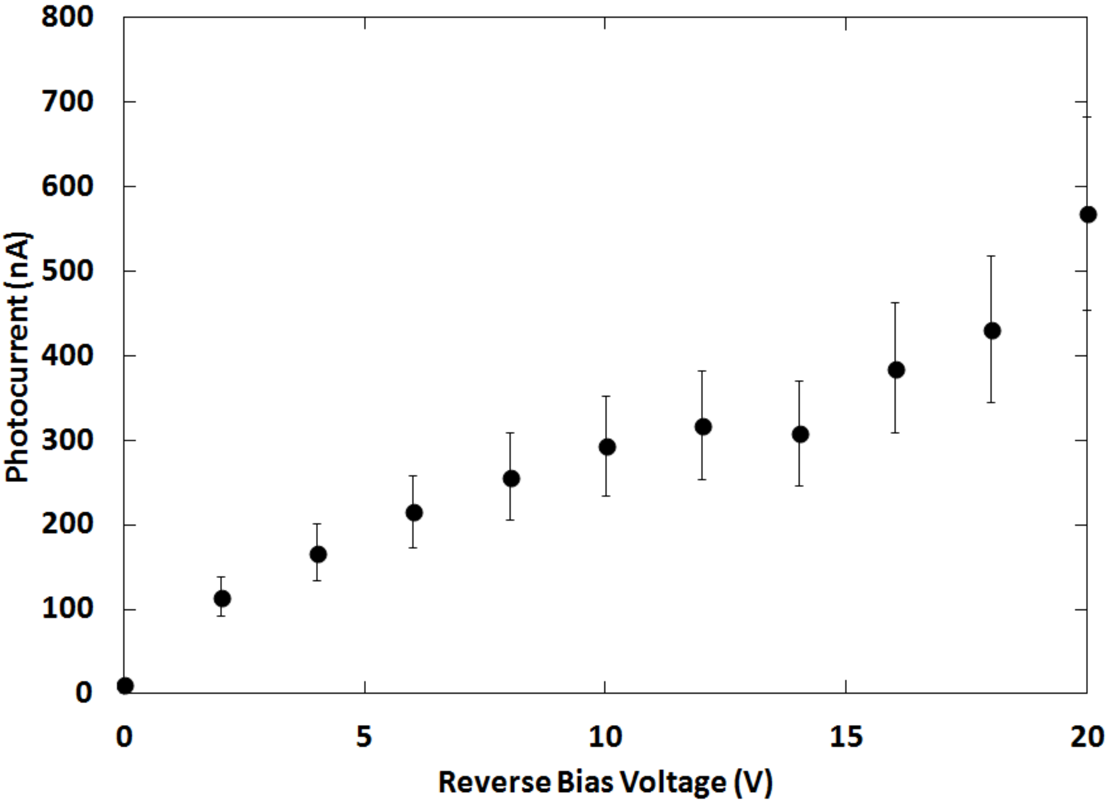


Figure 5.15: Photocurrent as a function of reverse bias voltage for a device with length $l = 480\mu$m and dopant separation $x = 4\mu$m in response to light of 1565nm wavelength. PCW hole radius and pitch are 113nm and 390nm, respectively ($r/a = 0.290$). A measurement error of 20% owing to Fabry-Perot oscillation is assigned to photocurrent.

(solid line). The effect of reverse bias on photocurrent is immediately obvious. At higher reverse biases, a greater number of photogenerated electrons and holes have sufficient drift velocity to separate and contribute to the photocurrent. For both devices, dark current was below the noise limit ($<$ 1nA) at 0V reverse bias. At 20V reverse bias, dark current increased to 6nA for $l = 80\mu$m and 10nA for $l = 480\mu$m. The direct proportionality of photocurrent on reverse bias is illustrated in figure 5.15, which plots photocurrent as a function of reverse bias for $l = 480\mu$m at a wavelength of 1568nm. Photocurrent and reverse bias were directly proportional for all device lengths and dopant separations.

Also apparent in figure 5.14 is the dependence of photocurrent on wavelength.

For $l = 80\mu$m, photocurrent is seen to resemble the parabolic evolution of the PCW's group index curve. This suggests that the slow-light effect enhances SRH carrier generation, as expected. Photocurrent peaks to a value 1100nA at cutoff, which here lies at a wavelength of ~1573nm, for 20V reverse bias. Photocurrent is seen to be modulated by an envelope with a periodicity of approximately 2.8nm. The device's transmission spectrum (not shown) is similarly modulated, with transmission peaks corresponding to photocurrent troughs, and vice versa. The precise origin of this modulation envelope is not known. With 2.8nm periodicity, the envelope corresponds to Fabry-Perot oscillation within a partially reflective cavity with an optical path length of 430nm. It is possible that such a cavity is formed from stitching errors between neighbouring 200μm-long write fields, or between a stitching error and an optical facet. Regardless, the cavity is believed to lie outside the PCW. Otherwise, the envelope's periodicity would shrink in concert with the PCW's modal group velocity.

Curiously, for $l = 480\mu$m, no peak in photocurrent at cutoff is present, yet average photocurrent is higher away from cutoff. The higher average photocurrent is attributed additional absorption of photons owing to a longer device length. The reason for the lack of a peak at cutoff, however, is unclear. At 480μm length, the PCW had to be stitched between three separate write fields of the electron beam writer. Scattering loss at these stitching boundaries is expected to be magnified at lower modal group velocities. Such loss could possibly explain why photocurrent is abnormally low near cutoff. Unfortunately, this conjecture cannot be qualified using the device's transmission curve, as an additive constant of similarly unclear origin masks the transmission curve's spectral features.

Photocurrent was seen to be directly proportional to PCW length l. Figure 5.16 shows a plot of photocurrent as a function of device length. For 1565nm-wavelength light, a device with $l = 20\mu$m provided ~80nA of photocurrent, whereas a device with $l = 480\mu$m provided ~520nA of current. This agrees with the theory of quantum efficiency η: longer devices absorb more passing light, so they yield higher photocurrent values. However, photocurrent is seen to scale sublinearly beyond $l = 80\mu$m. Excess loss, due to stitching errors or otherwise, may explain this sublinear scaling. From the sum of available data, no conclusions could be drawn about the relationship between

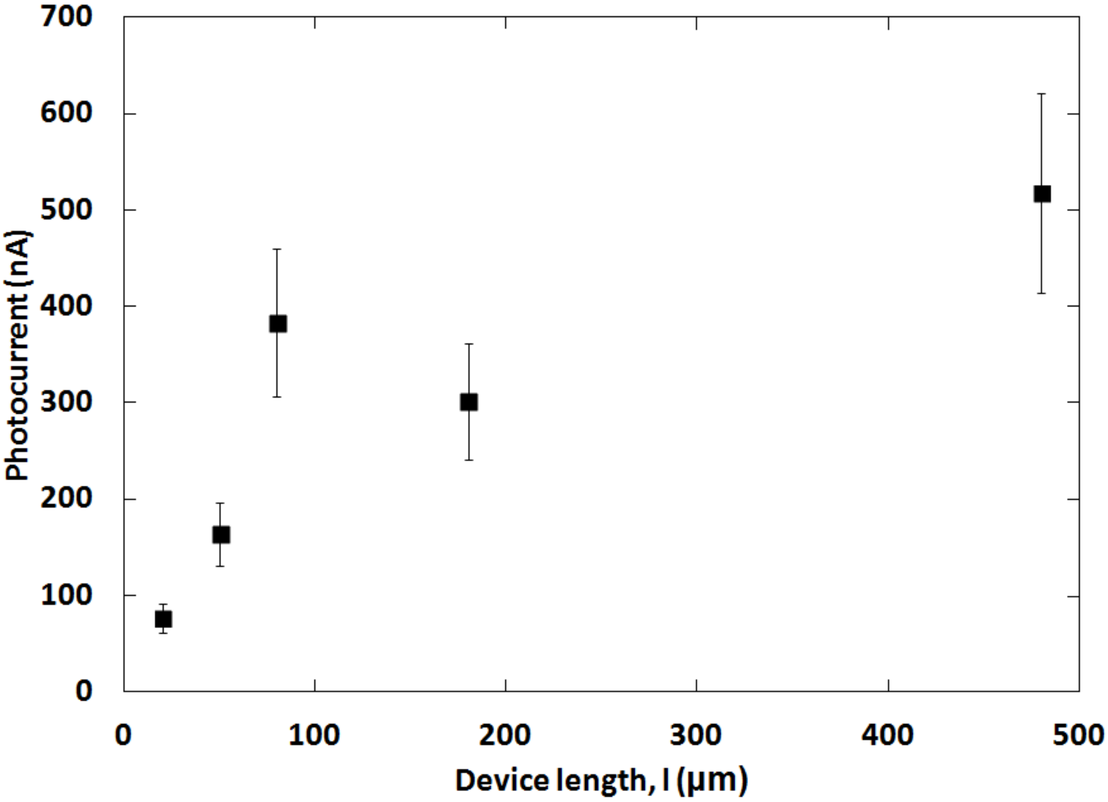


Figure 5.16: Photocurrent as a function of device length at a reverse bias of 20V and a wavelength of 1565nm. PCW hole radius and pitch are 113nm and 390nm, respectively ($r/a = 0.390$). A measurement error of 20% owing to Fabry-Perot oscillation is assigned to photocurrent.

photocurrent and dopant separation x; excess propagation loss experienced in devices with $x = 2\mu$m poisoned their comparison to $x = 4\mu$m and $x = 8\mu$m devices. Work published elsewhere [26] demonstrates that devices with shorter dopant separations yield higher photocurrent due to the shortened distance across which photogenerated carriers must drift. What may be assumed from the data, then, is that there exists some minimum x value beyond which the diminishing effect of free-carrier absorption begins to counteract the enhancing effect of smaller drift length. This minimum value is expected to be defined by the PCW's modal width, which widens for larger group indices. Additional experiments using a larger set of x values and tighter control of positional alignment during fabrication is necessary to investigate this relationship.

The most efficient device ($l = 80\mu$m and $x = 4\mu$m) was measured to generate a photocurrent of 1100nA near its PCW's cutoff wavelength at a reverse bias of 20V. Estimating the amount of power incident on the PCW to be 32μW, responsivity is calculated to be 34mA$\cdot$W^{-1}. The corresponding quantum efficiency is approximately 3%. Note, however, that the PCW is highly dispersive at cutoff. For telecommunications and photonic circuit applications, a more useful measure of the device's performance is at shorter wavelengths, where a signal's integrity may be preserved. At 1565nm, photocurrent was approximately 400nA, giving a responsivity and quantum efficiency of 13mA$\cdot$W^{-1} and 1%, respectively. (The device with $l = 480\mu$m – all other things equal – had a slightly higher responsivity of 16mA$\cdot$W^{-1}.) Propagation loss at 1565nm in this device is calculated to be approximately 110dB$\cdot$cm^{-1}. Estimating the coupling loss into the PCW to be 3dB and attributing all absorption to carrier generation, the photon-to-electron conversion efficiency ξ is approximately 10%. The total optical loss incurred from generating this photocurrent is $\sim$4dB, with an additional 3dB of loss if the light is to be coupled back out of the PCW. With such loss, it is recommended that this device is used either at the terminus of a photonic circuit or in combination with other photonic-crystal components. Nevertheless, this device is noteworthy for achieving a responsivity comparable with rib-waveguide sub-bandgap photodetectors (see, for example, [26] and [27]) while being of sub-millimetre length. Optimising dopant spacing and wire-waveguide-to-PCW coupling may yield responsivities that are yet higher.

Chapter 6

Summary

6.1 Conclusions

A CMOS-compatible silicon sub-bandgap attenuator and photodetector integrated in photonic-crystal waveguides have been demonstrated. The attenuator was shown to exhibit group-index-dependent absorption, with defect-induced propagation losses from 400dB$\cdot$cm^{-1} at 1530nm, or a group index of $\sim$8, to 1200dB$\cdot$cm^{-1} at 1563nm, or a group index of $\sim$88. This dependence is attributed to the slow-light effect. Propagation loss is seen to scale sub-linearly with group index at high index values, which is believed to be due to waveguide dispersion effects and the spreading of modal volume. The photodetector was shown to be sensitive to photons of sub-bandgap energies, demonstrating responsitivites of 14mA$\cdot$W^{-1} at 1565nm and 34mA$\cdot$W^{-1} at 1573nm, which was the photonic crystal's cutoff wavelength. These responsivities are noteworthy not for their magnitude, but from the fact that they are offered by a photodetector merely 80μm long. Thus, integrating a photonic crystal with a photodetector may help reduce the real estate of photonic circuits on a chip or, alternatively, reduce the reverse bias voltage required to attain a certain signal strength. Sensitivity may be enhanced by shrinking the separation between p and n dopant regions while keeping these regions from overlapping with the guided mode and by lengthening the device. This photodetector shows potential for integration at the terminus of densely-packed photonic circuits.

6.2 Future Work

Not all characteristics of the devices featured in this work and their fabrication processes were explored. Notably lacking are time-resolved measurements of the photodetectors. In order to be a competitive technology for integration with micro-electronics, these devices must exhibit temporal bandwidths on the order of gigahertz. Investigative techniques such as the inspection of their eye diagrams may be of assistance in this regard. Temporal bandwidth is expected to be strongly dependent on p-i-n junction capacitance, which is a function of both device length and dopant separation. Of equal importance is the investigation of the effect of dispersion on signal integrity. Theoretically, dispersion can be minimised by designing the photonic-crystal waveguide to exhibit a flat guided-mode dispersion curve. However, fabrication parameters must also be optimised so that the actual hole radii are on target with design. Obtaining such tight control requires rigorous characterisation of the lithography process used to define and transfer the photonic crystal pattern into the silicon layer.

Plenty of room remains in optimising the photodetectors' responsivities. Apart from shrinking dopant separations and increasing their length, the ideal concentration of lattice defects for maximising defect-mediated SRH carrier generation while minimising attendant defect-mediated absorption may be studied. It is known that divacancy-type defect transitions are activated at annealing temperatures on the order of 100 °C [26, 64]. Defect concentration may be gradually reduced by annealing at these temperatures over short, successive periods of time.

Finally, photodetectors integrated with different photonic crystal structures may be investigated. Of particular attention is the integration of defect-mediated carrier generation in a photonic crystal nanocavity. Such a nanocavity has recently been demonstrated to exhibit a Q value of over 10^5 [65]. The concentration of optical energy within a nanocavity, which offers spatial confinement on the order of $(\lambda/n)^3$, may enhance carrier generation in a way analogous to the slow-light effect.

Bibliography

[1] J. J. Saarinen, S. M. Weiss, P. M. Fauchet, and J. E. Sipe. “Optical sensor based on resonant porous silicon structures.” *Optics Express* **13**, 3754-3764 (2005).

[2] N. M. Jokerst et al. “Integrated planar lightwave bio/chem OEIC sensors on Si CMOS circuits.” *Proceedings of the SPIE* **5730**, 226-233 (2005).

[3] B. Jalali. “Can silicon change photonics?” *Physica Status Solidi (a)* **205**, 213-224 (2008).

[4] D. A. B. Miller. “Optical interconnects to silicon.” *IEEE Journal of Selected Topics in Quantum Electronics* **6**, 1312-1317 (2000).

[5] B. Jalali and S. Fathpour. “Silicon photonics.” *Journal of Lightwave Technology* **24**, 4600-4615 (2006).

[6] M. Haurylau et al. “On-chip optical interconnect roadmap: Challenges and critical directions.” *IEEE Journal of Selected Topics in Quantum Electronics* **12**, 1699-1705 (2006).

[7] H. Hogan. “Silicon photonics could save the computer industry.” *Photonics Spectra* **44**, March 2010.

[8] K. C. Saraswat and F. Mohammadi. “Effect of scaling of interconnections on the time delay of VLSI circuits.” *IEEE Transactions on Electron Devices* **29**, 645-650 (1982).

[9] D. A. B. Miller and H. M. Ozaktas. “Limit to the bit-rate capacity of electrical interconnects from the aspect ratio of the system architecture.” *Journal of Parallel and Distributed Computing* **41**, 42-52 (1997).

[10] International Technology Roadmap for Semiconductors. Publicly available at http://www.itrs.net/.

[11] M. Borselli, T. J. Johnson, and O. Painter. “Measuring the role of surface chemistry in silicon microphotonics.” *Applied Physics Letters* **88**, 131114 (2006).

[12] J. Takahashi, T. Tsuchizawa, T. Watanabe, and S. Itabashi. “Oxidation-induced improvement in the sidewall morphology and cross-sectional profile of silicon wire waveguides.” *Journal of Vacuum Science & Technology B* **22**, 2522-2525 (2004).

[13] R. A. Soref. “Silicon-based optoelectronics.” *Proceedings of the IEEE* **81**, 1687-1706 (1993).

[14] L. Tsybeskov, D. J. Lockwood, and M. Ichikazu. “Silicon photonics: CMOS going optical.” *Proceedings of the IEEE* **97**, 1161-1165 (2009).

[15] T. Baehr-Jones, T. Pinguet, P. L. Guo-Qiang, S. Danziger, D. Prather, and M. Hochberg. “Myths and rumours of silicon photonics.” *Nature Photonics* **6**, 206-208 (2012).

[16] G. G. Shahidi. “SOI technology for the GHz era.” *IBM Journal of Research and Development* **46**, 2/3, 121-131 (2002).

[17] G. K. Celler and S. Cristoloveanu. “Frontiers of silicon-on-insulator.” *Journal of Applied Physics* **93**, 4955-4978 (2003).

[18] K. V. Srikrishnan. ”Smark-Cut process for the production of thin semiconductor material films.” United States Patent No. 5882987, 16 March 1999.

[19] P. Dumon et al. ”Linear and nonlinear nanophotonic devices based on silicon-on-insulator wire waveguides.” *Japanese Journal of Applied Physics* **45**, 6589-6602 (2006).

[20] H. Park et al. "A hybrid AlGaInAs-silicon evanescent waveguide photodetector." *Optics Express* **15**, 6044-6052 (2010).

[21] S. Feng, Y. Geng, K. M. Lau, and A. W. Poon. "Epitaxial III-V-on-silicon waveguide butt-coupled photodetectors." *Optics Letters* **37**, 4035-4037 (2012).

[22] L. Colace, G. Masini, G. Assanto, H. C. Luan, K. Wada, and L. C. Kimerling. "Efficient high-speed near-infrared Ge photodetectors integrated on Si substrates." *Applied Physics Letters* **76**, 1231-1234 (2000).

[23] M. R. Reshotko, D. L. Kencke, and B. Block. "High-speed CMOS-compatible photodetectors for optical interconnects." *Proceedings of the SPIE* **5564**, 146-155 (2004).

[24] J. Michel, J. Liu, and L. C. Kimerling. "High-performance Ge-on-Si photodetectors." *Nature Photonics* **4**, 527-534 (2010).

[25] P. G. Kik, A. Polman, S. Libertino, and S. Coffa. "Design and performance of an erbium-doped silicon waveguide detector ooperating at 1.5 μm." *Journal of Lightwave Technology* **20**, 862 (2002).

[26] A. P. Knights, J. D. B. Bradley, S. H. Gou, and P. E. Jessop. "Monolithically integrated photodetectors for optical signal monitoring in silicon waveguides." *Silicon Photonics, Proceedings of the SPIE* **6125**, 119-130 (2006).

[27] M. W. Geis, S. J. Spector, M. E. Grein, R. T. Schulein, J. U. Yoon, D. M. Lennon, S. Deneault, F. Gan, F. X. Kaertner, and T. M. Lyszczarz. "CMOS-compatible all-Si high-speed waveguide photodiodes with high responsivity in near-infrared communication band." *IEEE Photonics Technology Letters* **19**, 152-154 (2007).

[28] H. Y. Fan and A. K. Ramdas. "Infrared absorption and photoconductivity in irradiated silicon." *Journal of Applied Physics* **30**, 1127 (1959).

[29] L. J. Cheng, J. C. Corelli, J. W. Corbett, and G. D. Watkins. "1.8-, 3.3-, and 3.9-μ bands in irradiated silicon: Correlations with the divacancy." *Physical Review* **152**, 761-774 (1966).

[30] A. H. Kalma and J. C. Corelli. "Photoconductivity studies of defects in silicon: divacancy-associated energy levels." *Physical Review* **173**, 734-745 (1968).

[31] R. C. Young and J. C Corelli. "Photoconductivity studies of radiation-induced defects in silicon." *Physical Review B* **5**, 1455-1467 (1972).

[32] L. J. Cheng and P Vajda. "Effect of polarized light on the 1.8-, 3.3-, and 3.9-μ radiation-induced absorption bands in silicon." *Physical Review* **186**, 816-823 (1969).

[33] B. G. Svensson and M. Willander. "Generation of divacancies in silicon irradiated by 2-MeV electrons: Depth and dose dependence." *Journal of Applied Physics* **62**, 2758 (1987).

[34] F. Carton-Merlet, B. Pajot, D. T. Don, C. Porte, B. Clerjaud, and P. M. Mooney. "Photo-induced changes in the charge state of the divacancy in neutron and electron irradiated silicon." *Journal of Physics C* **15**, 2239 (1982).

[35] A. O. Evwaraye and E. Sun. "Electron-irradiation-induced divacancy in lightly doped silicon." *Journal of Applied Physics* **47**, 3776 (1976).

[36] M. Notomi, K. Yamada, A. Shinya, J. Takahashi, C. Takahashi, and I. Yokohama. "Extremely large group-velocity dispersion of line-defect waveguides in photonic crystal slabs." *Physical Review Letters* **87**, 253902 (2001).

[37] L.H. Frandsen, A.V. Lavrinenko, J. Fage-Pedersen, and P.I. Borel. "Photonic crystal waveguides with semi-slow light and tailored dispersion properties." *Optics Express*, **14**, 9444-9450 (2006).

[38] T. Kawasaki, D. Mori, and T. Baba. "Experimental observation of slow light in photonic crystal coupled waveguides." *Optics Express*, **15**, 10274-10281 (2007).

[39] J. Li, T.P. White, L. O'Faolain, A. Gomez-Iglesias, and T.F. Krauss. "Systematic design of flat band slow light in photonic crystal waveguides." *Optics Express*, **16**, 6227-6232 (2008).

[40] T.F. Krauss. “Slow light in photonic crystals.” *Journal of Physics D: Applied Physics*, **40**, 2666-2670 (2007).

[41] C. Monat, B. Corcoran, M. Ebnali-Heidari, C. Grillet, B. J. Eggleton, T. P. White, L. O’Faolain, and T. F. Krauss. “Slow light enhancement of nonlinear effects in silicon engineered photoni crystal waveguides.” *Optics Express* **17**, 2944-2953 (2009).

[42] Y. Akahane, T. Asano, B. S. Song, and S. Noda. “High-Q photonic nanocavity in a two-dimensional photonic crystal.” *Nature* **425**, 944-947 (2003).

[43] B. S. Song, S. Noda, T. Asano, and Y. Akahane. “Ultra-high-Q photonic double-heterostructure nanocavity.” *Nature Materials* **4**, 207-210 (2005).

[44] N. Hauke, T. Zabel, K. Müller, M. Kaniber, A. Laucht, D. Bougeard, G. Abstreiter, J. J. Finley, and Y. Arakawa. “Enhanced photoluminescence emission from two-dimensional silicon photonic crystal cavities.” *New Journal of Physics* **12**, 053005 (2010).

[45] H. Altug, D. Englund, and J. Vučković . “Ultrafast photonic crystal nanocavity laser.” *Nature Physics* **2**, 484-488 (2006).

[46] K. Tanabe, M. Nomura, D. Guimard, S. Iwamoto, and Y. Arakawa. “Room temperature continuous wave operation of InAs/GaAs quantum dot photonic crystal nanocavity laser on silicon substrate.” *Optics Express* **17**, 7036-7042 (2009).

[47] B. Ellis, M. A. Mayer, G. Shambat, T. Sarmiento, J. Harris, E. E. Haller, and J. Vučković. “Ultralow-threshold electrically pumped quantum-dot photonic-crystal nanocavity laser.” *Nature Photonics* **5**, 297-300 (2011).

[48] T. Nakamura, T. Asano, K. Kojima, T. Kojima, and S. Noda. “Controlling the emission of quantum dots embedded in photonic crystal nanocavity by manipulating the Q-factor and detuning.” *Physical Review B* **84**, 245309 (2011).

[49] K. Nozaki et al. “Ultra-low power all-optical RAM based on nanocavities.” *Nature Photonics* **6**, 248-252 (2012).

[50] B. Saleh and M.C. Teich. Fundamentals of Photonics. John Wiley & Sons, Inc, 1991.

[51] A.S. Sedra and K.C. Smith. Microelectronic Circuits: Fifth Edition. Oxford University Press, 2004.

[52] B.G. Streetman and S.K. Banerjee. Solid State Electronic Devices, 6th edition. Prentice Hall, 2006.

[53] S.M. Sze. Physics of Semiconductor Devices: Third Edition. John Wiley & Sons, Inc., 2007.

[54] M.L. Cohen and J.R. Chelikowsky. Electronic Structure and Optical Properties of Semiconductors. Springer-Verlag, 1988.

[55] C. Hu. Modern Semiconductor Devices for Integrated Circuits. Prentice Hall, 2010.

[56] J.H. Svensson, B.G. Svensson, and B. Monemar. "Infrared absorption studies of the divacancy in silicon: New properties of the singly negative charge state." *Physical Review B*, **38**, 4192-4197 (1988).

[57] R. S. Jacobsen, A. V. Lavrinenko, L. H. Frandsen, C. Peucheret, B. Zsigri, G. Moulin, J. Fage-Pederson, and P. I. Borel. "Direct experimental and numerical determination of extremely high group indices in photonic crystal waveguides." *Optics Express* **13**, 7861-7871 (2005).

[58] M. Notomi, A. Shinya, S. Mitsugi, E. Kuramochi, and H. Ryu. "Waveguides, resonators and their coupled elements in photonic crystal slabs." *Optics Express* **12**, 1551-1561 (2004).

[59] A. Gomez-Iglesias, D. O'Brien, L. O'Faolain, A. Miller, and T. F. Krauss. "Direct measurement of the group index of photonic crystal waveguides via Fourier transform spectral interferometry." *Applied Physics Letters* **90**, 261107 (2007).

[60] J.D. Joannopoulos, S.G. Johnson, J.N. Winn, and R.D. Meade. Photonic Crystals – Molding the Flow of Light, 2^{nd} Edition. Princeton University Press, 2008.

[61] S. Kubo, D. Mori, and T. Baba. "Low-group-velocity and low-dispersion slow light in photonic crystal waveguides." *Optics Letters* **32**, 2981-2983 (2007).

[62] Y. Tanaka, T. Asano, R. Hatsuta, and S. Noda. "Investigation of point-defect cavity formed in two-dimensional photonic crystal slab with one-sided dielectric cladding." *Applied Physics Letters* **88**, 011112 (2006).

[63] S. Hughes, L. Ramunno, J. F. Young, and J. E. Sipe. "Extrinsic optical scattering loss in photonic crystal waveguides: Role of fabrication disorder and photon group velocity." *Physical Review Letters* **94**, 033903 (2005).

[64] R. Poirier, V. Avalos, S. Dannefaaer, F. Schiettekatte, and S. Roorda. "Annealing mechanisms of divacancies in silicon." *Physica B* **340-342**, 609-612 (2003).

[65] S. Jeon, J. Han, B. Song, and S. Noda. "Glass-embedded two-dimensional silicon photonic crystal devices with a broad bandwidth waveguide and a high quality nanocavity." *Optics Express* **18**, 19361-19366 (2010).