Center for Energy Efficient Electronics Science

Theme III – Nanophotonics eBook

Acknowledgements

This eBook was written by faculty, postdoctoral researchers, students, and staff of the Center for Energy Efficient Electronics Science (E^3S), a Science and Technology Center funded by the U.S. National Science Foundation (Award 0939514). The Center is a consortium of five world-class academic institutions: University of California at Berkeley, Massachusetts Institute of Technology, Stanford University, University of Texas at El Paso, and Florida International University. Researchers at E^3S are working in a collaborative and innovative environment to make fundamental and conceptual breakthroughs in the underlying physics, chemistry, and materials science of electronic systems, breakthroughs needed to reduce these systems' energy consumption by orders of magnitude.

The goal of the Nanophotonics team is to enable optical communications between switches on a chip at unprecedented efficiency levels. In fact, E^3S researchers pursue the ultimate goal of experimentally approaching the quantum limit of photons-per-bit in a data-link. To meet this goal, the Center for E^3S strives to improve energy efficiency and sensitivity of both the emitter and the photo-receiver. Central to the E^3S nanophotonics research goals has been the demonstration that spontaneous emission from antenna enhanced nano-LEDs can be faster and more energy efficient than the stimulated emission of lasers, the ubiquitous light source in optical communications today. The team is led by UC Berkeley Professor Ming Wu.

Thank you to Dr. Velveth Klee for her thorough review of this book. Dr. Klee is a former participant of the E^3S Research Experiences for Teachers program and a Professor of Physics and Engineering at Los Angeles Trade Technical College.

Table of Contents

Chapter 0: What is Light?

Sean Hooten

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

0.0 INTRODUCTION

What is light? At first glance this question appears simple. Phenomenologically, we understand light to be the "stuff" that allows us to see – it seems to fill darkness and reveal underlying colors, shapes, and textures of objects. Upon further scrutiny, we notice that light can have its own distinct colors (rather than just the white light from the sun), it can be "pointed" in specific directions, and it interacts uniquely with matter and surfaces like mirrors. What causes these behaviors? Why does light interact with objects in different ways, and what separates red light from blue light? How are we able to use light to capture and communicate information, as in a camera image or the display of your computer screen? Why does laser light seem so different than sunlight? Indeed, what was originally a simple question brings up many more complicated questions after further investigation. In fact, the scientific study of light began as early as the 1600s and continues to this very day. Many of the greatest physicists in human history studied light and contributed to the modern theory, which has literally and figuratively allowed us to shed light on our understanding of the universe.

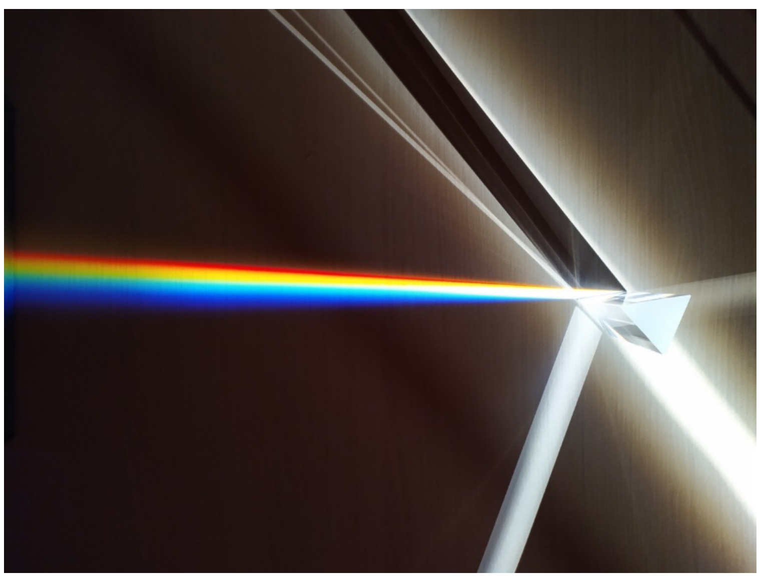

Figure 0.1: A glass prism separates white light into its many colors. This phenomenon was first demonstrated by Isaac Newton in 1666.

What is the modern theory and understanding of light? Without going into too much detail (as this would require several college-level courses for a complete understanding), light is a form of electromagnetic energy. "Electromagnetic" means light is closely related to electricity and magnetism. Light behaves in distinct ways when interacting with objects because of the electric and magnetic properties of those objects. Furthermore, light is a wave, much like an ocean wave, except that it can propagate in many types of matter or even the emptiness of space. A propagating light wave's "energy" is what gives it color (lower energy light is red, higher energy light is blue, and combining many different colored light waves produces white light as in Figure 0.1). After propagation, the energy of light can be converted into different forms of energy like heat or electricity. The inverse is also possible, whereby light can be created from electricity, for example. These last two concepts will be the most important subjects in this eBook.

0.1 OUTLOOK OF THE E3S THEME III "NANOPHOTONICS" EBOOK

"Nanophotonics" refers to the study and use of nanoscale devices that can interact with, create, and detect light in novel ways. This can be understood by noticing that a "photon" is a particle of light and hence "photonics" refers to methods for light manipulation, much like "electronics". In particular, the goal of the E3S Theme III research is to use light for communication (i.e. to send digital information) on a very small scale. Light is already used for communication on a large scale and the modern internet would not be possible without it, but the small scale brings a unique set of challenges that require the research of special materials and devices. We are interested in this application because communication using light can be very "efficient" – meaning that it uses less energy than alternatives such as electrical wires that are currently used for integrated circuits (e.g. on a computer chip). For the practical replacement of these on-chip electrical wires, nanoscale light communication must not only be efficient, but fast and easily manufacturable.

The ensuing chapters of this eBook will delve deeper into the physics of light and its use for nanoscale (on-chip) optical communications. Chapter 1 discusses optoelectronic materials – materials that can be used to create and detect light using electricity. Chapter 2 gives a broader overview of the physics and devices for light creation, detection, and manipulation as well as the tools that researchers use to develop new optical devices. Chapter 3 discusses the antenna-LED – a nanoscale device that can potentially communicate light at ultra-fast speeds very efficiently. Chapter 4 details nontraditional "monolayer" optoelectronic materials. Chapter 5 demonstrates how an antenna-LED can be connected to an optical wire ("waveguide"), in order to transport light across a chip. Finally, Chapter 6 briefly discusses novel and advanced directions of E3S Theme III research.

Chapter 1: Optoelectronic Materials

Yunjo Kim and Jeehwan Kim

Department of Mechanical Engineering
Massachusetts Institute of Technology

1.0 INTRODUCTION

This chapter will introduce concepts surrounding semiconductor materials that are central to the operation of optoelectronic devices. Basic background in chemistry will be helpful for understanding, but it is not required. By the end of this chapter, we will have covered how electrons can be used to generate light in semiconductors and how the color of the emitted light can be tuned in semiconductors.

1.1 OPTOELECTRONIC MATERIALS TODAY

The topic of optoelectronics came into the vernacular in the 1960's [1-4], but half a century later it is difficult to imagine a modern world without the existence of optoelectronic devices. These devices are the primary means through which electronic devices interact with light. These are the devices responsible for generating light on our displays, sensing light through our cameras and transmitting information using light that keep our world connected. It was in the 1950's that scientists developed our understanding of how certain electronic materials were able to interact with photons [6, 7]. Over time, through innovations in material science, scientists were able to design and create a plethora of electronic materials for a wide range of applications. Among these applications, the most impactful invention was the laser, which is primarily used in data communication today [8-14]. We can transmit data by pulsing the light source and measuring the changing intensity of the light from a detector. While any light source can be used to achieve this, lasers offer the fastest means of transmitting data capable of transmitting up to 40 Gb/s on a single channel [15, 16]. It is no wonder lasers and fiber optic communication is the standard through which data is communicated. Moreover, optoelectronic materials are extremely energy efficient in converting electrical energy into light and vice versa. Due to its energy efficiency, light emitting diodes (LEDs) are quickly taking over the lighting industry [17]. It is difficult to ignore their presence when they are being used to light our phones, computer monitors and our rooms. Optoelectronic materials are also used as light sensors such as the cameras in our phones, but it has far ranging applications in medical imaging, space surveillance, and temperature sensing to name a few [8-20]. While all optoelectronic materials behave similar to each other, it is quite surprising how scientists were able to apply it in such a large application space. So, what is the characteristic of an optoelectronic material that enables so many device technologies? In order to answer this question, we will give a discussion to the properties of an optoelectronic material and the basic physical mechanisms that are observed in these materials.

1.2 SEMICONDUCTOR PROPERTIES

In optoelectronics, semiconductors are an important class of materials that enable scientists to control the interaction between electrons and light. Thus, understanding semiconductor properties is essential to develop our understanding in optoelectronics. Throughout this chapter, we will find how characteristics of certain semiconductors enable the emission of colored lights that we observe in everyday lighting systems.

First, we must define what is a semiconductor and what are its unique electronic properties that distinguishes them from other materials. One can view semiconductors as having electronic characteristics that lie somewhere between metals and insulators. In metals, the electrons are loosely held by the atoms. The weak binding energy allows for electrons to freely move inside conductors. Inversely, in insulators the electrons are tightly bound by atoms and does not allow free movement of electrons. Semiconductors have a moderate binding energy with electrons that allow for their limited movement and depending on how scientists design semiconductor devices, the electrons can be controlled to behave as insulators or metals.

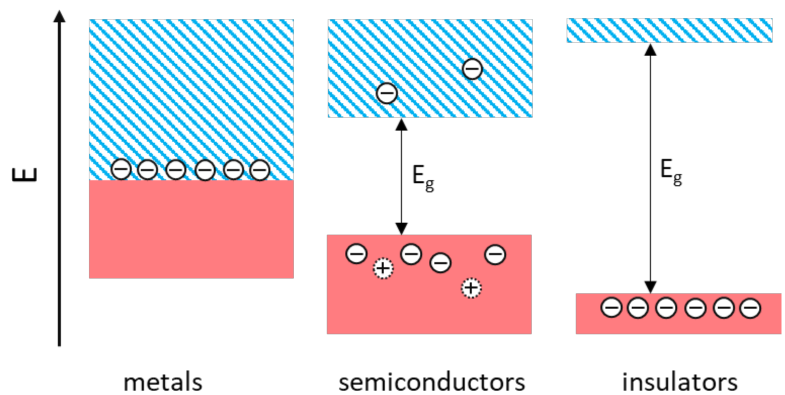

Figure 1.1. Band occupation of varying materials, red-solid shading represented energy states occupied by electrons, blue-dashed shading represents empty energy states.

One can visualize the binding energies of electrons through energy levels within electronic materials. In electronic materials there exists a set of allowable energy levels for which electrons can occupy. In a material with many atoms, these energy levels exist close to each other and can be considered as a continuum of energy levels which are referred to as 'bands' [21]. Electrons, in an attempt to minimize their energy, fill the lower energy bands first. In Fig 1.1, the red-solid boxes indicate energy bands that are packed with electrons. These bands are referred to as valence bands. In order for electrons to freely move and for the material to conduct, the electrons must be adopted into energy levels that are unoccupied by other electrons. These unoccupied energy bands are called conduction bands and they lie above the valence bands as illustrated in blue-dashed boxes in Fig 1.1. As an analogy, we can imagine electrons as cars that are initially stuck in a congested road lane (the valence band) and next to the lane is an empty road (the conduction band). In order for the cars to flow out of traffic, they must change lanes into the empty lane. The cars that have successfully changed lanes are now able to facilitate the flow of traffic. In metals, electrons easily escape into the conduction band and can freely move. In insulators, electrons are fully packed inside their valence band and a significant energy is necessary to promote the electrons into the conduction band to facilitate conduction. This domain of forbidden energy levels is referred to as the bandgap of the material and it is characterized by the bandgap energy E_g, which is defined as the energy difference between the top-most energy level of the valence band and bottom-lowest energy level of the conduction band. In semiconductors the bandgap between the valence band and conduction band is small enough such that electrons in the valence band are readily promoted into the conduction band. Consequently, the promotion of an electron into the conduction band leaves behind a vacant position in the valence band. Charge-neutrality dictates that this vacant position must hold a positive charge. The behavior of these positive charges in the valence band are of significant importance in semiconductors and are referred to as ‘holes’. The electronic behavior of semiconductors is made simple, by understanding holes as a positively charged particles that freely moves in the valence band. Thus, the conduction of semiconductors is facilitated by the movement of electrons in the conduction band and the movement of holes in the valence band. In optoelectronics, it is the interaction of the hole-electron pair that generate light.

For electrons in the valence band to be promoted into the conduction band, the electron must be supplied sufficient energy that is larger than the bandgap energy E_g. One way to supply this energy is by absorbing the energy from photons. The energy of a photon is directly related to its wavelength $E_{ph} = hc/\lambda$, where E_{ph} is the energy of photon, h is the Planck’s constant, c is the speed of light and λ is the wavelength of the light. The maximum wavelength of light necessary to create conduction electrons is given as $\lambda_g = hc/E_g$, referred to as the gap wavelength. A wavelength of light less than or equal to the gap wavelength can be absorbed by the electron in the valence band to generate an electron-hole pair. Alternatively, electrons in the conduction band can recombine with a hole in the valence band and release its energy by emitting a photon with wavelength λ_g. The emission of light by combination of an electron-hole pair is known as radiative recombination [22]. It is this absorption and emission of photons that is central to the operation of optoelectronic devices.

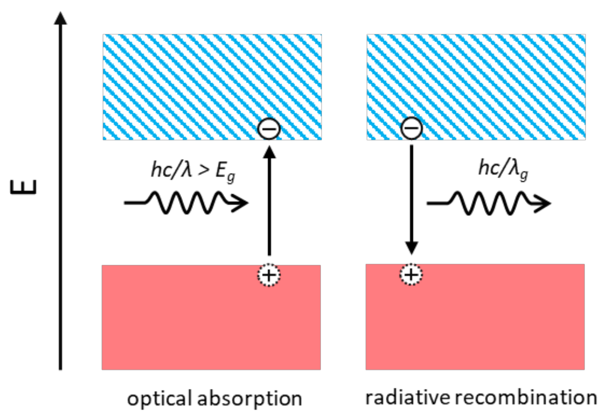

Figure 1.2. Band-to-band optical absorption and radiative recombination mechanism.

1.3 SEMICONDUCTOR ALLOYS IN OPTOELECTRONIC DEVICES

Based on this fundamental relationship between bandgap energy and photon energy we can tailor the bandgap of a semiconductor through material synthesis in order to create an emission at a desired wavelength. Fig 1.3 illustrates a map of various semiconductor materials characterized by their bandgap energy in the y-axis and lattice constant in the x-axis. The lattice constant refers to the dimension of the basic unit cell of a semiconductor crystal, for now we can ignore this value. As can be seen in the map, Si emits light in the infrared domain at 1130 nm wavelength below the range of visible light.

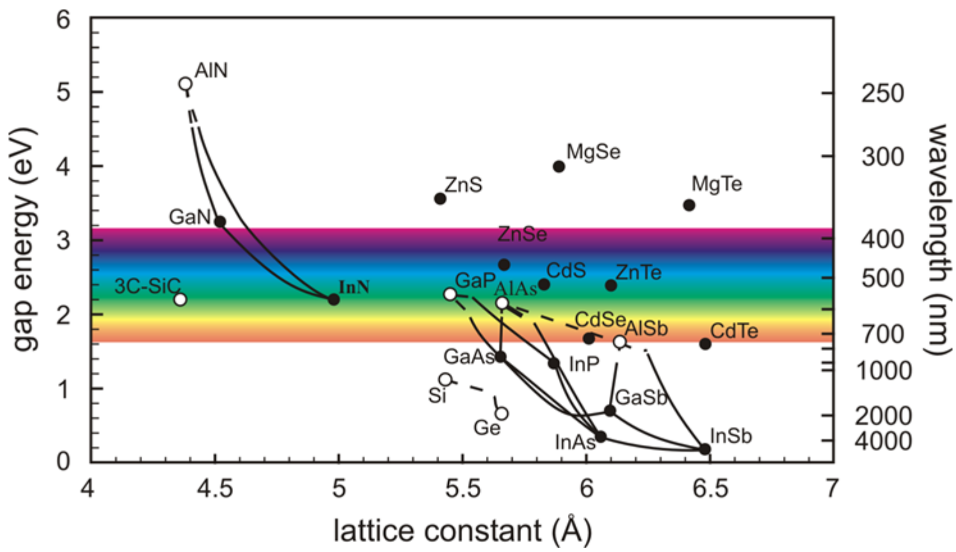

Figure 1.2. Bandgap vs lattice constant of various semiconductor materials [23]

There are also a wide variety of semiconductors composed of two elemental species from Group III and Group V elements or Group II and Group VI elements on the periodic table. These are indicated in the map, such as GaAs, InP, GaN, etc… Semiconductors composed of two elemental species are referred to as compound semiconductors. These compound semiconductors can be made as a mixture of other compound semiconductors to yield a semiconductor alloy with a modulated bandgap. For example, GaAs compound semiconductors can be mixed with GaP to yield $GaAs_{1-x}P_x$ where x denotes the mole fraction. GaAsP semiconductors can be used to emit light with a wavelength that lie within the gap wavelength of GaP and GaAs. GaAsP semiconductors was the base material used to fabricate the first commercial red light emitting diodes (LEDs) [24, 25], Fig 1.4.

Figure 1.3. Red LED display of a TI-30 scientific calculator [26]

As evident today, a wide variety of LEDs are in production due to progress in manufacturing high quality semiconductor alloys. Today, InAlGaP alloys are used to make red and yellow LEDs [7] and InAlGaN alloys are used to produce blue and green LEDs [27]. By modulating the mole fraction of each elemental component, a semiconductor alloy can be engineered to emit a wide range of wavelengths. The following chapters will discuss the mechanisms of optoelectronic devices based on the platform of semiconductor materials.

1.4 INFRARED OPTOELECTRONIC MATERIALS

Sensing and transmission of infrared (IR) light is another important field in optoelectronics that has applications in thermal imagining, proximity sensing and data communication. In this field, GaSb and InAs have been investigated for IR sensing in the short wavelength IR (SWIR) range 1.4–3 µm and mid wavelength IR (MWIR) range 3–8 µm. GaSb and InAs are suitable materials for sensors as their band wavelengths are 1.70 µm and 3.50 µm, respectively [19, 20, 28-30]. In the E3S Theme III research, we specifically investigate the emission properties of semiconductors that emit below the visible regime of light in the IR domain. In concluding this chapter let us discuss some of the properties and applications of IR optoelectronic materials, specifically InGaAs.

1.5 INGAAS EMITTERS FOR ON-CHIP OPTICAL COMMUNICATION

Optoelectronic materials that emit infrared (IR) light are particularly desirable for optical data communication on silicon chips. Silicon is a necessary platform for on-chip optoelectronic devices since all digital computation is done using Si-based devices. However, integrating optoelectronics on Si limits our material choice for fabricating optoelectronic devices since Si absorbs light with energy greater than ~1 eV, thus it is easier to utilize IR light for data communication. Due to this restriction, InGaAs semiconductor is used to emit optical data since the energy of emitted photons fall below 1 eV [30].

InGaAs is a ternary alloy composed of In and Ga from the group III elements and As from the group V element. The In/Ga mole fractions can be tailored to modulate the bandgap energy from 1.441 eV which corresponds to GaAs, down to 0.35 eV which corresponds to InAs. However, designing ternary material systems have many challenges from a processing perspective. While binary semiconductor alloys such as GaAs and InAs can be made as a bulk crystal, ternary alloys are most often produced through deposition processes, which builds the crystal through one atomic layer at a time. The process of growing semiconductor crystals layer-by-layer is known as epitaxy [31-33]. Epitaxy of InGaAs is easily achieved by utilizing an InP crystal as a starting platform for growth. This is because InGaAs shares a similar crystal dimension as InP, a lattice constant of 5.8687 Å, shown in Fig 1.3. In epitaxy, crystals of similar dimensions interface with each other easily since the crystal structure is not required to change shape as the material system abruptly transitions from InP to InGaAs. The alloy composition of $In_{0.53}Ga_{0.47}As$ has the same characteristic dimensions of InP and has a bandgap energy of 0.74 eV which emits at the infrared wavelength at 1675 nm. In the E3S Theme III research, InGaAs plays a prominent role as an IR emitter for many of the novel emitter designs that are investigated.

REFERENCES

[1] Yuri R. Nosov, "History of optoelectronics: general review," Proc. SPIE 4761, Second Conference on Photonics for Transportation, Apr. 2002.
[2] A. Yariv and P. Yeh, Photonics—Optical Electronics in Modern Communications, 6^{th} ed., Oxford University Press, New York, 2007.
[3] M. Balkanski, P.Laleman, "Photonics", Mir, Moscow, 1978
[4] W. T. Tsang, Volume Editor, Part D, Photodetectors, in Vol. 22, Lightwave Communications Technology, R. K. Willardson and A. C. Beer, Eds., Semiconductors and Semimetals, Academic Press, New York, 1985.
[5] B. E. A. Saleh and M. C. Teich, Fundamentals of Photonics, 2nd ed., Wiley, New York, 2007.
[6] K. Lehovec, C. A. Accardo, and E. Jamgochian, "Injected Light Emission of Silicon Carbide Crystals", Phys. Rev. 83, 603, Aug. 1951.
[7] M. G. Craford, N. Holonyak, Jr., and F. A. Kish, "In pursue of the ultimate lamp," Sei. Am. 284, 63-67 (2001).
[8] K. Iga, Fundamentals of Laser Optics, Plenum, New York, 19
[9] A. Schawlow and C. H. Townes, "Infrared and optical masers," Phys. Rev. 112, 1940-1949(1958).
[10] J. Hecht, "Winning the laser patent war," Laser Focus World, 49-51, December 1994.
[11] T. H. Maiman, "Stimulated optical radiation in ruby," Nature 187, 493-494 (1960).
[12] R. N. Hall, G. E. Fenner, J. D. Kingsley, T. J. Soltys, and R. O. Carlson, "Coherent light emission from GaAs junctions," Phys. Rev. Lett. 9, 366-368 (1962)
[13] M. I. Nathan, W. P. Dumke, G. Burns, F. H. Dill, Jr., and G. Lasher, "Simulated emission of radiation from GaAs p-n junctions," Appl. Phys. Lett. 1, 62-64 (1962).
[14] R. D. Dupuis, "The diode laser—the first 30 days, 40 years ago," Optics and Photonics News, April, 30-35 (2004).
[15] J. C. Palais, "Fiber Optic Communication Systems", Springer-Verlag, 2005.
[16] C-F. Lin, "Optical Components for Communications: Principles and Applications", Springer-Verlag, 2004.
[17] Eissler, D., S. Illek, and G. Bogner. "LED technology trends." The International Conference on Compound Semiconductor Manufacturing Technology. Vol. 4. 2005.
[18] A. Rogalski, Ed., Selected Papers on Semiconductor Infrared Detectors, SPIE Milestone Series, Vol. MS66, SPIE Optical Engineering Press, Bellingham, WA, 1992.
[19] J. D. Vincent, Fundamentals of Infrared Detector Operation and Testing, Wiley, New York, 1990.
[20] T. P. Lee and T. Li, "Photodetectors," pp. 593-626 in S. E. Miller and A. G. Chynoweth, Eds., Optical Fiber Telecommunications, Academic, New York, 1979.
[21] S. M. Sze and K. K. Ng, Physics of Semiconductor Devices, Wiley, New York, 2007.
[22] G. Lasher and F. Stern, "Spontaneous and stimulated recombination radiation in semiconductors," Phys. Rev. 133, A553-A563 (1964).
[23] A. Rizzi, Private communications, 2001.
[24] N. Holonyak, Jr., "From transistors to lasers and light-emitting diodes," Mater. Res. Bull. 30, 509-517 (2005).
[25] R. D. Dupuis and M. R. Krames, "History, development, and applications of high brightness visible light-emitting diodes," J. Lightwave Technol. 26, 1154-1171 (2008).
[26] D. R. Tribble, "Seven-segment display digits of an original Texas Instruments TI-30 scientific calculator (circa 1979)." Wikimedia Commons, Mar. 2012.

[27] Nakamura, Shuji, and Gerhard Fasol. The blue laser diode: GaN based light emitters and lasers. Springer Science & Business Media, 2013.
[28] J. E. Bowers and C. A. Burrus, Jr., "Ultrawide-band long-wavelength p-i-n photodetectors," J. Lightwave Technol. LT-5, 1339-1350 (1987).
[29] G. E. Stillman and C. M. Wolfe, "Avalanche photodiodes," pp. 291-393, Infrared Detectors II, Vol. 12 in R. K. Willardson and A. C. Beer, Eds., Semiconductors and Semimetals, Academic Press, New York, 1977.
[30] T. P. Lee and T. Li, "Photodetectors," pp. 593-626 in S. E. Miller and A. G. Chynoweth, Eds., Optical Fiber Telecommunications, Academic, New York, 1979.
[31] G. B. Stringfellow, International Review of Science, Inorganic Chemistry, Series 2, Vol. 10 (Butterworths, London, 1975), pp. 111-145.
[32] G. B. Stringfellow, Gerald B. *Organometallic vapor-phase epitaxy: theory and practice*. Elsevier, 1999.
[33] Vossen, John L., Werner Kern, and Walter Kern, eds. *Thin film processes II*. Vol. 2. Gulf Professional Publishing, 1991.

Chapter 2: Tutorials on Optics and Optoelectronics

Sean Hooten

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

2.0 INTRODUCTION

Light is prevalent in many aspects of our everyday lives. Many tools and devices explicitly use light for operation such as mirrors, digital displays, laser pointers, glasses, cameras, polarizers, and much more. Even the internet requires light for communications (via light wires or "optical fibers"). The general term and scientific field referring to the study of the behavior of light and its manipulation using physical objects is called **optics.** Using optics and optical devices, the properties of light can be measured, altered, or generated according to a desired application. When the optical properties of objects are controlled using electrical circuits it is referred to as **optoelectronics**. By advancing optics and optoelectronics, new technologies can be developed for use in our everyday world and scientific endeavors can be expanded.

In this chapter, various optics and optical devices that are common in science and industry applications will be explored. In doing so, we will learn first-hand about how light can be physically manipulated, and about exciting emerging technologies in the field of optics and optoelectronics. The intention of doing this will be to eventually motivate the fundamental physics and applications of the ultra-fast light sources known as Antenna-LEDs, which are the main subject of the E3S Theme III research. This chapter will assume a basic understanding of energy conservation, wave mechanics, simple principles of optics (e.g. white light consists of many colors), and electrical circuits. Some knowledge of semiconductor material properties would also be useful but is not required. In Section 2.1 the most common ways that light is generated in matter will be discussed and then we will go on to describe the fundamentals of LEDs and laser physics. In Section 2.2 we will describe how light can be guided, or transported, by using optical fibers and waveguides. In Section 2.3, we will show the common devices used in the detection and absorption of light. Finally, in Section 2.4 we will introduce the concept of an electromagnetic simulation.

Key Terms:
Optics
Spontaneous emission
Stimulated emission
Photon
Absorption
Blackbody radiation
Valence band
Conduction band
Hole
Optical pumping
Electrical injection
P-N junction
Diode
Quantum efficiency
Recombination
Optical cavity
Gain
Gain threshold
Threshold current
Wall-plug efficiency
Optical fiber
Total internal reflection
Snell's Law
Refractive index
Waveguide
Electromagnetic simulation

We are familiar with many different light sources. The sun is probably the most obvious, but there is also starlight in the night sky, fire, light bulbs of various types, lasers, liquid crystal displays

(LCDs), and more such as bioluminescence in the animal kingdom (pictured below in Fig. 2.1). What distinguishes different light sources? There are probably many characteristics that are obvious to the reader, for example some light sources have different colors, some are bright, some flicker and change, and some seem to shine in only one direction (lasers specifically). In this section we will explore what is behind these differences and then delve into the fundamental physics of lasers and LEDs which are the subject of most of the ongoing research in E3S Theme III.

Figure 2.1. Bioluminescent jellyfish.

Fundamentally, there are two main processes for the generation of light in matter: **spontaneous emission** and **stimulated emission**. Spontaneous emission is the most common process and is responsible for almost all light that we witness in our everyday lives including starlight (including the sun), all types of lightbulbs, TVs and laptop monitors, fire, the red glow of molten metal, and more. Stimulated emission, on the other hand, is the major mechanism for light emission from lasers. A diagram explaining the fundamental emission and absorption processes is given in Figure 2.2.

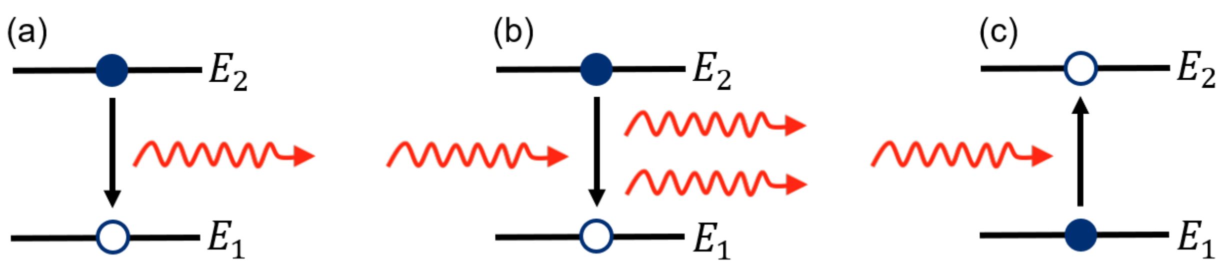

Figure 2.2. Common light-matter interactions including (a) spontaneous emission, (b) stimulated emission, and (c) absorption. Red squiggly lines indicate photons, blue filled circles represent electron occupied energy states, and white empty circles represent empty energy states.

Each diagram has the same basic elements: the black lines represent energy levels ($E_2 > E_1$), the red squiggly lines depict **photons** (particles of light), the black arrows show electron energy level transitions, blue filled circles are states occupied by electrons, and empty circles are empty states at a specific energy level. In general, absorption and emission can only occur when occupied and corresponding empty states are available according to conservation of energy and momentum.

Fig. 2.2(a) shows the process of spontaneous emission. Spontaneous emission occurs when an excited electron *spontaneously* falls to an empty state at a lower energy level. Due to conservation of energy, the change of the electron energy creates a photon with energy $E_{photon} = E_2 - E_1$. This is in contrast to Fig. 2.2(b) which shows stimulated emission. For stimulated emission to occur, the excited electron state only falls to the empty state after it has been *stimulated* by an external photon (incident from the left in this diagram). Once again, due to conservation of energy an additional photon is created, so two photons leave to the right. Stimulated emission is considered inverse to the process of **absorption** which is given in Fig. 2.2(c). In this case, the electron begins in the lower energy state. When a photon is incident on the system with the correct energy $E_{photon} = E_2 - E_1$, the electron is excited to the upper energy state. Due to conservation of energy, the photon is absorbed in the process and no photons exit the system.

Each of these processes shown in Fig. 2.2 represents a single light emission or absorption event which could be occurring in a single atom or molecule. In macroscopic reality, *many* of these events are happening simultaneously amongst trillions and trillions of molecules. This is why lightwaves (from a lightbulb for example) appear to be continuous and do not appear to take on the discrete nature of these individual events. Furthermore, note that Figure 2.2 depicts a system that consists of only two energy levels (also known as a two-level system). Most macroscopic objects consist of systems of molecules with many energy levels, all of which can participate in the above absorption and emission processes. This is why the sun emits white light, consisting of many different energies (colors). Blue light is high energy, while red light is low energy.

Perhaps the reader may be wondering, why do physical objects have color—are they emitting light? Most of the visual light and colors that can be seen on physical objects (that are not light sources) is actually due to absorption of some colors in white light and selective reflection of other colors. The absorbed light is eventually reradiated by the material in the form of spontaneous emission since excited states are generated by absorption, but this usually occurs at energies that are too small for our eyes to perceive (most of the energy from absorption is converted to heat). This is why sitting next to a lightbulb will cause your body to warm up, and you don't begin to glow like a lightbulb yourself (though by using a thermal or infrared camera, radiation from your body can be seen).

Another way that excited states can be filled, as alluded to above, is through increasing the temperature of the system (e.g. for an incandescent lightbulb, molten metal, or the sun). This type of emission is known as **blackbody radiation**. As the temperature of an object is increased, higher energy states become occupied and spontaneously emit photons. If the object is hot enough, visual radiation can be observed.

The final way that will be discussed here, and the one of most interest for the investigation of LEDs and lasers, is through optical and electrical injection of semiconductors. The following subsections will explain the working principles of LEDs and lasers in more detail.

2.1.1 SEMICONDUCTOR MATERIAL PROPERTIES AND LIGHT EMISSION

Before discussing LEDs and lasers, the electronic and optical characteristics of semiconductors will be introduced. A semiconductor is a material that contains an energy bandgap, or a gap in which energy levels and hence electron occupied states are not allowed. To describe semiconductors, E-k diagrams are typically used, an example of which is shown in Fig. 2.3(a).

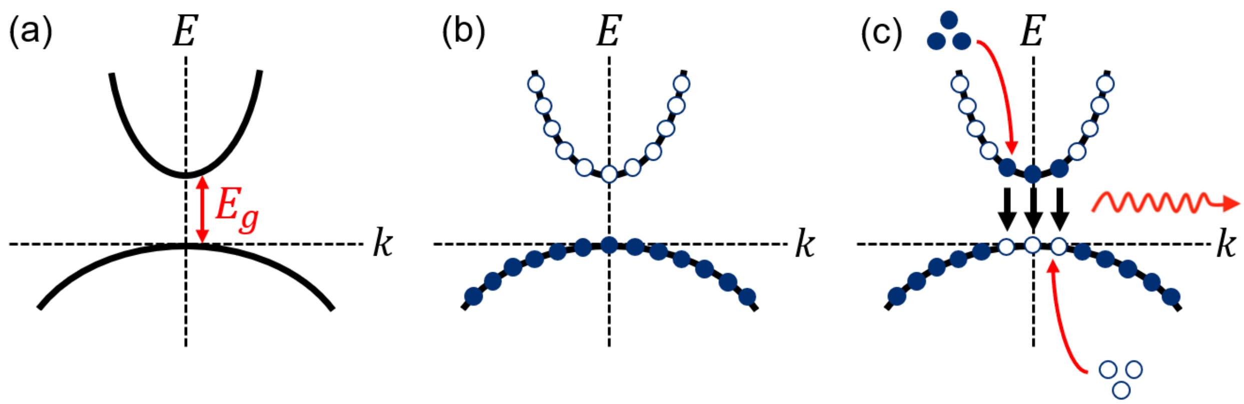

Figure 2.3. E-k diagram of a simple semiconductor. (a) Diagram shows which energy states are allowed to be occupied (indicated by black lines), as well as the forbidden energy bandgap. (b) Typical occupation of a semiconductor at low temperature. Electrons occupy the low energy levels (known as the valence band) while the excited energy levels (known as the conduction band) are empty. (c) By pumping electrons into higher energy states and holes into lower energy states, light emission can occur by spontaneous or stimulated emission.

Where E represents energy, k represents a quantity known as the wavevector, E_g is the forbidden energy bandgap, and black lines represent energy states that are allowed to be occupied. The line that curves downwards is known as the **valence band** and the line that curves upwards is known as the **conduction band**. The significance of the wavevector is not important for this chapter, but it is related to the momentum of electrons and holes in the semiconductor crystal. For the purposes of this chapter, a **hole** can be thought of as "empty" states in the valence band with a charge that is equal in magnitude to an electron but opposite in polarity.

The occupation of a typical semiconductor at low temperature (e.g. room temperature), is given in Fig. 2.3(b). As in Fig. 2.2, filled blue circles represent occupied electron states, and empty circles represent empty states. Since no (in reality, very few) electron states are occupied at higher energy, there is almost no radiative emission. However, as depicted in Fig. 2.3(c), if the semiconductor is pumped by injecting electrons into the higher energy levels and holes into the lower energy levels, radiative emission can be achieved by spontaneous emission.

Notice that spontaneous emission occurs across the bandgap in Fig. 2.3(c). This is an important property of semiconductors: the energy of radiative emission is given by the bandgap energy. By using different semiconductor materials with different bandgaps, the energy (color) of light emission can be tuned as desired. Current smartphone displays exploit this by using red, blue, and green colored microscale organic LEDs for individual pixels. By combining the light from these LEDs to varying degrees, any desired output color pixel can be achieved.

Pumping of a semiconductor, as in Fig. 2.3(c), can be achieved by **optical pumping** or **electrical injection.** Optical pumping simply refers to causing many absorption events in the material, usually by shining a laser on it with photon energy that is greater than the bandgap energy. Sometimes this type of light emission from a semiconductor is called an optically-pumped LED, though technically an LED requires a P-N junction which will be described in the next subsection.

2.1.2 LIGHT-EMITTING DIODE (LED)

Electrical injection occurs when electrons and holes are brought into the semiconductor emitter via an electrical current. Typically, this is done using a **P-N junction** which is a type of **diode**, or a device that tends to conduct current mainly in one direction. An example of a semiconductor P-N junction light-emitting diode is given in Fig. 2.4. A P-N junction consists of a semiconductor that is asymmetrically doped. Essentially this means that on one side the semiconductor has excess excited electrons (n-type side) and on the other it has excess holes (p-type side). When a positive voltage is applied to the P-N junction, excess electrons from the n-doped side flow to the p-type side and excess holes from the p-type side flow to the n-type side. Thus, current flows from p-type to n-type. At the junction between n- and p-type sides, there is both excess electrons and holes which allows for radiative light emission according to spontaneous emission, as shown previously in Fig. 2.3(c).

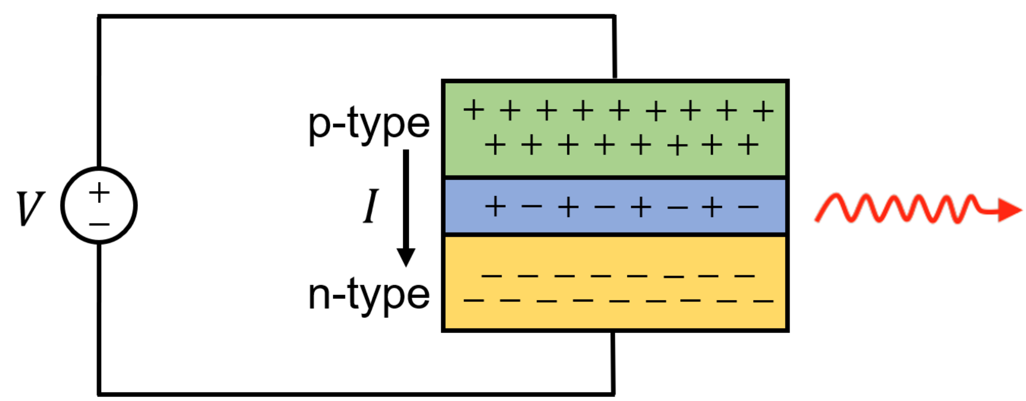

Figure 2.4. A simple light-emitting diode (LED), consisting of a P-N junction. When a positive voltage is applied (forward bias), light radiates out via spontaneous emission.

The amount of optical power that leaves the LED is typically represented by an L-I curve, depicted in Fig. 2.5. In this case, L represents optical light intensity in terms of the number of photons emitted per unit time and I represents electrical current flowing through the diode. Depending on the **quantum efficiency** of the LED, more photons will exit the device per unit time given a specific electrical current. The quantum efficiency of the LED is largely determined by the ratio of radiative recombination events to total recombination events, where **recombination** refers to the general process when excited electron states fall into empty hole states, similar to that shown in Fig. 2.2(a)-(b). However, unlike Fig. 2.2(a)-(b), not all recombination events will result in light emission (when this occurs it is called non-radiative recombination). Non-radiative recombination is an inelastic process where heat is generated in the device, which is undesirable but inevitable in practice. The quantum efficiency can be improved by using high-quality materials and through careful engineering of the P-N junction. It can also be improved through spontaneous emission enhancement, which is the subject of E3S Theme III research and will be discussed more thoroughly in Chapter 6.

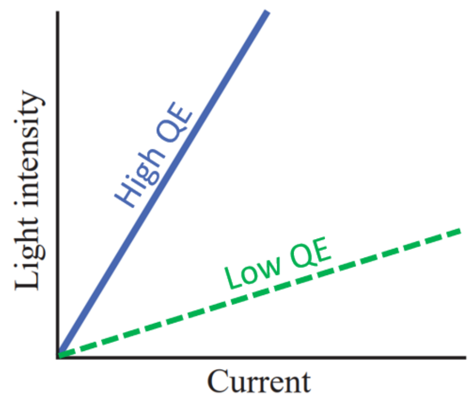

Figure 2.5. L-I curve of an LED where QE refers to quantum efficiency. Printed with permission from [1].

2.1.3 LASER

The term laser is actually an acronym that means "light amplification by stimulated emission of radiation". Hence, a laser is any device that operates on the principle of stimulated emission as shown in Figure 2.2(b). Most lasers that are manufactured in the modern day consist of semiconductors and a P-N junction, but there are actually many types such as atomic lasers which consist of a material such as helium-neon (HeNe) and no P-N junction. A laser consists of a material with optical **gain** that resides in an **optical cavity**. This is depicted in Fig. 2.6.

Fig. 2.6(a) shows a generalized laser consisting of a gain material between two (imperfect) mirrors. Light is contained in the optical cavity by bouncing between the two mirrors, and it passes through the gain material. Some amount of light escapes from out of each mirror, which is what we typically observe as laser light. Figure 2.6(b) shows a realistic manifestation of a semiconductor laser consisting of a semiconductor gain material with facets cleaved at each end. The cleaved facets act as imperfect mirrors, containing the mode propagating through the semiconductor waveguide (waveguides will be discussed in the next section).

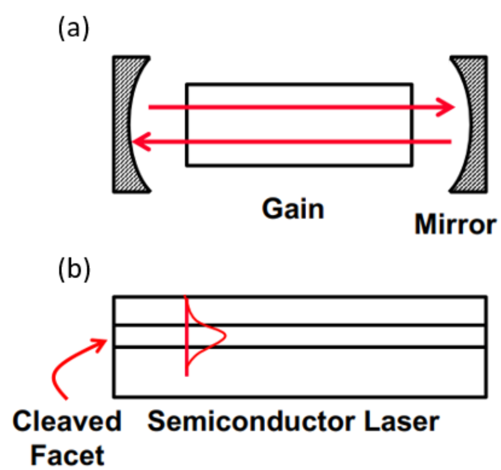

Figure 2.6. (a) A generalized laser, consisting of a gain material and an optical cavity. Some light leaks out of the mirrors, which is what we call laser light. (b) A semiconductor laser, where an optical cavity is formed by cleaving off two ends of the semiconductor. Printed with permission from [2].

As light propagates in the optical cavity, it interacts with the gain material which amplifies the optical signal via stimulated emission. Gain is determined by the degree to which the material has been pumped, optically or electrically. To understand this, refer back to Figure 2.2(b)-(c) which shows the processes of stimulated emission and absorption. In the case of absorption, there are more electrons in lower energy states than excited energy states. In the case of stimulated emission, there are more electrons in excited energy states than lower energy states. The latter situation is called a population inversion. Hence, gain can be thought of as “negative absorption” and is given by the amount which a population of electrons has been inverted. In a real laser, absorption and stimulated emission are always occurring simultaneously, but if a population inversion has been created than statistically more stimulated emission events will occur than absorption events, leading to light amplification.

However, this isn’t the end of the story, because lasing can only occur when the amount of light amplification equals the loss of photons transmitted through the mirrors and through other non-ideal absorption mechanisms. This condition is called the **gain threshold** and occurs when the gain matches all other forms of loss in the optical cavity.

In a semiconductor laser, generally the device will be pumped via electrical injection to reach gain threshold. The amount of current required to reach gain threshold is called the **threshold current**. Before the threshold current has been reached, the device will mostly emit light in the form of spontaneous emission, which tends to be small. Hence, the L-I curve for the laser appears as the following in Fig. 2.7:

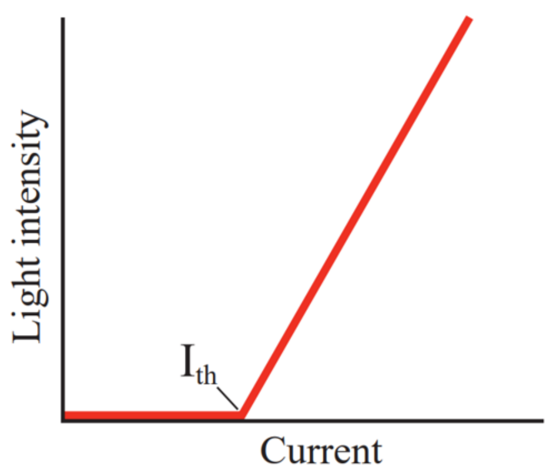

Figure 2.7. L-I curve for a laser. Light intensity as a function of current is small before threshold has been reached. Printed with permission from [1].

As can be seen here, the optical power of the laser increases rapidly after the threshold current has been reached. It can be said that the quantum efficiency of a laser is very high after threshold. However, the **wall-plug efficiency** of the device takes the threshold current into account, and is the more important metric to determine the energy efficiency of a laser. When a laser has a low threshold current it tends to have better wall-plug efficiency, however in practice the threshold current is always finite. This is in contrast with the LED from the previous section, where we showed that light is emitted for any amount of current that is injected in the device. Thus, if the quantum efficiency of an LED can be increased to unity, it can be more efficient than a laser because it requires no threshold condition to be reached.

2.2 GUIDING LIGHT

We are most commonly familiar with light traveling freely through the air. For instance, when a lightbulb is turned on, its light will propagate in all directions until it is absorbed, reflected, or transmitted by a physical object. The same concept applies to laser light, where the light tends to propagate freely through the air but it travels in a straight line until interacting with a physical object. Can the direction of light be manipulated, guiding it from a source to where one desires? In this section, two particularly useful tools that accomplish this functionality will be introduced: the optical fiber and the waveguide. More detail will be presented on this subject in Chapter 5, where the integration of an optical source with a waveguide will be discussed.

2.2.1 OPTICAL FIBER

An **optical fiber** is essentially a thin deformable fiber composed of glass or plastic. Fig. 2.8 depicts a diagram of what a simple optical fiber might look like.

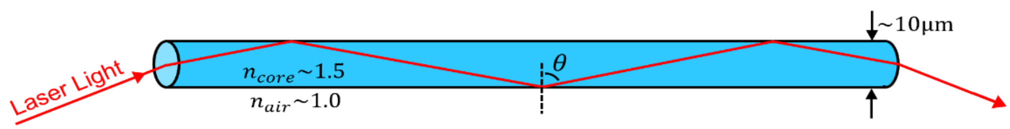

Figure 2.8. Simple optical fiber consisting of a core with refractive index of 1.5 and no cladding. Laser light is incident on the left and is internally reflected until it exits at the right.

As can be seen here, laser light is injected at one end of the fiber and the light subsequently bounces around without leaving the fiber until it exits at the other end. The diameter of an optical fiber core is typically $10\mu m$, though it should be noted that practical optical fibers contain many layers of cladding such that the fiber is at least 0.5mm in diameter. For an optical fiber to work correctly without leaking, light must be injected into the fiber within a specific angle for **total internal reflection** to occur. Total internal reflection is a consequence of **Snell's Law**, which is written here:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

Without going into too much detail, Snell's Law states that if a ray of light is incident on the interface between two materials with **refractive index** n_1 and n_2 respectively, then the angle which the light is transmitted through the interface, θ_2, is related to the angle of incidence, θ_1, by the above equation. In other words, light bends when it enters a new material which can easily be observed by submerging objects in water or by using glasses. An interesting consequence of Snell's Law is that if light is incident on the interface from a material with $n_1 > n_2$ then it is possible for no light to be transmitted through the interface, and all light will be reflected. This (total internal reflection) condition occurs when,

$$\sin \theta > \frac{n_2}{n_1}$$

Thus, we see in the above Fig. 2.8, the angle θ must be $\sin\theta > 1.0/1.5$, or, $\theta > 41.8°$. For the same reason, the optical fiber cannot be bent considerably, or else the angle of incidence on the fiber wall could become too small and light will leak out.

Using optical fibers, light can be guided to great distances with high efficiency and information density. In fact, all modern telecommunications rely on optical fibers for communication across the world, including across the Pacific and Atlantic Oceans. More recently, optical fibers have been introduced for shorter distances to provide internet and networking for individual homes and between computing servers in large data centers.

2.2.2 WAVEGUIDE

Currently, most communication on a computer chip is done using electrical wires. This is because there are processes that make it possible for very dense electrical wiring to be fabricated on an integrated circuit, allowing for many billions of transistors to be interconnected. However, electrical wires are inherently lossy due to parasitic capacitive charging. Consequently, the high-performance computing industry is moving towards using optical wires on chip, which could be more energy efficient at approximately the millimeter length scale or smaller.

The technical term for an integrated optical wire is called a **waveguide**. A waveguide is very similar to an optical fiber, except that its size is close to or even smaller than the wavelength of light that is propagating through it. An example cross-section of a waveguide is provided here in Fig. 2.9:

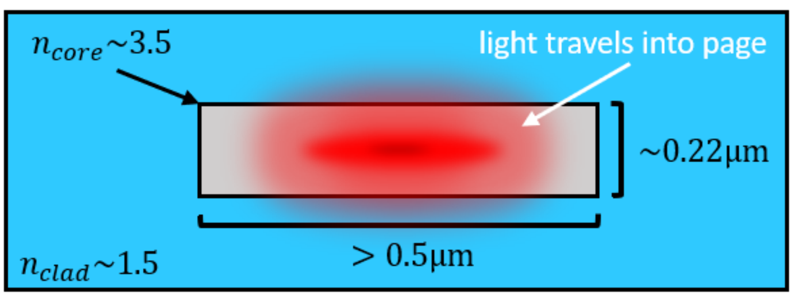

Figure 2.9. The cross-section of an optical waveguide consisting of a core with refractive index of 3.5 (similar to silicon), clad in a material with refractive index of 1.5 (similar to glass). Light propagates into the page, contained in the core.

This waveguide consists of a rectangular layer surrounded by cladding in all directions. Typically, the waveguide core consists of silicon or a III-V semiconductor such as InP, and the cladding consists of glass. Note that the refractive indices used in waveguides are often much larger than those in optical fibers from Fig. 2.8 above. This allows for strong optical confinement of the waveguide mode, which is pictorially illustrated as the red dot in Fig. 2.9 which propagates into the page. Also note that the dimensions of the waveguide are far smaller than that of the optical fiber from Fig. 2.8.

Due to their small scale and capability for microfabrication, waveguides are well-suited to serve as efficient wiring in an integrated circuit (also known as a photonic integrated circuit). The ultimate goal of E3S Theme III is to propose a viable optical source that can be integrated with a waveguide to provide efficient on-chip optical communications. More details on waveguide integration will be discussed in Chapter 5.

2.3 DETECTING/ABSORBING LIGHT

All materials absorb electromagnetic waves with at least some range of energies. This was shown diagrammatically in Fig. 2.2(c). When light is absorbed, it excites electrons to higher energy states. This can lead to chemical changes in the material, temperature rise, re-radiation of energy, or increased conductivity. Thus, absorption can be used for many different applications, which will be discussed in this section.

2.3.1 CAMERAS

Cameras are devices that can produce a photograph through the sampling (e.g. absorbing) of the color spectrum and intensity of incident light at various spatial locations typically along a 2D plane. Cameras can be engineered to be sensitive to different energies of incident electromagnetic waves (e.g. infrared, visible light, or ultraviolet cameras). In this case we will only consider typical cameras that photograph visual light of which there are two main types: film and digital.

Film cameras use a strip of photographic film that is chemically sensitive to the incident intensity of light. More intensity of light will cause more chemical changes as a result of absorption. Color film generally consists of three layers that are sensitive to red, green, and blue light respectively. After exposure and development, the layers will have produced dyes that are complementary to the colors of light that shined on them, thus producing a film negative of the image. This negative can then be used to create color photos.

Digital cameras use a 2D array of microscale physical electronic devices known as charge-coupled devices (CCDs). CCDs are essentially capacitors that can be charged in the presence of an incident flux of light, and then subsequently read out in a digital format using an integrated circuit. By placing colored filters above the array (pictured in Fig. 2.10), each CCD can be made sensitive to a different color of light. Thus, during digital processing of the CCD readout, the individual colors incident at each pixel of the image can be deciphered and recorded. Since the CCDs are very small due to microfabrication processing capabilities, extremely high-resolution digital images can be recorded. Furthermore, since the CCD array sensor is electronic, many images can be photographed and stored almost instantly without any film development required.

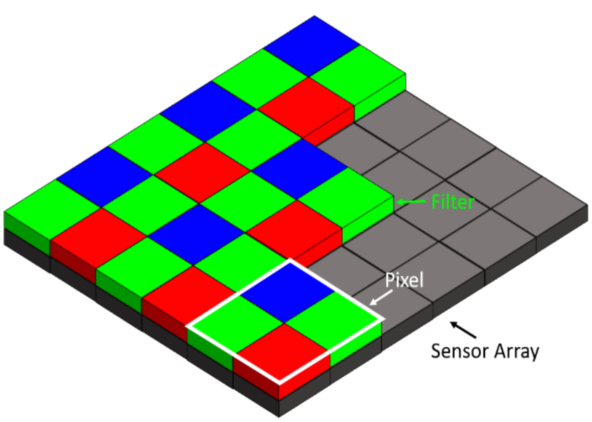

Figure 2.10. A Bayer filter over a CCD sensor array. Each pixel consists of 2 green filters, 1 blue filter, and 1 red filter. There are more green filters, because the human eye is most sensitive to the color green, and hence more importance is placed on green light. By combining and interpolating information from the sensors under each filter, the color and intensity of incident light can be digitally read to great precision.

2.3.2 PHOTOVOLTAICS (SOLAR CELLS)

Photovoltaic (PV) cells are devices that can be used to harvest electrical energy from light sources such as the sun. In its simplest form, a PV cell is a P-N junction (Fig. 2.4), except in this case light is incident on the cell, as depicted in Fig 2.11 below.

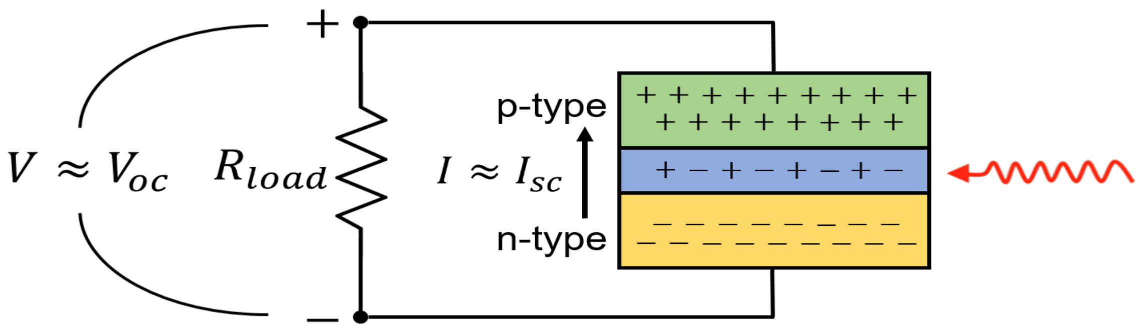

Figure 2.11. Simple photovoltaic cell consisting of a P-N junction and a resistive load. Light that is incident on the P-N junction produces current and voltage given approximately by I_{sc} and V_{oc} respectively.

Note that current now flows in the opposite direction, and is close to a quantity known as the short circuit current, I_{sc}, which is almost equivalent to the photogenerated current generated by the incident light being absorbed in the P-N junction. This induces a potential called the open circuit voltage, V_{oc}, which can be dropped across a resistor load R_{load}. Essentially, the PV cell acts as an engine with output generated electrical power given by,

$$P = IV \approx I_{sc}V_{oc}$$

that can be used to power any electrical device such as an electrical grid, a vehicle, or a hand-held calculator.

2.3.3 PHOTODETECTORS

Photodetectors are devices that are similar to photovoltaics except they are not designed for energy harvesting; instead they are used for accurately measuring the intensity of incident light in a time-correlated manner. Typically, they are tuned to absorb light most efficiently at a small number of photon energies, and they are much smaller in size such that they can be integrated on a photonic chip. Consequently, they can be used for optical communications. The development of photodetectors that are nanoscale and low capacitance is pivotal for the success of efficient and fast on-chip optical interconnects that are being researched in E3S Theme III.

2.3.4 MODULATORS

Optical modulators are devices that can be dynamically switched to be transparent or absorbing of incident light using an electrical input signal. In the context of photonics and on-chip optical interconnects, a modulator is a device that would be integrated in the path of a waveguide. Thus, by changing the absorbance of the modulator dynamically, the input light can be *modulated* to encode a digital signal which can be sent to another part of the chip. Modulators can be very fast, and hence are an attractive option for optical communications.

On-chip optical modulators are typically semiconductors with tunable effective bandgap energies (though other types exist). A simplified version of this concept is depicted in Fig. 2.12. Incident photons of a specific energy can be either absorbed as in Fig. 2.12(a) or pass through as in Fig. 2.12(b) based on the effective bandgap energy. The effective bandgap energy can be altered using an electrical bias.

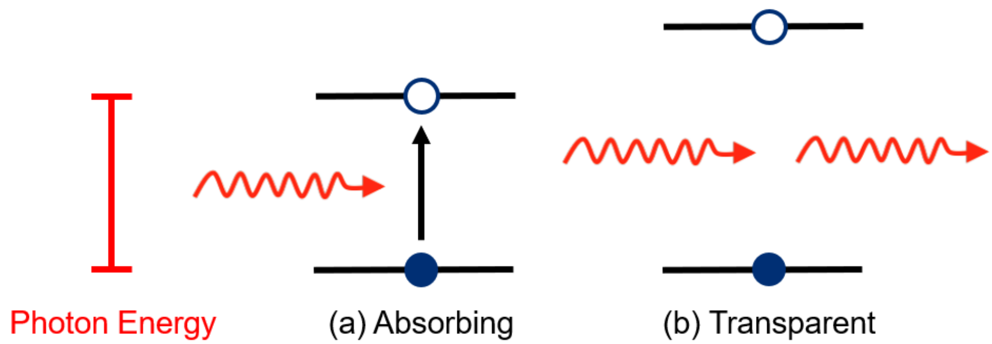

Figure 2.12. Basic principle of an electroabsorption modulator. The bandgap energy of the modulator can be tuned to be either (a) absorbing or (b) transparent to incoming photons with a specific energy.

While (intensity) modulators can be switched rapidly, they are inherently inefficient because the energy of absorbed light is lost to heat. Ideally, the optical signal in a photonic circuit would be directly modulated using the light source (by turning it on and off, for example). However, there are fundamental limits on the speed by which optical sources can be modulated; in particular, LEDs are naturally very slow. However, by exploiting spontaneous emission enhancement LEDs can be made very fast, making them competitive with lasers and modulators in terms of both speed and efficiency. This will be discussed thoroughly in the ensuing chapters.

2.4 ELECTROMAGNETIC SIMULATIONS

As a final topic in this chapter, we will briefly discuss **electromagnetic simulations**. A simulation is essentially an imitation of a physical system or process, which usually involves the numerical calculation of a system of equations. In particular, in optics and optoelectronics we are interested in solving Maxwell's Equations:

$$\nabla \cdot D = \rho$$
$$\nabla \cdot B = 0$$
$$\nabla \times E = -\frac{dB}{dt}$$
$$\nabla \times H = J + \frac{dD}{dt}$$

A detailed discussion of Maxwell's Equations is out of the scope of this chapter. But essentially, these four equations describe all classical electromagnetic phenomena in the universe. Each equation consists of electric fields, magnetic fields, material values, and sources defined for any position and polarization.

It is easier to understand simulations by example; Fig. 2.13 shows an example of the simulation of a waveguide splitter.

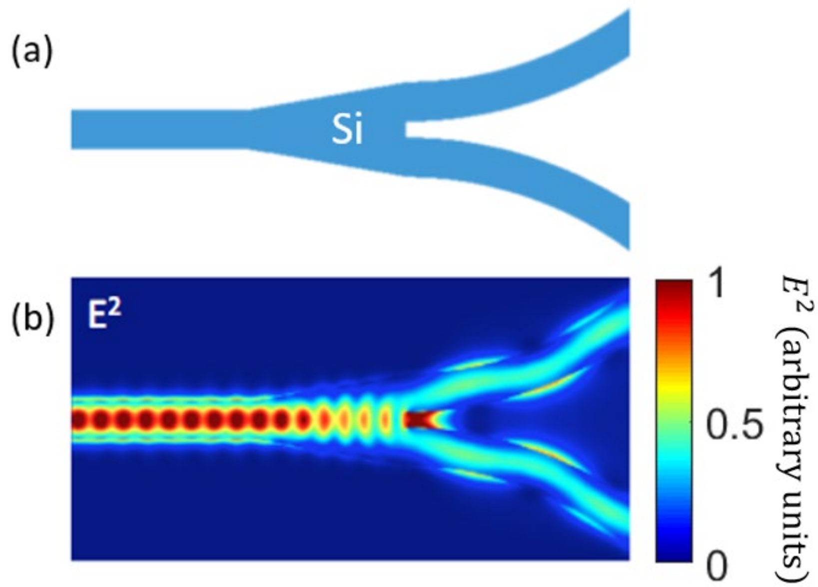

Figure 2.13. Waveguide splitter simulation. (a) the splitter consists of an input silicon waveguide, a taper splitter, and two output silicon waveguides. (b) Electric-field intensity of the splitter after simulation, where red indicates high intensity and blue indicates low intensity. The input light intensity splits equally between the two output branches.

Figure 2.13(a) depicts the geometry of the splitter. It consists of a silicon waveguide that is split into two output waveguides using a short taper. Figure 2.13(b) shows the result of the electromagnetic simulation after injecting light in the fundamental resonance of the waveguide from the left side. Specifically, it shows the intensity of the electric field in the device, where blue represents small intensity and red represents large intensity. Intuitively, we find that the intensity of the electric field in the waveguide input is larger than the two outputs, since the incident power is approximately evenly split between them.

Using the information from this simulation, we can extract quantities such as the splitting efficiency of the waveguide splitter, and the amount of power that gets reflected back to the input. It would be very difficult to calculate such quantities analytically by hand, so electromagnetic simulations are a powerful tool to assist in the design of optical devices. Simulations are heavily relied upon in E3S Theme III research, and results from such simulations will be presented in the following chapters. In Chapter 6, additional detail about computational electromagnetics will be discussed.

REFERENCES

[1] S. Fortuna, "Integrated Nanoscale Antenna-LED for On-Chip Optical Communication," Dissertation.
[2] M. Wu, UC Berkeley EE232 Course Notes, http://www-inst.eecs.berkeley.edu/~ee232/sp18/.
[3] S. L. Chuang, *Physics of Photonic Devices*. John Wiley & Sons, 2012, vol. 80.

Chapter 3: Antenna LED

Seth Fortuna

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

3.0 INTRODUCTION

In this chapter the basic concept of the antenna-LED will be introduced. The reader is expected to have read the previous chapters in this e-book. Some knowledge on basic atomic physics (high-school level) will be helpful.

3.1 LIGHT EMITTING DIODE AS A COMMUNICATION DEVICE

As described in Chapter 0, there is a recent trend toward miniaturizing optical communication links so they are on the same size-scale (< 1 cm in length) of an integrated circuit. By doing this, it will be possible to replace traditional on-chip metal wiring with optical links and reduce the power consumption on the integrated circuit. Traditional (> 1 m in length) optical communication links contain a semiconductor laser diode which is used to encode digital data on a light-wave (Figure 3.1). However, laser diodes have a few non-ideal properties. As discussed in the previous chapter, a minimum amount of energy needs to be supplied for the device to work (i.e. the laser threshold condition). This energy is not converted into usable light and is wasted as heat. Secondly, it is difficult to decrease the size of the laser much below the volume λ^3 where λ is the wavelength of light which the laser emits ($\lambda \sim 1\ \mu m$). This volume is dictated by the so-called "diffraction limit" which places an upper limit on the volume in which you can confine a light wave. There are methods to get around the diffraction limit, but they tend to introduce additional inefficiencies which reduces the amount of usable light and increases wasted heat.

Semiconductor-based light emitting diodes (LEDs) are inherently efficient devices. This is the primary reason that LEDs have started to replace traditional incandescent and fluorescent lighting (LED lighting is sometimes also called solid-state lighting). Unlike the laser, there is no equivalent threshold condition for the LED to operate. Furthermore, LEDs are not bound by the diffraction-limit and can therefore be easily scaled to extremely small volumes ($V \ll \lambda^3$).

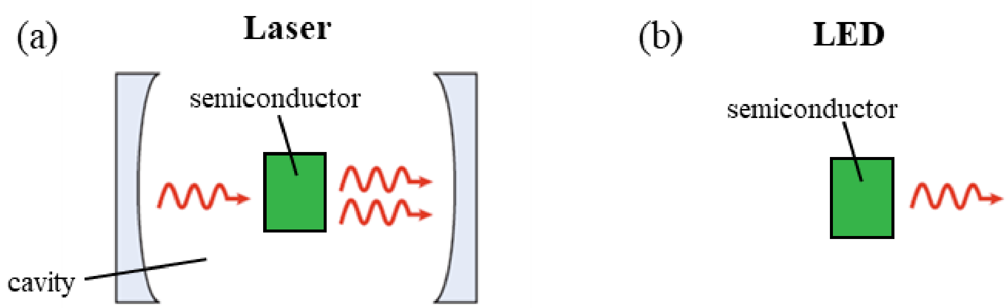

Figure 3.4. Simplified drawing of (a) Laser; and (b) light emitting diode LED. Both consist of semiconductor light emitting materials. The laser has a cavity whereas the LED does not.

Unfortunately, LEDs are not fast emitters of light. One way to observe this behavior is to turn on an LED and then rapidly turn it off by quickly applying and then removing a voltage across the LED. We will observe that the emitted light does not rapidly turn on and off: it takes about a nanosecond or so for the LED to start and stop emitting light. If we were to use an LED in an optical communication link, the amount of digital data per unit time (bits per second) which we can transmit in this optical link is given roughly by the inverse of this turn-on/off time. This turns out to be about 100 megabits per second or 10^8 bits per second. This seems like a lot of data, but integrated circuits already communicate with about 50 times higher data rate (and sometimes even higher) using existing electrical wiring on chip. The laser can operate at data rates that are 100 to 500 times faster than the LED. So, despite other benefits, the basic LED is usually overlooked as a light emitter for an on-chip optical communication link.

3.2 WHY IS AN LED SLOW?

Light is emitted from an LED by a process called spontaneous emission. The basic process is shown in Figure 3.2. We consider a simple system consisting of two possible electron states. The electron in this system can reside in either the ground state or excited state. Light emission occurs when an electron residing in the excited state spontaneously jumps (or decays) into the ground state. Conservation of energy requires that the electron must give up a photon with energy that is equal to the energy separation between the excited and ground state. The spontaneous emission lifetime is the average amount of time it takes for the electron to decay into the ground state; this is typically a few nanoseconds.

Contrast this with the stimulated emission process shown in Figure 3.2b. In this case, an external photon stimulates the transition from the excited state to the ground state. As before, a photon is emitted during the transition. This is the mechanism by which a laser emits light. As you can imagine, the stimulated transition is more likely to occur if there is more than one external photon present. Therefore, if we wish to speed up the stimulated emission process, we can simply confine more light in our laser. The stimulated emission lifetime can be as fast as a few picoseconds; much faster than stimulated emission. We can also equivalently say that the spontaneous emission rate is smaller than the stimulated emission rate where the emission rates are the inverse of the emission lifetimes.

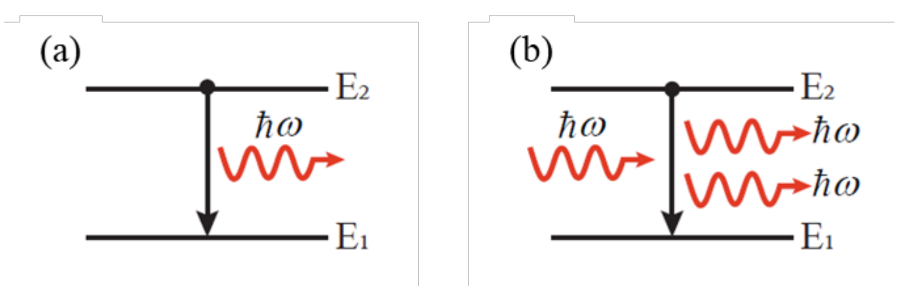

Figure 3.5. Two-level system. (a) Spontaneous emission; (b) Stimulated emission

So, an LED is slow because spontaneous emission is a slow process, but why is that? A rigorous explanation requires quantum physics, but a simple model shown in Figure 3.3 can capture most of the relevant details. Consider a single atom with two available electron states as before. We shine light on the atom to place the atom in the excited state. The electron prefers to be in the ground state and will eventually transition from the excited to the ground state. As this transition occurs, the electron is neither entirely in the excited state or entirely in the ground state. Instead the electron is sloshing periodically back and forth between two states. You can imagine the electron (really the electron cloud) sloshing back and forth around the positively charged nucleus. A negative charge separated a distance from a positive charge is known as a dipole in electromagnetic theory. The frequency of this dipole oscillation is given by the energy separation between the excited state and ground state (recall $E = h\nu$ where h is Planck's constant, ν is the oscillation frequency and E is the energy separation). So, we have charge moving back and forth in the atom. Movement of charge is a current, and we know that oscillating current can radiate energy because this is how an antenna works. In other words, the atom looks like a small radiating antenna! Here, our atomic antenna is radiating light waves instead of radio waves which we usually associate with an antenna. Nonetheless, similar physics apply.

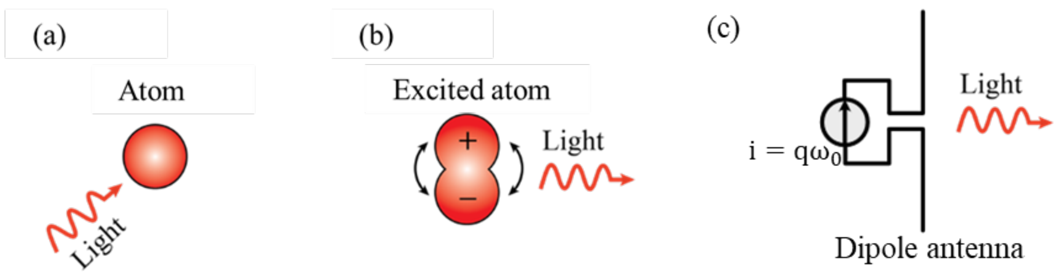

Figure 3.6. Antenna model for spontaneous emission from an atom; (a) external energy is used to excite the atom; (b) atom transitions from the excited to ground state forming an oscillating dipole; (c) equivalent representation of an oscillating dipole as a dipole antenna.

Unfortunately, our atomic antenna is not a great antenna. According to the Larmor formula, the power radiated by our atomic antenna is proportional to $\left(\frac{x_0}{\lambda}\right)^2$ where x_0 is roughly the same as the radius of the electron orbital. Now, λ is about a thousand times longer than x_0 so $\left(\frac{x_0}{\lambda}\right)^2$ is a very small number and therefore the power radiated by our atom is very small. There is a large mismatch between the two sizes: the atom is not a very good antenna for radiating light. Another way to think about this, is that the atom likes to store energy for a very long time before releasing that energy in the form of light. Recall that the units for power are Joules per second, in other words, a rate times energy where the rate is called the spontaneous emission rate and the energy is the photon energy. In other words, low power means low spontaneous emission rate. Therefore, using a very simple atomic model we have proven that spontaneous emission is a slow process. Now, semiconductor LEDs are of course comprised of many atoms and our simple picture gets more complicated; however, it turns out that the behavior predicted by our simple atomic model is also observed in semiconductor materials.

3.3 ANTENNA-ENHANCED SPONTANEOUS EMISSION

So, we have determined that light emission in semiconductor LEDs is slow because spontaneous emission is slow. And spontaneous emission is slow because atoms are poor antennas for their own light emission. What happens if we somehow attach a "good" antenna to a semiconductor? Take for example a simple dipole antenna which consist of two metal arms separated by a gap (Figure 4). If we insert the semiconductor inside the gap of the dipole antenna, remarkably, we will find that the spontaneous emission rate increases. This rate increase can be observed by measuring the light emission of the antenna-enhanced semiconductor and comparing the result to case without the antenna present. The light emission intensity is often measured as a function of wavelength or a function of time. The former is called a spectral measurement whereas the latter is known as a time-dependent measurement. In both cases, the semiconductor is excited by either applying a voltage across the semiconductor or by shining another light source on the semiconductor. A continuous excitation is commonly employed for the spectral measurement whereas a pulsed excitation is used for the time-dependent measurement.

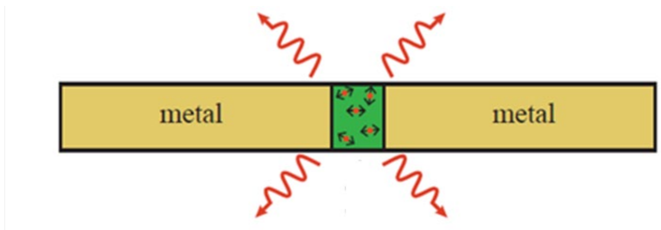

Figure 3.7. Semiconductor inserted into the gap of a metal dipole antenna

The antenna enhancement is a figure of merit that describes how strongly the antenna enhances the spontaneous emission and can be written as $F = \frac{R_{rad}}{R_{rad,0}}$ where R_{rad} and $R_{rad,0}$ are the spontaneous emission rate with and without an antenna respectively. The largest antenna enhancement values are achieved when three conditions are met: (1) the antenna gap spacing is much smaller than the wavelength of light; (2) the length of the antenna is about $\frac{\lambda}{2}$ where lambda is the peak emission wavelength of the semiconductor; and (3) the metal that comprises the antenna is relatively free of loss.

Antenna length

Most antennas perform better when the size of the antenna is about equal to the wavelength of whatever radiation the antenna is designed to emit. In the case of the dipole antenna described above, the ideal length is closer to $\lambda/2$ where λ is the peak emission wavelength of the semiconductor (the length is the tip-to-tip distance of both antenna arms). The peak emission wavelength is roughly equal to hc/E_g where E_g is the bandgap energy of the semiconductor. Semiconductors have a bandgap energy on the order of 1 eV, and therefore the dipole antenna is about 500 nanometers (nm) long or 10^{-9} meters. The antennas are extremely small and require sophisticated tools to manufacture them in a clean room environment.

Antenna gap spacing

It can be shown that the antenna enhancement is proportional to $\left(\frac{\lambda}{d}\right)^2$ where d is the antenna gap spacing (the distance between the two metal arms shown in Figure 4). Therefore, a very small antenna gap is needed to achieve the largest antenna enhancement. Shown in Figure 5 is a plot of the expected antenna enhancement as function of antenna gap spacing. We see that the antenna enhancement increases dramatically when the antenna gap is less than 30 nm. This is extremely small! For a gap spacing of 10nm the antenna enhancement is about 1000.

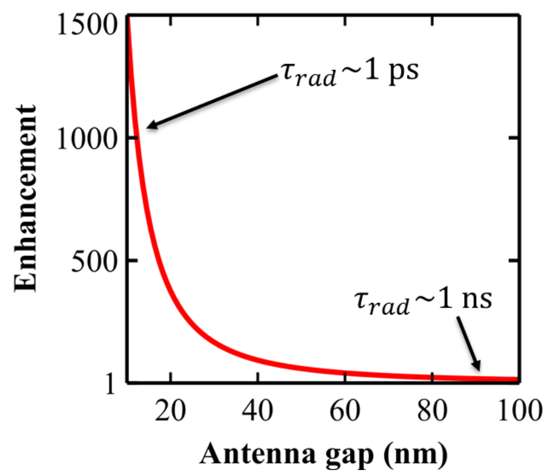

Figure 3.5. Antenna enhancement plotted as a function of the antenna gap.

As described before, the typical spontaneous emission lifetime of a semiconductor LED (without an antenna) is about 1 ns which allows for about 1 gigabit-per-second data rates. Enhancing the spontaneous emission by 1000 times would lead to an LED with a data rate of ~1000 gigabits-per-second. This is a remarkable speedup but we must not also forget about the power emitted by such a small LED. By scaling down the antenna gap we effectively scale down the volume of the LED thereby decreasing the total power since there is less volume of material that is emitting light. If the LED is too small, it will be difficult find a photoreceiver that is sensitive enough to detect the emitted light on the other end of an optical communication link.

Antenna metal

Most metals will work reasonably well for an antenna that is designed to work at radio frequencies. However, the conductivity of metal drops as the frequency is increased. The frequency of a light wave is about a thousand time larger than the frequency of a radio wave and therefore the conductivity of metal that is at optical frequencies can be rather low. The noble metals silver (Ag) and gold (Au) are an exception and have reasonably high conductivities and are often used as the antenna metal at optical frequencies. Nonetheless, these metals still have conductivity at optical frequencies that is lower than the conductivity of most metals at radio frequencies. Because of this, optical antennas tend to have more resistive loss than their radio frequency counterparts. Careful design of the antenna can help negate the effect of this low conductivity but is beyond the scope of this chapter.

Metals also exhibit another behavior called kinetic inductance that is observed at optical frequencies but that is not observed at radio frequencies. Normally, when an electromagnetic wave impinges on a metal, the electrons within the metal respond immediately to the electric field in the wave since electrons can move freely within the metal. At optical frequencies, the electromagnetic wave can be oscillating so rapidly that the electrons are a little slow to respond to the electric field. If the electrons lag very far behind the electric field than we say that the metal is “plasmonic”. This is often the case for metals at optical frequencies and so researchers have coined a term called “plasmonics” to describe the field that involves the research of devices that use metal at optical frequencies. The plasmonic effect is often undesirable but it has been used for desirable purposes such as defeating the diffraction limit described earlier in this chapter.

3.4 EXPERIMENTAL DEMONSTRATION OF ANTENNA-ENHANCED SPONTANEOUS EMISSION

Shown in Figure 3.6a is a scanning electron microscope image of a semiconductor that is inserted into the gap of a dipole antenna. The InGaAsP semiconductor is composed of indium, gallium, arsenic, and phosphorous and is known as a III-V semiconductor because each of these elements sit in either the III-column or V-column of the periodic table. InGaAsP is a direct bandgap semiconductor that efficiently emits light with a wavelength near 1.55 µm. Plotted in Figure 6b is the spectral measurement of the light emission from the semiconductor with and without an antenna. Strong light emission occurs in the semiconductor with antenna indicating an increase in the efficiency of light emission. The efficiency of the light emission can be written as $\eta = \frac{R_{rad}}{R_{nr}+R_{rad}}$ where R_{nr} is called the non-radiative rate. R_{nr} is the rate at which electrons jump to the ground state (i.e. the valence band) without photon emission. This is an undesirable process but cannot be avoided in most materials. As described before, the antenna enhances the spontaneous emission thereby increasing the radiative rate; however, the non-radiative rate remains the same. We can see that by increasing R_{rad} the efficiency of the light emission increases therefore increasing the light emission intensity.

In this experiment, the semiconductor was excited by shining the light from a tabletop laser onto the semiconductor. The spectrum shown in Figure 3.6b is commonly called a photoluminescence or PL spectrum. It is convenient to use an external laser in the research lab, but it is not practical for a commercial device as this usually requires an electrical source of excitation which is significantly more compact and efficient. Such an electrically-injected device can be created by inserting a pn-junction in the semiconductor material, thereby forming an LED. The LED is "turned-on" by applying a positive voltage to the p-side of the pn-junction with respect to the n-side. This will cause electrons to travel across the junction to recombine with holes on the p-side and vice- versa. When the electrons and holes recombine, light is emitted. This light emission is called electroluminescence or EL. Electroluminescence is the method by which traditional LEDs emit light.

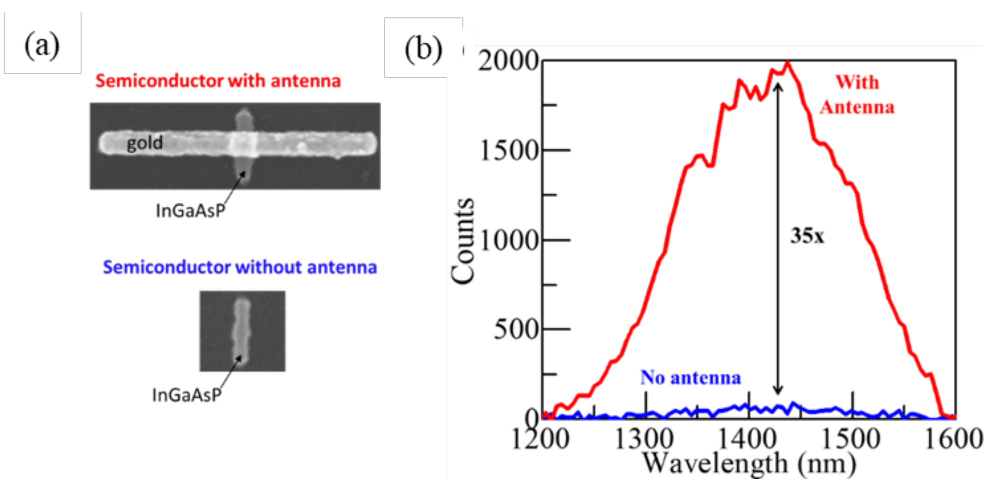

Figure 3.6. Experimental demonstration of antenna-enhanced spontaneous emission (a) image of a III-V semiconductor with an without an antenna (b) Light emission spectra showing large increase in the light emission intensity from semiconductor with antenna.

3.5 ELECTRICALLY-INJECTED III-V ANTENNA-LED

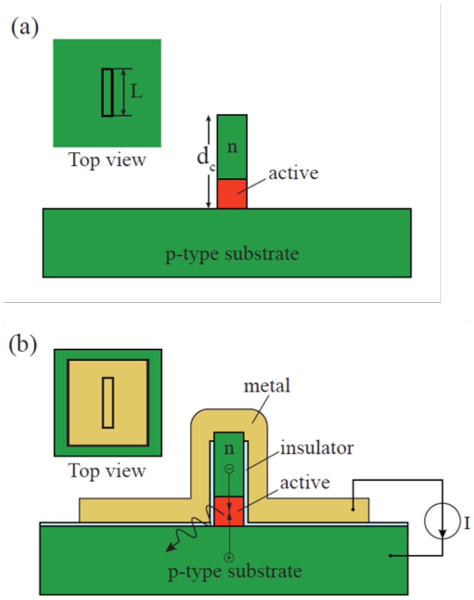

Figure 3.7. Schematic of the eletrically-injected antenna-LED. (a) bare LED ridge without antenna. (b) LED ridge with slot antenna.

Now, there is a problem with inserting this pn-junction because in addition to the antenna metal, we must somehow bring in two electrodes to contact the p and n-type regions of the pn-junction. This is challenging because the pn-junction is extremely small, and the placement of the electrodes may affect the antenna if not done properly. Instead of a dipole antenna, it is much easier to use a slot antenna. The slot antenna is simply a metal sheet with a rectangular hole cut out of it. The basic idea is shown in Figure 3.7. A small semiconductor ridge composed of the pn-junction is etched into a semiconductor wafer. The ridge is then covered with a sheet of metal. The sheet of metal is continuous except at the semiconductor ridge where a small protrusion is made into the metal. This discontinuity in the metal creates an opening that forms the slot of the slot antenna. Importantly, the antenna metal contacts the n-side of the pn-junction ridge allowing for the metal to serve as an electrode for the pn-junction. The entire substrate on which the device sits is p-type and electrically connected to the p-side of the pn-junction. It is straightforward to deposit an electrode on the p-type semiconductor on some other part of the substrate far away from the antenna.

The manufactured antenna-LED is shown in Figure 3.8. Figure 3.8a shows a scanning electron microscope image of the antenna-LED as viewed from above. Figure 3.8b shows the same device in a cross-sectional view so that inner detail of the device can be more clearly seen. This antenna-LED uses the III-V semiconductor InGaAs as the light emitting material.

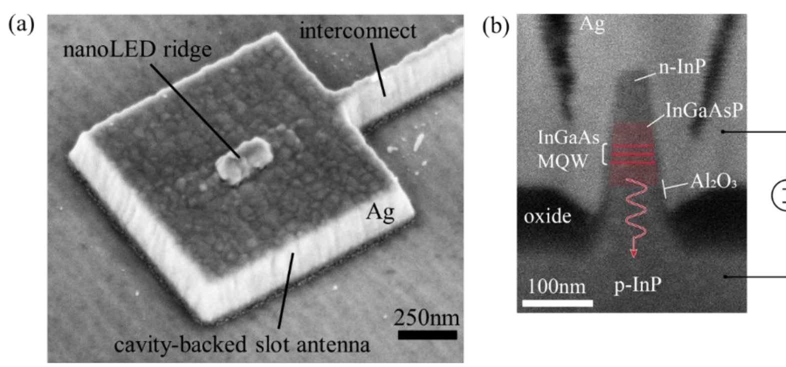

Figure 3.8. Manufactured electrically-injected antenna-LED (a) Angled view (b) Cross-sectional view.

A semiconductor parameter analyzer is used to apply a constant current to the antenna-LED. After the current is applied, the antenna-LED will emit light. This light will be captured by a microscope objective and then routed to a spectrograph or infrared (IR) camera. The spectrograph is used for the spectral measurement (intensity vs. wavelength) whereas the IR camera is used for taking an IR image of the antenna-LED. Shown in Figure 3.10 is the spectral measurement and IR image of the antenna-LED. We observe again that the light emission intensity increases with respect to an LED without antenna.

3.6 TIME-RESOLVED MEASUREMENT

The electroluminescence measurement described in Figure 3.9 can be thought of as a "DC" measurement since the supplied current to the device is continuous and does not change with time. It is also useful to measure the dynamic time-dependent response of the antenna-LED to better understand how the device can be turned on and off (modulated) for use in optical communication. Such a time-resolved measurement is often done by first pulsing the device with a light or current pulse and then measuring the emitted light as a function of time using a fast photodetector. A typical result is shown in Figure 3.10. The light emission from the antenna-LED decreases faster compared with the LED without antenna as expected due to the increased spontaneous emission rate. Interpreting the time-resolved measurement data is not necessarily straightforward. Without the antenna, the non-radiative rate dominates the dynamic behavior of the device. With the antenna, the radiative rate catches up with the non-radiative and dominates the dynamic behavior. As a result, the actual speedup of the antenna-LED is several times faster than what is implied in Figure 10.

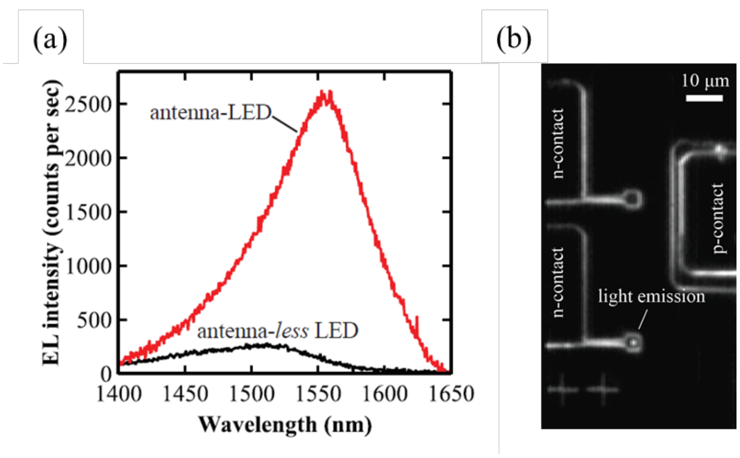

Figure 3.9. Electroluminescence from electrically-injected antenna-LED. (a) Enhanced light emission is observed from the antenna-LED (b) IR image showing tiny area of light emission from the antenna-LED.

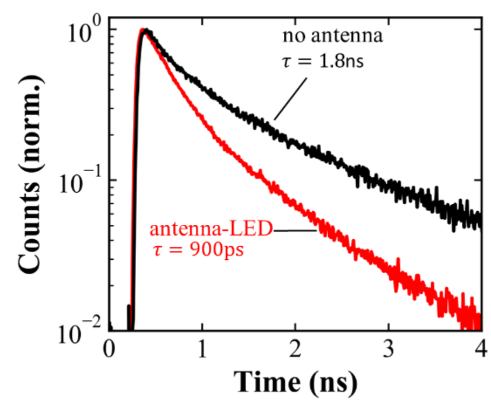

Figure 3.10. Time-resolved light emission measurement

3.7 OUTLOOK

In this chapter, basic theory and experimental results of the antenna-LED was introduced. It was shown that an antenna can significantly speed up the rate of spontaneous emission and therefore the modulation speed on an LED. Antenna-LEDs are expected to have applications in short-range (< 1cm) optical communication links where their small size and high speed can be used for energy efficient communication. Several challenges remain before implementation of antenna-LEDs into optical links will be possible. The very high modulation speeds predicted by the theory have yet to be demonstrated. This will require very small devices (as predicted by Figure 3.5) which poses manufacturing and measurement challenges. The efficiency of the antenna-LED is still smaller than desired. This is largely due to high non-radiative recombination of electrons and holes at the semiconductor surface given the large surface-to-volume ratio of the devices. Finally, the devices described in this chapter are not integrated into any larger systems and are instead "free-space coupled". It is more desirable to integrate the antenna-LED with an optical waveguide for use in integrated circuits. This will be the topic of a later chapter in this e-book.

3.8 ADDENDUM: REMAINING CHALLENGES FOR THE ANTENNA-LED

Sean Hooten

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

The antenna-LED was introduced and discussed in detail this chapter, and it was argued that it has many desirable properties in comparison to lasers for use in on-chip optical interconnects. Nevertheless, there are several remaining fundamental challenges for the implementation of antenna-LEDs in a realistic optical link. The challenges are listed below, with possible solutions.

1. *Can antenna-LEDs really be as fast or faster than lasers?* Lasers have well known modulation speeds (the time it takes to turn them on and off) and can be extremely fast. In fact, the most widely accepted model for laser modulation speed (small-signal 3dB modulation rate) would seem to indicate that they can always be faster than antenna-LEDs. However, it turns out that this model is incorrect when considering large-signal digital modulation* – in other words, a more realistic operating mode for a laser in an on-chip optical interconnect. Without going into the details, lasers and antenna-LEDs are limited by very similar physics in this situation. Therefore, so long as significant rate enhancement can be demonstrated, the antenna-LED can in fact be as fast as a laser. *Reference: R. S. Tucker, *Electronics Letters* 20, 802–803 (1984).

2. *Do antenna-LEDs have enough signal for optical links?* Earlier in this chapter it was emphasized that the antenna-LED is extremely small, on the order of nanometers in size. Consequently, there is not very much optical material available to emit photons. Under a conservative estimate, antenna-LEDs could emit approximately 100 photons per transmitted bit. Modern large-scale optical links require approximately 10,000 photons per bit for adequate signal, so improving the antenna-LED power or the detector sensitivity is a major priority for feasibility.

3. *Are antenna-LEDs efficient enough?* Part of the original argument for adopting on-chip optical interconnects was that they could be more efficient than electrical wires. Therefore, the efficiency of the optical source should be maximized. Lasers and LEDs have somewhat different challenges in this regard. Lasers must overcome a threshold current, and there are additional optical and material losses that must be considered. Antenna-LEDs, on the other hand, are sensitive to material defects and there are optical losses due to Ohmic resistance in the metallic antenna. A careful analysis of the antenna-LED shows that it can reach good efficiency, perhaps out-pacing the laser, but material defects must be significantly improved.

Chapter 4: Monolayer Semiconductor Emitter

Kevin Han

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

4.0 INTRODUCTION

As mentioned in the previous chapter, a compact, high-speed and high-efficiency on-chip light emitter can potentially reduce energy consumption of communication compared to electrical wires. Apart from traditional III-V semiconductors, it is interesting to explore other material systems which may present new advantages as well as open up new applications. One such set of materials are the monolayer *transition-metal dichalcogenides (TMDCs)*, which will be the focus of this chapter.

This chapter assumes the knowledge of the previous chapters, and high-school knowledge of chemistry and electromagnetics. After finishing the chapter, the reader will have an understanding of 2D semiconductors and their use in light-emitting devices.

4.1 MONOLAYER SEMICONDUCTORS

As their name suggests, TMDCs are composed of one transition metal element (such as W or Mo) and two chalcogens (O, S, Se, or Te). Common examples include WSe_2 and MoS_2. They occur naturally in layered form. Each molecular layer is separated from the others by weak van der Waals forces, instead of strong covalent or ionic bonds. This is shown in Fig. 4.1. A more familiar layered material may be graphite, which is composed of only carbon.

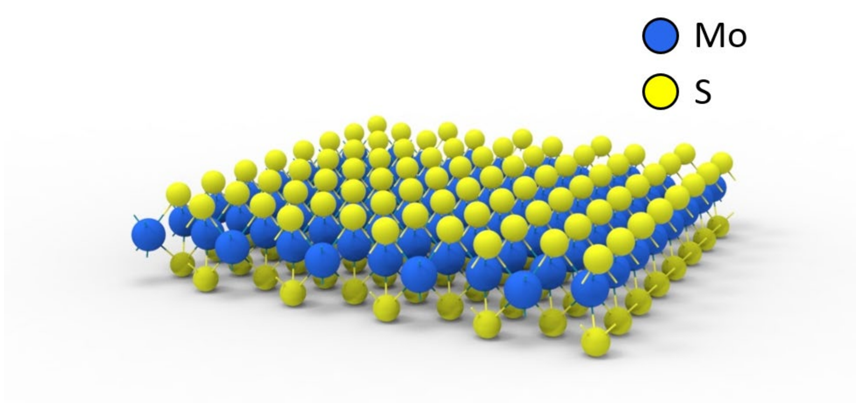

Figure 4.1. Structure of monolayer MoS_2, a commonly-studied TMDC.

When thinned down to a single layer, TMDCs change their properties drastically. In their bulk form, they are *indirect-gap* semiconductors, which means they are inefficient light emitters. A common indirect-gap semiconductor is silicon, which is used for electronics but not for LEDs or lasers. However, when TMDCs are thinned to a monolayer, they become *direct-gap* semiconductors like the III-V materials in the last chapter, so they can be used for light emission. Due to their molecular thickness, they are also known as “2D materials”.

Such 2D semiconductors present several possible advantages relative to traditional bulk, or 3D, III-V emitters.

1. **Lack of surface recombination.** TMDCs lack dangling bonds on their surface since layers are only held together by van der Waals forces. Therefore, surface recombination is avoided in this dimension. Recall from the last chapter that surface recombination is an important problem for III-V nanoscale devices due to their high surface to volume ratio. Surface recombination causes injected current to be lost at the surface of the device instead of emitting light.
2. **Material integration**. They can be easily transferred or grown onto diverse substrates. One problem in III-V optoelectronics is integration with photonic devices on silicon. Silicon is attractive for passive elements such as waveguides due to its low cost and mature processing technology, while III-Vs are required for light emitters due to silicon’s indirect band gap. However, growing III-Vs directly on silicon leads to reduced material quality and less efficient light emission. 2D materials can be transferred without significant loss of material quality. Moreover, a rich library of 2D materials exists, including insulators (BN), conductors (graphene), and semiconductors (TMDCs). This can lead to diverse functionality such as sources with multiple colors of emission in one chip.
3. **New applications**. The monolayer thickness of TMDCs opens up applications in several areas. In optics, they can potentially be used for transparent displays. They are also being investigated for sensing due to their high sensitivity to the environment. For electronics, they enable scaling down of transistors due to their efficient electrostatic gating possibility. They can handle enormous amounts of strain without breaking (up to ~10%) relative to bulk semiconductors (~1%), enabling flexible or strain-tunable devices.

4.2 LIGHT-EMITTING DEVICES USING 2D SEMICONDUCTORS

There has been extensive research on the light emission properties of 2D semiconductors. Most studies involve optical pumping, known as *photoluminescence*. In this process, light of a certain wavelength goes into the material, excites electrons inside, these electrons lose energy, and light of a higher wavelength (lower energy) comes out. However, for practical applications, we want to turn electricity into light (*electroluminescence*) instead of light into light. This turns out to be more challenging due to the requirement for electrical injection of both electrons and holes. To meet this challenge, there have been three main device designs demonstrated: a lateral injection structure, a vertical injection structure, and a unipolar light-emitting capacitor. The schematics of each device are shown in Figure 4.2.

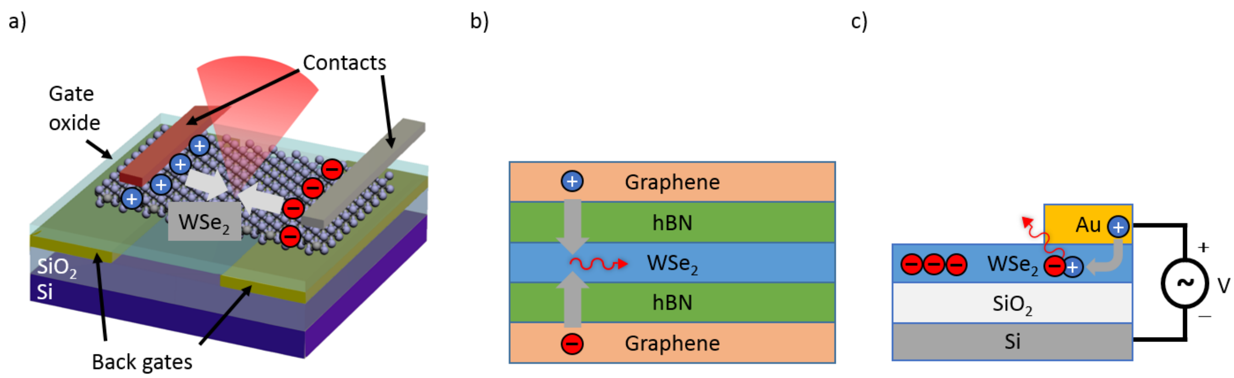

Figure 4.2. Electrically injected TMDC light emitters. a) Lateral p-n junction design. b) Side view of vertical injection structure with hBN tunnel barriers. c) Side view of light-emitting capacitor, depicting the p-injection cycle. Holes are injected and recombine with electrons in the monolayer populated from the previous, n-injection, cycle.

The lateral injection scheme was demonstrated first in 2014 [1-3]. In this approach, electrons and holes are injected from two contacts to the monolayer separated by some distance, usually a few micrometers (microns). Recall that conventional semiconductor emitters require a p-n junction: a hole-rich (p-doped) region next to an electron-rich (n-doped) region. However, the manufacturing technologies used to form these regions are challenging to adapt to monolayer semiconductors due to their thinness. One way to overcome this is to form two capacitors on each side of the device near the contacts. Opposite voltages are applied to each capacitor to attract positive and negative charge into the monolayer ($Q = CV$). This is realized using a gate dielectric underneath the monolayer, with separate metal conductors (back gates) under the dielectric, one near each contact. The p contact should be surrounded by holes, so a negative voltage with respect to the p contact is applied to the p back gate, and vice versa for the n contact. In this way, a p-n junction is realized. Fig. 4.3 shows emission from a real device. Note that in this device there is only one back gate (i.e. the two back gates are shorted to each other), which is simpler to fabricate and allows adequate carrier injection and light emission.

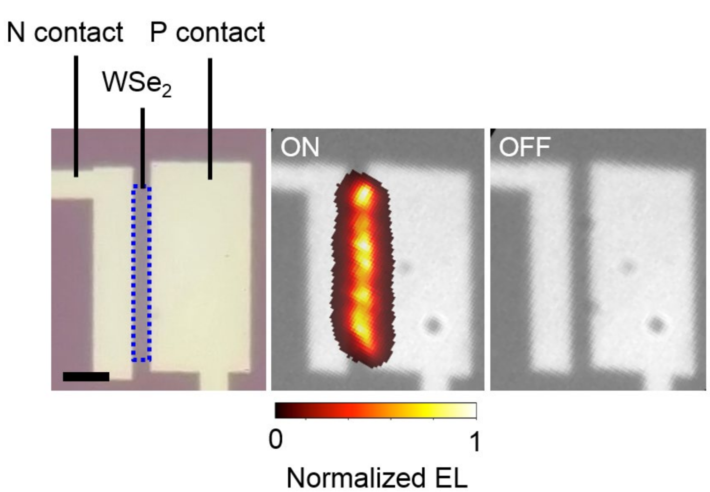

Figure 4.3. Light emission from lateral p-n junction TMDC LED. Left: image of the device. Middle: emission intensity overlaid on image. Emission occurs in the region between the two contacts. Right: device in the off state with no bias applied. Printed with permission from [5].

The vertical injection scheme was demonstrated in 2015 [4]. In this approach, no explicit p-n junction is formed. Instead, the monolayer is sandwiched between few-layer hexagonal boron nitride (hBN) that is 1-3 nm thick. hBN is another layered material but with insulating properties (band gap of ~6 eV). This stack is in turn sandwiched by graphene electrodes. Due to the thinness of the hBN, holes and electrons can quantum tunnel into the semiconductor with sufficient voltage applied between the electrodes. These same insulators then act as tunnel barriers to keep the carriers in the monolayer from escaping to the opposite electrode. This design is more popular in recent years due to its ease of fabrication and higher brightness. The higher brightness is due to its larger injected area, since the whole area can be injected at once, similar to a conventional quantum well LED. However, the hBN layers act as very thin capacitors, leading to high capacitance and lower modulation speeds than lateral junctions [5], [6].

Finally, in 2018 a novel light-emitting capacitor design was demonstrated (Fig. 4.2c) [7]. A metal-oxide-semiconductor (MOS) capacitor structure is formed using a back gate metal, gate oxide, and TMDC. Only one top contact to the TMDC is required. Then, an alternating polarity pulsed voltage is applied between the top contact and the back gate to inject holes and electrons in alternating cycles. During the hole (electron) injection, some electrons (holes) from the previous cycle will recombine with the incoming carriers, emitting light. This has the advantage of simple fabrication compared with the other two schemes. However, the device efficiency is somewhat lower (~0.1% compared with ~1-5% for the other types). This is because of the inherently poor carrier confinement. While injecting holes, some of the electrons from the previous cycle may diffuse back into the contact without recombining with the holes coming in.

4.3 2D SEMICONDUCTORS COUPLED TO OPTICAL CAVITIES

Just like bulk semiconductors discussed in the last chapter, 2D materials can be coupled to optical cavities, including optical antennas, to enhance light-matter interaction. One way to couple a monolayer to an antenna is shown in Fig. 4.4a. Just as in the previous chapter, a cavity-backed slot antenna can be covered with a monolayer. The spontaneous emission rate inside the slot is enhanced. This design using a silver slot antenna has shown >300x peak rate enhancement compared to a non-antenna-coupled monolayer [8].

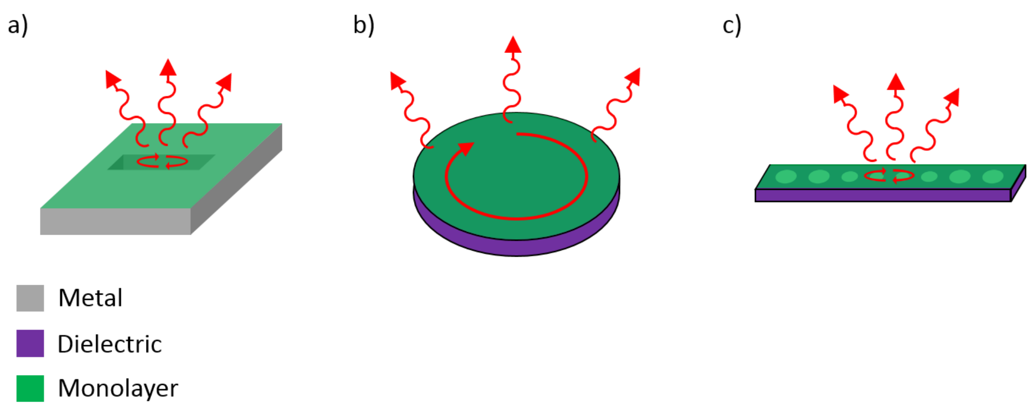

Figure 4.4. Antenna and cavity structures coupled to 2D materials. a) Slot antenna with monolayer on top. b) Dielectric microdisk laser. c) Photonic crystal nanobeam laser.

Besides LEDs, many lasers using 2D materials have been demonstrated [9-16]. The rate enhancement does not necessarily have to be high for this application. What is more important is how well the light emitted stays inside the cavity to continue interacting with the monolayer, causing stimulated emission. Fig. 4.4b shows a microdisk made of a dielectric material, with the monolayer atop the disk. The dielectric is chosen to have low optical absorption so the light propagating inside is not lost by absorption. The light circulates around the edge of the disk, as shown in the figure. Fig. 4.4c shows a less intuitive structure, a photonic crystal nanobeam with a monolayer on top. This consists of a periodic array of holes etched into a dielectric beam. The size of the holes is varied in the center to form a cavity, and destructive interference from the surrounding holes causes the light to stay in the central region (see next chapter for more on interference). For a totally periodic structure with all holes the same size, there is total destructive interference and nowhere for the light to propagate inside the beam.

All of the lasers shown to date have been optically-injected. One significant milestone would be to realize an electrically-injected laser. The challenge here is to integrate the cavity with an electrical injection structure that can supply enough current to turn on the laser. Many cavity structures such as microdisks are simple to optically inject but hard to contact electrically.

4.4 OUTLOOK

Ever since the discovery of monolayer MoS_2 in 2010 [17], the field of 2D materials has exploded, with many promising results demonstrated in a range of areas but still much to be understood. In fact, a recent theoretical study predicts over 1000 2D layered materials exist, while only a handful have been intensely studied [18]. For 2D light emitters, perhaps the biggest challenge is to increase the internal quantum efficiency of electroluminescence from a few percent to near 100%, like in traditional III-V LEDs. Chemical treatments and electrical doping techniques have shown that the photoluminescence efficiency can be near 100% [19], [20], but these have yet to be integrated with electrically injected devices.

REFERENCES

[1] B. W. H. Baugher, H. O. H. Churchill, Y. Yang, and P. Jarillo-Herrero, "Optoelectronic devices based on electrically tunable p-n diodes in a monolayer dichalcogenide," *Nat Nano*, vol. 9, no. 4, pp. 262–267, Apr. 2014.

[2] J. S. Ross *et al.*, "Electrically tunable excitonic light-emitting diodes based on monolayer WSe2 p-n junctions," *Nat Nano*, vol. 9, no. 4, pp. 268–272, Apr. 2014.

[3] A. Pospischil, M. M. Furchi, and T. Mueller, "Solar-energy conversion and light emission in an atomic monolayer p–n diode," *Nature Nanotechnology*, vol. 9, no. 4, pp. 257–261, Apr. 2014, doi: 10.1038/nnano.2014.14.

[4] F. Withers *et al.*, "WSe2 Light-Emitting Tunneling Transistors with Enhanced Brightness at Room Temperature," *Nano Lett.*, vol. 15, no. 12, pp. 8223–8228, Dec. 2015.

[5] K. Han *et al.*, "Bright electroluminescence in ambient conditions from WSe2 p-n diodes using pulsed injection," *Appl. Phys. Lett.*, vol. 115, no. 1, p. 011103, Jul. 2019, doi: 10.1063/1.5100306.

[6] C.-H. Liu *et al.*, "Nanocavity Integrated van der Waals Heterostructure Light-Emitting Tunneling Diode," *Nano Lett.*, vol. 17, no. 1, pp. 200–205, Jan. 2017, doi: 10.1021/acs.nanolett.6b03801.

[7] D.-H. Lien *et al.*, "Large-area and bright pulsed electroluminescence in monolayer semiconductors," *Nature Communications*, vol. 9, no. 1, p. 1229, Mar. 2018.

[8] M. S. Eggleston *et al.*, "Ultrafast Spontaneous Emission from a Slot-Antenna Coupled WSe2 Monolayer," *ACS Photonics*, vol. 5, no. 7, pp. 2701–2705, Jul. 2018, doi: 10.1021/acsphotonics.8b00381.

[9] E. Y. Paik, L. Zhang, G. W. Burg, R. Gogna, E. Tutuc, and H. Deng, "Interlayer exciton laser of extended spatial coherence in atomically thin heterostructures," Nature, vol. 576, no. 7785, pp. 80–84, Dec. 2019, doi: 10.1038/s41586-019-1779-x.

[10] X. Ge, M. Minkov, S. Fan, X. Li, and W. Zhou, "Laterally confined photonic crystal surface emitting laser incorporating monolayer tungsten disulfide," npj 2D Materials and Applications, vol. 3, no. 1, p. 16, Apr. 2019, doi: 10.1038/s41699-019-0099-1.

[11] Y. Liu et al., "Room temperature nanocavity laser with interlayer excitons in 2D heterostructures," Science Advances, vol. 5, no. 4, p. eaav4506, Apr. 2019, doi: 10.1126/sciadv.aav4506.

[12] H. Fang et al., "1305 nm Few-Layer MoTe2-on-Silicon Laser-Like Emission," Laser & Photonics Reviews, vol. 12, no. 6, p. 1800015, 2018, doi: 10.1002/lpor.201800015.

[13] Y. Li et al., "Room-temperature continuous-wave lasing from monolayer molybdenum ditelluride integrated with a silicon nanobeam cavity," Nature Nanotechnology, vol. 12, no. 10, pp. 987–992, Oct. 2017, doi: 10.1038/nnano.2017.128.

[14] Y. Ye et al., "Monolayer excitonic laser," Nature Photonics, vol. 9, no. 11, pp. 733–737, Nov. 2015, doi: 10.1038/nphoton.2015.197.

[15] O. Salehzadeh, M. Djavid, N. H. Tran, I. Shih, and Z. Mi, "Optically Pumped Two-Dimensional MoS2 Lasers Operating at Room-Temperature," Nano Lett., vol. 15, no. 8, pp. 5302–5306, Aug. 2015, doi: 10.1021/acs.nanolett.5b01665.

[16] S. Wu et al., "Monolayer semiconductor nanocavity lasers with ultralow thresholds," Nature, vol. 520, no. 7545, pp. 69–72, Apr. 2015, doi: 10.1038/nature14290.

[17] K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, "Atomically Thin MoS2: A New Direct-Gap Semiconductor," *Phys. Rev. Lett.*, vol. 105, no. 13, p. 136805, Sep. 2010.

[18] G. Cheon, K.-A. N. Duerloo, A. D. Sendek, C. Porter, Y. Chen, and E. J. Reed, "Data Mining for New Two- and One-Dimensional Weakly Bonded Solids and Lattice-Commensurate Heterostructures," *Nano Lett.*, vol. 17, no. 3, pp. 1915–1923, Mar. 2017.

[19] M. Amani *et al.*, "Near-unity photoluminescence quantum yield in MoS2," *Science*, vol. 350, no. 6264, pp. 1065–1068, Nov. 2015.

[20] D.-H. Lien *et al.*, "Electrical suppression of all nonradiative recombination pathways in monolayer semiconductors," *Science*, vol. 364, no. 6439, pp. 468–471, May 2019.

Chapter 5: Waveguide Integration

Nicolas M. Andrade

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

5.0 INTRODUCTION

The previous two chapters have described novel light emitters; however, in order to create a link, we need to be able to guide the light to a photodetector. We can achieve this by coupling light into a waveguide (see chapter 2.2). In this chapter we will describe ways to manipulate electromagnetic fields in order to couple the light emitted from a cavity backed slot antenna to a single mode waveguide.

The reader is expected to have knowledge of electricity and magnetism and to have read Chapter 2 to understand the basic principles behind waveguides.

5.1 BACKGROUND

Before discussing on-chip links it may be useful to look at an analogous macroscopic picture. The internet relies heavily on coupling laser light into a low loss optical fiber which guides the light to the receiver on the other end. Although our design is on a much smaller scale, it can be informed by the basic principle behind coupling lasers to optical fibers.

The most important parameters for coupling laser light into optical fibers is **mode overlap** and **power coupling**, which describes how similar the two modes are. The equation for mode overlap is given below in Eq. 5.1, where we want the value to be as close to one as possible - which would indicate the electric (E) and magnetic (H) fields are the same [1]. To get the power coupling we include **Fresnel Reflection**, which is the amount of power reflected when light goes from one material to another, in Eq. 5.2 we give the power coupling at normal incidence.

$$\eta_{overlap} = Real\left[\frac{(\int \overrightarrow{E_1}\times\overrightarrow{H_2^*}\cdot d\vec{S})(\int \overrightarrow{E_2}\times\overrightarrow{H_1^*}\cdot d\vec{S})}{(\int \overrightarrow{E_1}\times\overrightarrow{H_1^*}\cdot d\vec{S})(\int \overrightarrow{E_2}\times\overrightarrow{H_2^*}\cdot d\vec{S})}\right] \qquad \text{Eq. 5.1}$$

$$\eta_{power\ coupling} = \eta_{overlap}\left(1 - \left|\frac{n_1-n_2}{n_1+n_2}\right|^2\right) \qquad \text{Eq. 5.2}$$

Where n is the refractive index of the material.

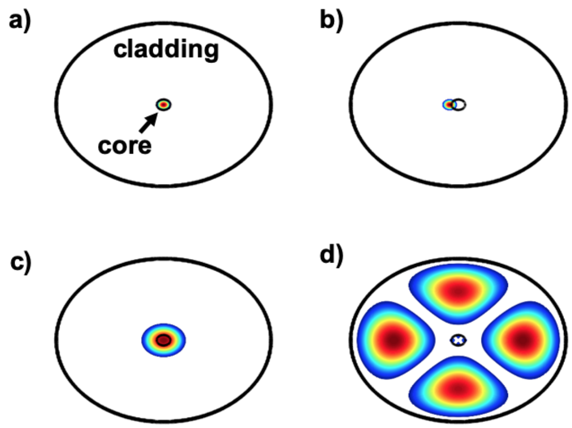

Figure 5.1. Representation of problems when coupling to a single mode optical fiber. a) mode with perfect shape and location, b) mode offset from core of fiber, c) mode larger than core of fiber, d) mode with wrong profile

A more intuitive representation of power coupling is shown in Fig. 5.1. In order to get good power coupling (Fig. 5.1a) we need to engineer for the mode size, offset, and profile. Any deviation from the ideal mode can lead to large decreases in the power coupling efficiency.

The following section will describe the different ways we try to position and shape the output mode from our optical antenna-LEDs.

5.2 WAVEGUIDE COUPLING

In the following subsections we will describe the theory behind coupling the cavity backed slot antenna to a waveguide, as shown in Fig. 5.2. In the figure, we see that the emission is directed straight down towards the substrate, making it a nontrivial problem to turn the emission profile 90° and couple it to a single mode waveguide. While the simulations are specific to this structure, the theory can be applied to more general problems.

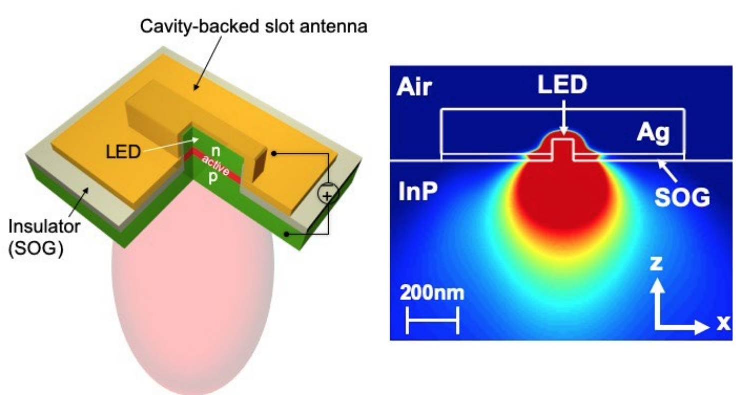

Figure 5.2. Sketch and simulation of cavity-backed slot antenna radiation pattern on the left and right, respectively.

5.2.2 CONSTRUCTIVE INTERFERENCE

Light is a wave with an amplitude, frequency, and phase. Light waves traveling in the same direction can be summed to give the total wavefront. This is known as superposition. As shown in Fig. 5.3, if these two waves have the same phase (in-phase), the total field is stronger or we can say there is **constructive interference**. If they are 180˚ out of phase (π phase shift), the fields will cancel each other and the light will not go in that direction; this phenomenon is known as **destructive interference**. Since waves are periodically adding or subtracting, a period (360 degrees or 2π radians) doesn't change the wave at all, so all phase shifts are given between 0-360 degrees (0-2π radians).

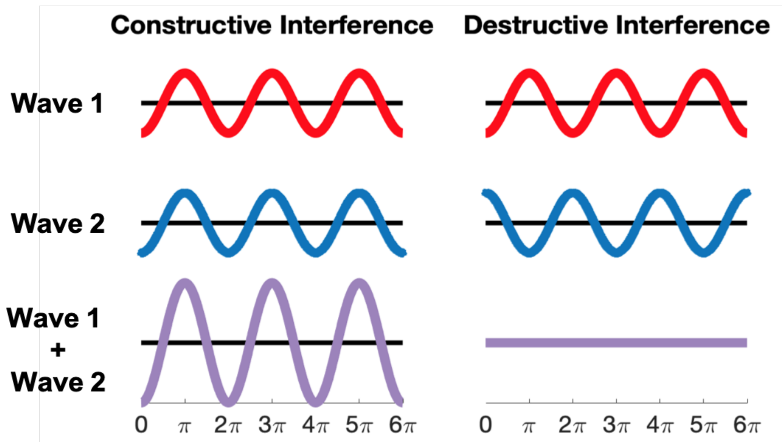

Figure 5.3. Constructive and destructive interference

In our case, we want the light going straight down to destructively interfere. We can accomplish this by increasing the phase offset between the initial beam and its reflection off the silver (Ag) surface; the easiest way to control the interference is by changing the layer thickness. As the layer thickness increases by $\lambda_0/2n$ the phase increases by π, where λ_0 is the wavelength of light in vacuum and n is the refractive index of the layer. In Fig. 5.4, we find that in order to achieve destructive interference for light propagating down into the SiO2 we want the waveguide layer to be a multiple of $\lambda_0/2n$ since it leads to an overall π phase shift for light propagating straight down - note that we pick up an additional π phase shift for the reflection off the metal interface. Additionally, at higher angles we will get constructive interference in our waveguide layer, which is exactly what we want for higher waveguide coupling.

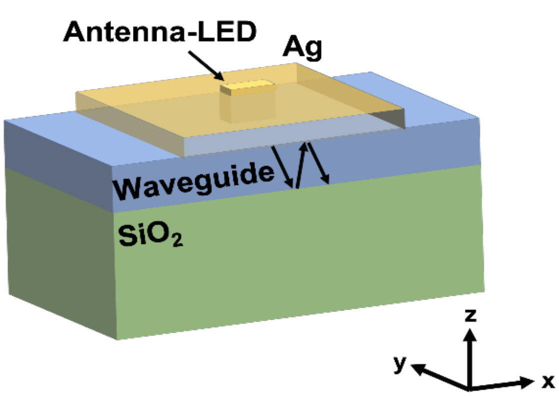

Figure 5.4-Cavity backed slot antenna-LED on single mode InP waveguide which is sitting on a SiO2 ridge. The black arrows in the waveguide show a basic ray bounce diagram, for a thickness of $\lambda_0/2n$ each black arrow shows a π phase shift. With the additional π phase shift for reflecting off the silver (Ag) surface we get destructive interference for light propagating down towards the SiO2 ridge.

This picture looks nice for the first reflection; however, the metal keeps adding a π phase shift - so for additional bounces it alternates between destructive and constructive interference. This may be why the waveguide coupling efficiency for this structure is limited to around 48% (24% in each direction).

5.2.3 IMAGE THEORY

In a perfect metal, electrons can redistribute themselves in response to an electric field, which results in no field penetrating into the metal. So, at the boundary of the metal and the dielectric the field must be zero. As shown in Fig. 5.5, in order to satisfy these boundary conditions and solve for the fields in all space outside the metal we can place an **image charge** inside the metal with the opposite charge. As shown in Figure 5.5b this concept also applies to dipoles.

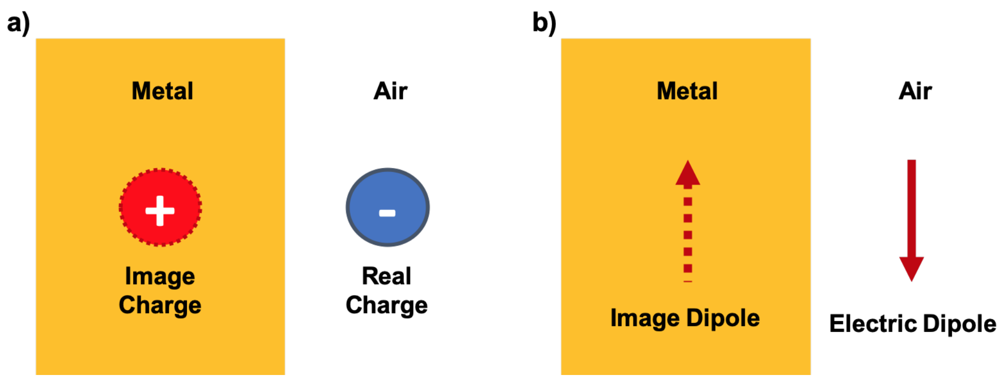

Figure 5.5. a) Charge and its induced image charge in the metal, and b) electric dipole and induced image dipole in the metal

While the theory is for an infinite metal plane, it will hold approximately true for our structure. In Fig. 5.6 we took our structure from the previous subsection and truncated one end of the waveguide and wrapped silver around this facet. In Fig 5.6a, we overlaid both the cavity backed slot antenna-LED electric dipole and the expected image dipole in the silver back plane. As shown in Fig. 5.6b, we were able to achieve a waveguide coupling efficiency of 74% with the optimized structure. As expected, we only achieve constructive interference for fields in the waveguide by designing the separation between the antenna-LED and the facet, shown in Fig. 5.6c - a separation of approximately $\lambda_0/2n$ gives us destructive interference in the waveguide, which is undesired.

While this structure was able to achieve a relatively high and unidirectional waveguide coupling efficiency, further improvements can be made since not all the power near the waveguide was in the correct mode.

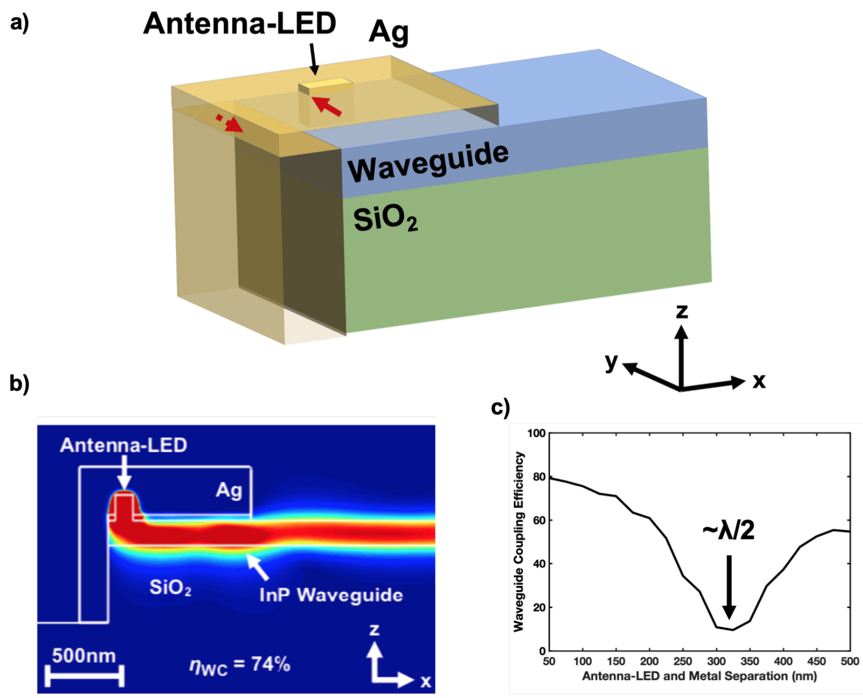

Figure 5.6. a) Perspective view of cavity backed slot antenna-LED on InP waveguide with one end truncated and silver (Ag) wrapped around facet. Additionally, electric dipole and image dipole are shown in solid and dashed red, respectively. b) The power flow for the optimized structure, with a waveguide coupling efficiency (η_{WC}) of 74%. c) Waveguide coupling efficiency of structure as a function of separation between antenna-LED and waveguide facet.

5.2.4 MODE SHAPING

In this subsection we will describe ways to prohibit higher order modes. In other words, how to shape the emission so it can couple to the desired waveguide mode. By using image theory, we can direct more light towards the waveguide; however, image theory alone cannot guarantee that we are coupling to the correct mode. In Fig. 5.7, we show two potential modes that exist under in the waveguide coupler section; however, only Fig. 5.7a will efficiently couple to the InP waveguide after the metal is removed.

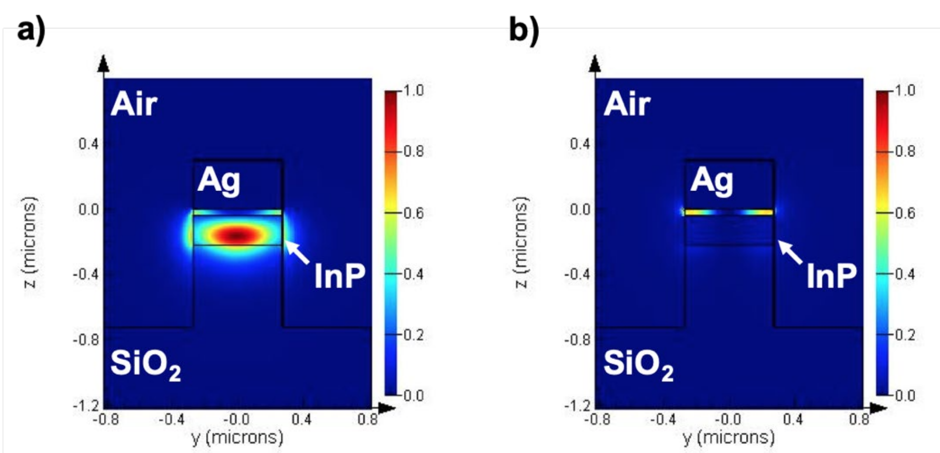

Figure 5.7. Potential modes under the metal. a) Waveguide mode that couples well to InP waveguide and b) waveguide mode that does not couple well to InP waveguide

From Maxwell's equations we know that at the boundary between a perfect conductor and dielectric the only electric field component that can exist is the normal component - the tangential component must be zero. As shown in Fig. 5.7b, our undesired modes have an electric field normal to the top metal contact and near the edge of the waveguide. We found that by wrapping metal around the waveguide sidewalls and tapering the waveguide width down to 300nm we can cancel the undesired mode since the electric field would now be tangential to the waveguide sidewall. We can now calculate the modes for the new structure and see that only one mode can propagate under the metal, shown in Fig 5.8.

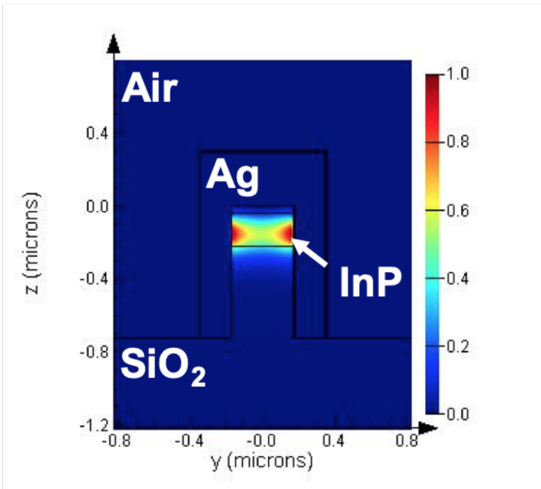

Figure 5.8. Mode in waveguide when metal is wrapped around the waveguide sidewalls. This mode is the only mode that exists at this wavelength.

While the overlap for this mode is lower, it is the only mode able to propagate under the metal so the waveguide coupling efficiency is increased. Using this structure, we can achieve a waveguide coupling efficiency of 90%.

5.2.5 COUPLED CAVITIES

In the previous subsection we did a simple taper of the waveguide due to the ease of simulating the structure; however, that geometry is not the only one that could provide high waveguide coupling efficiency. In order to explore a larger design space for the shape, we used the adjoint method [3,4,5], commonly referred to as inverse design—see Chapter 6 for more information on this method. By running inverse design optimization on our structure, we came up with the following shapes, shown in Fig. 5.9.

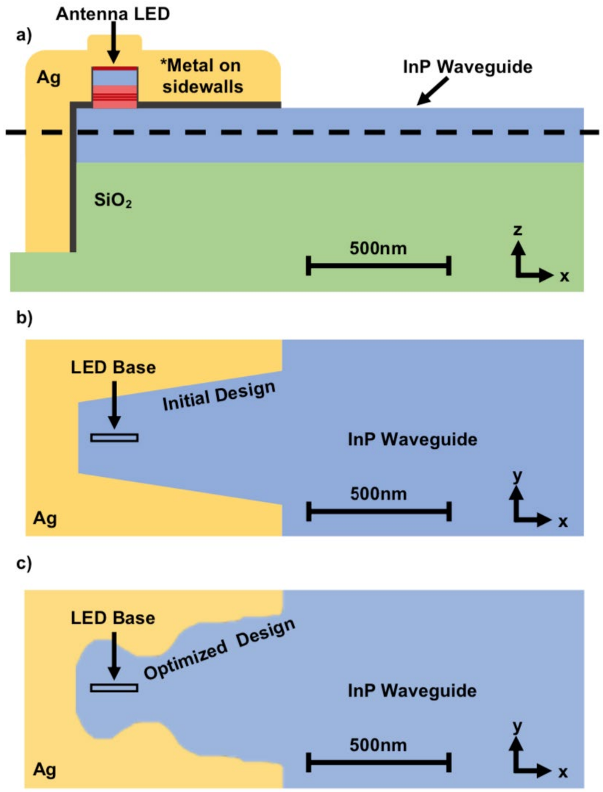

Figure 5.9. a) Cross section schematic (XZ) of tapered waveguide coupler showing dashed cutline, and b) top view XY cross section of waveguide along dashed cutline. c) XY cross section of coupler after optimization, showing perturbations to Ag-InP boundary. Note b) and c) also show the projection of the LED base. Printed with permission from [7].

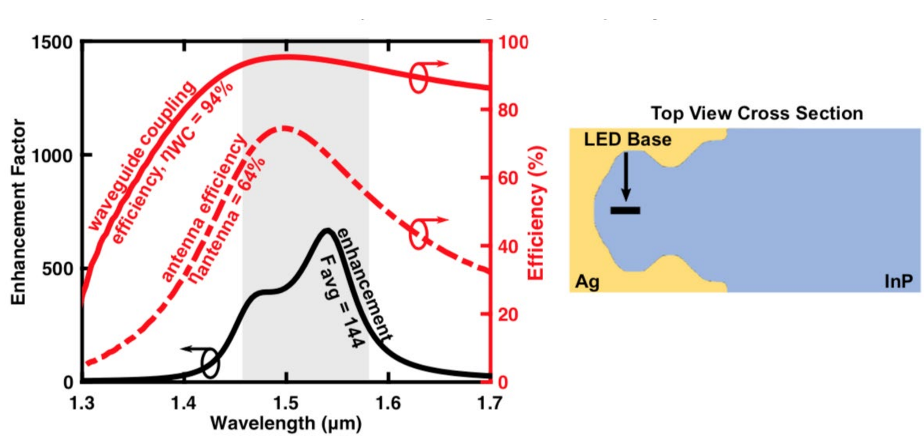

Figure 5.10. Enhancement, antenna efficiency, waveguide coupling efficiency spectra and top view XY cross sections for multi frequency optimization. For reference, the LED material spectrum [L(ω)] between its 50% power points is shown by the gray shaded region. Printed with permission from [7].

In Fig. 5.10, we see that the enhancement spectra for this sample shows two distinct peaks. This can be explained by thinking of the antenna LED and coupler section (see inset Fig. 5.11(a)) as coupled resonators. When they have the same resonance frequency, it will lead to a frequency split that can be observed in the enhancement spectra. This was confirmed by sweeping the LED length in the multi frequency design, which resulted in an avoided crossing between the antenna LED resonance and the coupler section resonance, as shown in Fig. 5.11. The dashed black line was generated by sweeping the length of the antenna LED on a bulk InP substrate. The dashed green line was created by placing a dipole in the coupler section (see inset) and sweeping the length of an off-resonance antenna LED. During the length sweep we found that the antenna efficiency always peaked at the coupler section resonance.

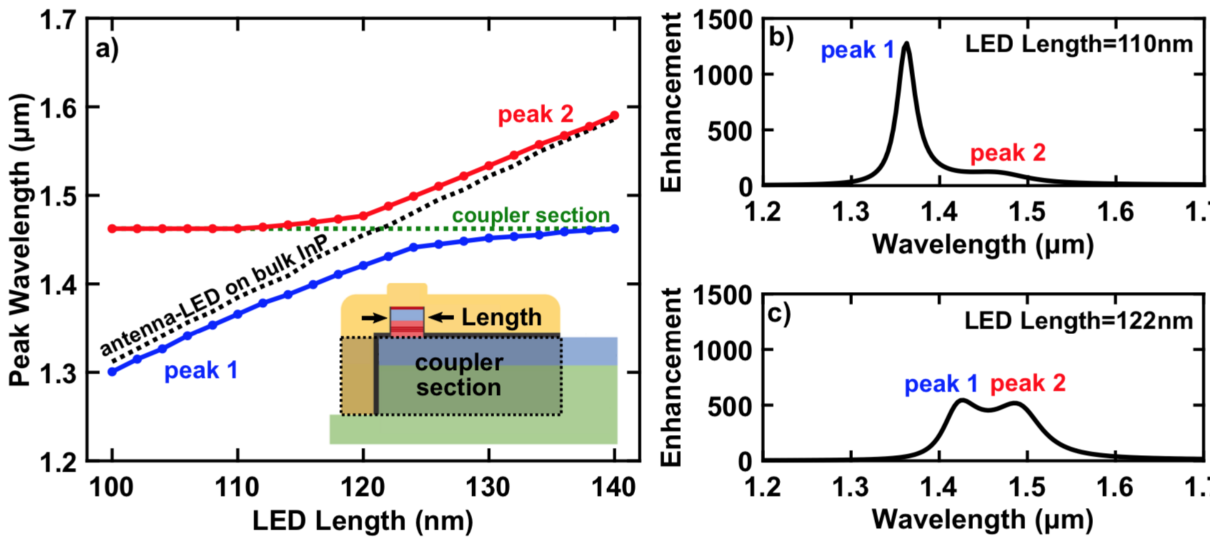

Figure 5.11. a) Avoided crossing between the optical antenna resonance and the inverse design coupler resonance. For reference, dashed black and green lines show independent resonances of the antenna-LED on a bulk InP substrate and the coupler section as a function of LED length, respectively. Enhancement spectra for LED lengths of b) 110nm and c) 122nm. Printed with permission from [7].

Essentially, the spectra of the waveguide coupling designs can be explained by considering the antenna LED and the coupler section as coupled resonators. When the resonances are tuned, we have an impedance match and frequency splitting. Due to the impedance match, the optical power is able to quickly leave the lossy antenna LED (lower Q factor) resulting in less metal loss (higher antenna efficiency). A similar conclusion was reached in [6], where detuned resonators were exploited to achieve higher peak enhancement. Note that this design also yielded a higher waveguide coupling efficiency than the tapered coupler at 94%.

5.3 EXPERIMENTAL METHODS

We fabricated the tapered structure; in Fig 5.12 the sample was imaged immediately after metal liftoff and before substrate removal to form the waveguide.

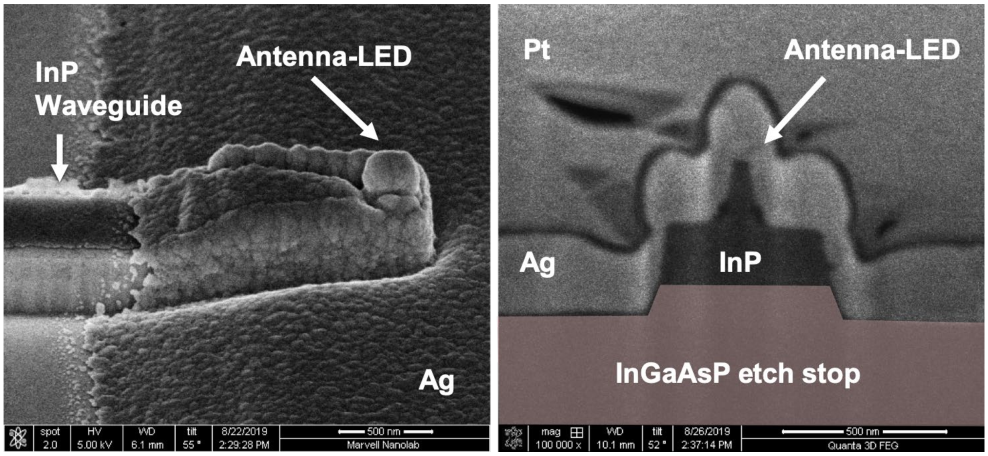

Figure 5.12. Scanning electron micrographs of the sample after metal liftoff, but before substrate removal. Left shows tilt view of the tapered waveguide coupler and the InP waveguide. Right shows a focused ion beam cross section of the waveguide and antenna-LED width.

In simulations we can place monitors to know exactly where power is going, which lets us get an exact number for the waveguide coupling efficiency. However, after fabricating the device we can only measure a fraction of the light. In Fig 5.13, we show the two main sources of light collected on the air-side using red arrows, labeled not-coupled and grating. Grating couplers are commonly used to direct light out of the waveguides by using interference, the light collected from the grating coupler gives us an estimate of the amount of light that was coupled to the waveguide. Likewise, we can collect the light that propagates straight down, which is labeled not-coupled.

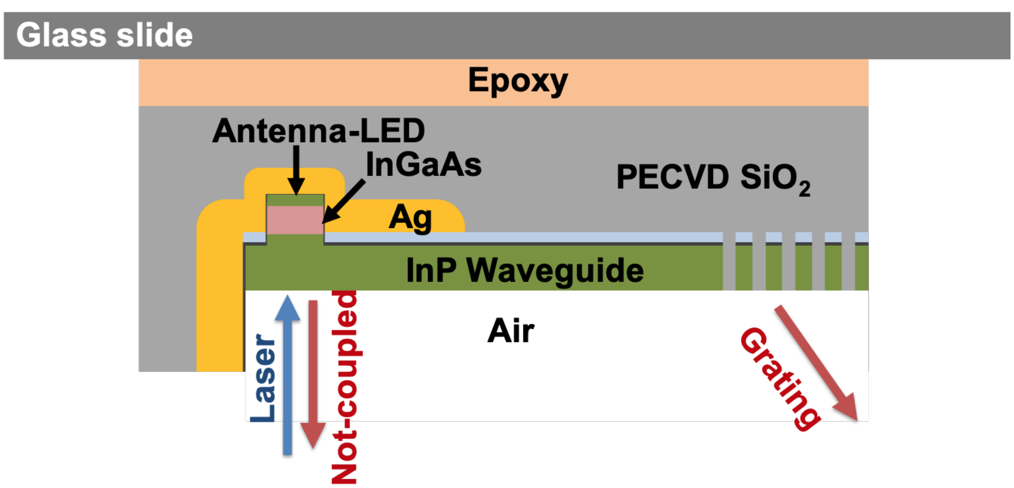

Figure 5.13. Cross section schematic of the final device, showing laser pumping in blue and light collection from not-coupled counts and grating in red.

In Fig 5.14a, we show an optical microscope image of the fabricated device, with the antenna-LED and grating labeled. In Fig 5.14b, the emission profile from the device is overlaid as a heatmap on the dark field image. Likewise, the 1D counts along the waveguide is overlaid as a solid blue line. We can take the ratio between the light coming from the grating to the total light collected from this side to get an estimate of the waveguide coupling efficiency, which we found to be 85.9%.

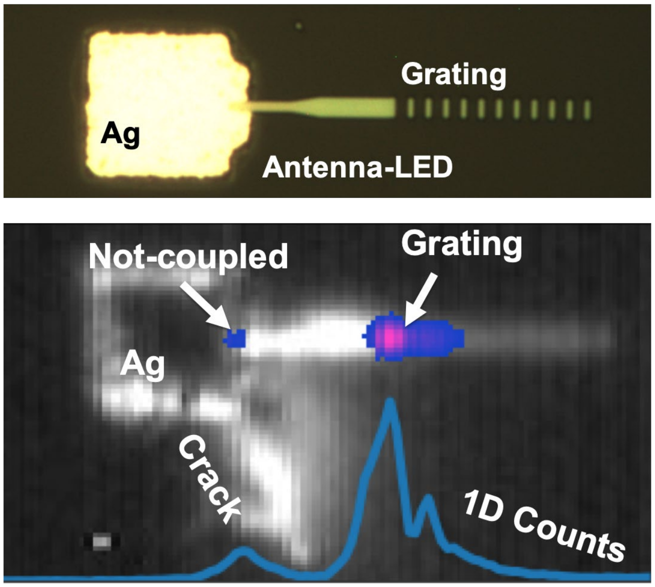

Figure 5.14. a) Optical microscope image of fabricated device after substrate removal and b) dark field image overlaid with heatmap of emission and solid blue line showing counts along 1D pixel line.

This is a promising proof of concept of our waveguide coupling methodology; however, more work will need to be done to fully characterize the waveguide coupling efficiency and create a full link.

5.4 CONCLUSION

In this chapter we explored using destructive interference, the image theory, and mode shaping to manipulate electromagnetic fields to couple the radiation from the cavity backed slot antenna-LED into a single mode waveguide. Additionally, we found that by using coupled cavities we can optimize the structure for either high enhancement or high efficiency. The concepts used in this chapter could be applied towards manipulating the fields from other single mode light sources.

REFERENCES

[1] “Overlap analysis - Modal Analysis Tab,” *Lumerical Support*. [Online]. Available: http://support.lumerical.com/hc/en-us/articles/360034917453-Overlap-analysis-Modal-Analysis-Tab.

[2] M. S. Eggleston and M. C. Wu, “Efficient Coupling of an Antenna-Enhanced nanoLED into an Integrated InP Waveguide,” *Nano Lett.*, vol. 15, no. 5, pp. 3329–3333, May 2015, doi: 10.1021/acs.nanolett.5b00574.

[3]C. M. Lalau-Keraly, S. Bhargava, O. D. Miller, and E. Yablonovitch, “Adjoint shape optimization applied to electromagnetic design,” *Opt. Express, OE*, vol. 21, no. 18, pp. 21693–21701, Sep. 2013, doi: 10.1364/OE.21.021693.

[4]A. Y. Piggott, J. Petykiewicz, L. Su, and J. Vučković, “Fabrication-constrained nanophotonic inverse design,” *Scientific Reports*, vol. 7, no. 1, pp. 1–7, May 2017, doi: 10.1038/s41598-017-01939-2.

[5] A. Michaels and E. Yablonovitch, “Leveraging continuous material averaging for inverse electromagnetic design,” *Opt. Express, OE*, vol. 26, no. 24, pp. 31717–31737, Nov. 2018, doi: 10.1364/OE.26.031717.

[6] H. M. Doeleman, E. Verhagen, and A. F. Koenderink, “Antenna–Cavity Hybrids: Matching Polar Opposites for Purcell Enhancements at Any Linewidth,” *ACS Photonics*, vol. 3, no. 10, pp. 1943–1951, Oct. 2016, doi: 10.1021/acsphotonics.6b00453.

[7] N. M. Andrade, S. Hooten, S. A. Fortuna, K. Han, E. Yablonovitch, and M. C. Wu, “Inverse design optimization for efficient coupling of an electrically injected optical antenna-LED to a single-mode waveguide,” *Opt. Express, OE*, vol. 27, no. 14, pp. 19802–19814, Jul. 2019, doi: 10.1364/OE.27.019802.

Chapter 6: Novel Optical Design

Sean Hooten

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

6.0 INTRODUCTION

This chapter will discuss some advanced concepts in electromagnetic design that are being explored in E3S Theme III research. Section 6.1 will discuss the theoretical design of an ultra-high enhancement antenna-LED. Section 6.2 will discuss advanced concepts in computational electromagnetics, including methods for designing electromagnetic devices. This chapter is intended to be conceptual without dwelling on mathematical rigor. It will require an understanding of the concept of electromagnetic simulation as introduced in Chapter 2, and the antenna-LED introduced in Chapter 3.

6.1 ADVANCED TOPICS IN SPONTANEOUS EMISSION ENHANCEMENT

In the previous chapters the antenna-LED was introduced and described in detail. In this section we will explore the fundamental physics of spontaneous emission and the antenna-LED from a slightly different perspective. In doing so, we will show how ultra-high enhancement and efficient antenna-LEDs can theoretically be produced.

6.1.1 QUANTUM MECHANICAL ORIGIN OF SPONTANEOUS EMISSION

Interestingly, both stimulated emission and absorption do not require any quantum mechanical treatment and are compatible with classical conceptions of physics. Spontaneous emission, on the other hand, cannot be explained classically. This is because spontaneous emission happens *spontaneously* with seemingly no external perturbation. However, if the problem is explored from a quantum mechanical point-of-view, we find that there actually is an external perturbation known as the zero-point electromagnetic field. The meaning of the zero-point electromagnetic field is that any given volume of space has a finite amount of electromagnetic energy even when it is seemingly empty (e.g. contains no light or static fields). The implications of this fact are enormous, potentially explaining phenomena beyond spontaneous emission such as the accelerating expansion of the universe (also called "dark energy") [1]. As a consequence, it can be argued that spontaneous emission is the stimulated emission due to zero-point photons [2].

Using this theory, the rate of spontaneous emission of a system can be derived. The rate can be found using an equation called Fermi's Golden Rule, and has the following form,

$$1/\tau_{sp} = \frac{2\pi}{\hbar}|qx \cdot E_{zpf}|^2 \rho_f$$

where τ_{sp} is the spontaneous emission lifetime, $\hbar$ is Planck's constant, q is the elementary charge, x is the dipole moment of the emitting material (usually the size of a crystal unit cell in a semiconductor), E_{zpf} is the electric field due to zero-point electromagnetic fluctuations, and ρ_f is known as the optical density of states. For brevity, we will not describe each of these quantities in detail, but some immediate intuition about spontaneous emission can be gained from viewing this equation: the rate of spontaneous emission can be increased by increasing the dipole moment, the zero-point electric field, or the optical density of states. For the purposes of this chapter, the dipole moment and the optical density of states can be considered constant and fundamental to a material. The zero-point field, however, can be tuned through proper engineering of an electromagnetic device. The concentration of the zero-point electric field is often given in terms of a quantity called effective volume,

$$V_{eff} = \frac{\int_V \quad \varepsilon E^2}{\varepsilon {E_p}^2}$$

where the numerator gives the total electromagnetic energy in the antenna mode and the denominator gives the energy density at the peak of the mode. Thus, the smaller the effective volume, the larger the zero-point electric field. The next subsection will detail how the zero-point electric field can be concentrated (and hence the effective volume of the device can be reduced).

6.1.2 METAL-DIELECTRIC ANTENNA

In Chapter 2, we briefly explained how spontaneous emission enhancement increases as the metal gap in an antenna decreases. This is repeated below for a dipole antenna in Figure 6.1.

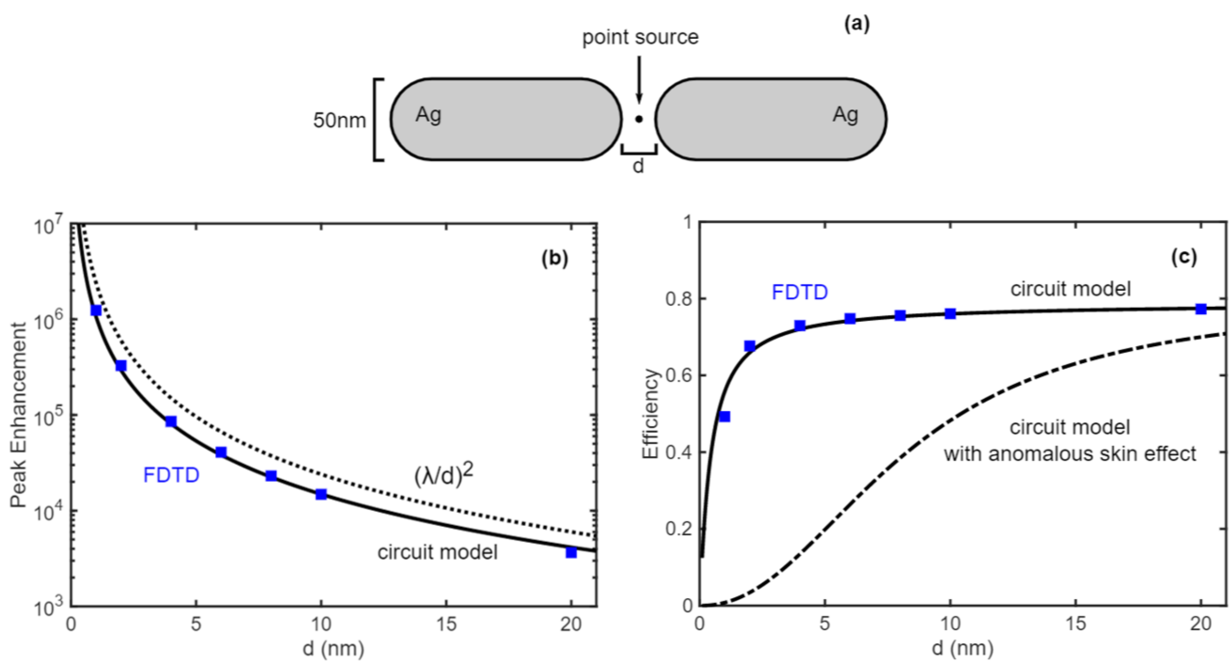

Figure 6.1. The enhancement and efficiency of a dipole antenna. (a) The antenna consists of two metal rods separated by a gap, d, where an optical source is placed. (b) The enhancement of the antenna follows a quadratic dependence on the gap. Ultra-high enhancement can be obtained for small d. (c) The efficiency of the antenna follows the reverse trend, tending towards 0 as the gap is decreased.

As the metal gap decreases, the enhancement of the dipole antenna increases quadratically. However, in Figure 6.1(c), we find that the efficiency of the antenna trades off with enhancement, ultimately becoming impractical below a gap of about 10nm. Thus, the maximum enhancement becomes capped at a value of 10^4(note that the enhancement is so large in this case because the antenna emits directly into free space, the enhancement is significantly lower when emitting into a semiconductor).

As indicated in the last section, enhancement increases with the magnitude of the zero-point electric field. This is exactly the reason why the enhancement increases with decreasing metal gap size: as the metal surfaces get closer and closer, the electric field in the antenna gets more and more concentrated near the point source. Is there a way to further increase the electric field concentration without negatively affecting the antenna efficiency?

Recent research in this area has shown that the electric field concentration can be increased dramatically by using pointy dielectric tips [3,4]. Dielectrics are essentially lossless, so this field concentration comes nearly for free (at the cost of difficult fabrication). By leveraging this technology with the metallic antenna, the enhancement can be greatly improved. The structure of this metal-dielectric antenna, in comparison to the regular metallic dipole antenna, is shown in Figure 6.2.

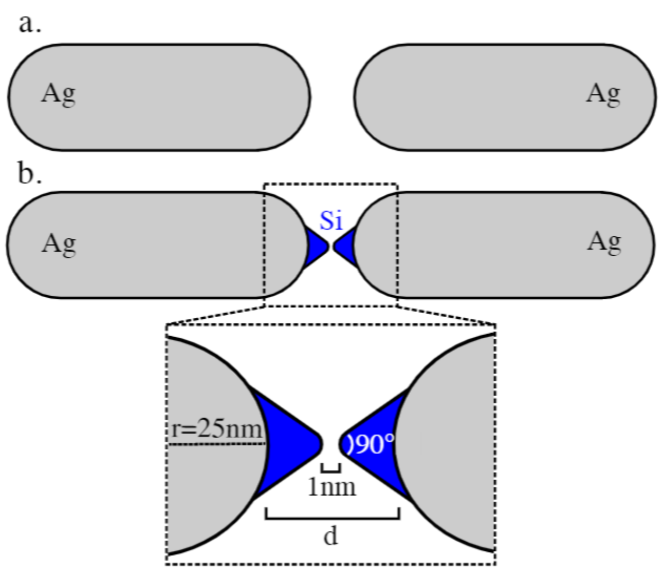

Figure 6.2. (a) A metal dipole antenna compared to a (b) metal-dielectric dipole antenna. The metal-dielectric antenna uses pointy dielectric tips in the vicinity of a source.

The dielectric tips are separated by just 1nm in proximity to a source. The metal gap distance, d, is varied. Figure 6.2, shows the enhancement antenna as d is varied.

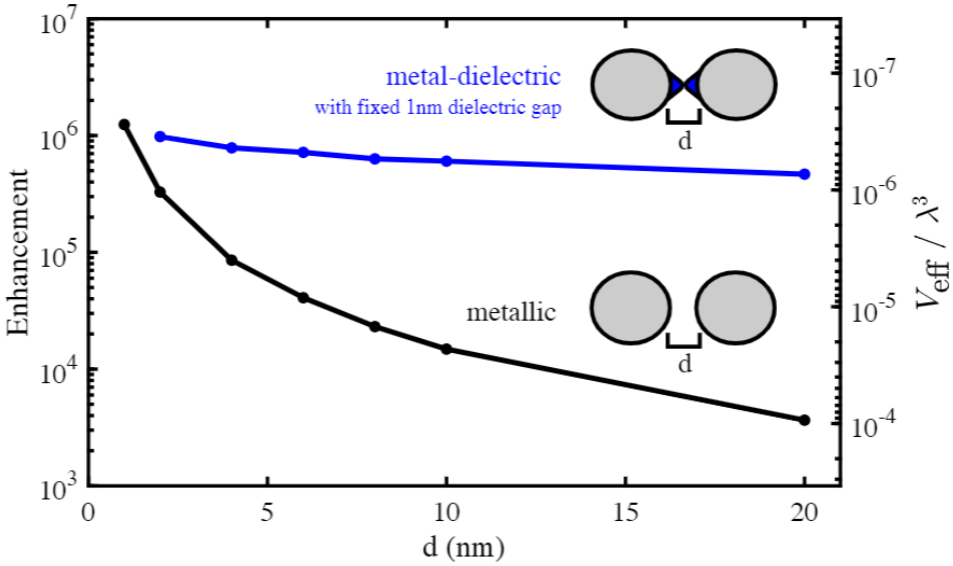

Figure 6.3. Enhancement versus metal gap distance of the all metal antenna and metal-dielectric antennas from Fig. 6.2. Even for large metal gap, d, the metal-dielectric antenna maintains a large enhancement without any compromise to efficiency.

As a result of the dielectric tips, the metal-dielectric antenna maintains a large enhancement even for large metal gaps. The efficiency of the metal-dielectric antenna looks similar to the plot given in Fig. 6.1(c), so large efficiency and enhancement values can be obtained simultaneously. The electric field concentration of the metal-dielectric antenna can be seen in Fig. 6.4, where the electric field of the metallic and metal-dielectric antennas are compared for two different metal gaps in a zoomed-in picture of the gap. Also shown here is a comparison of the effective volumes of the antenna, revealing that the antenna mode is extremely dense for the metal-dielectric antenna. Thus, ultra-high enhancement and efficiency antennas can be obtained by leveraging the use of dielectric structures.

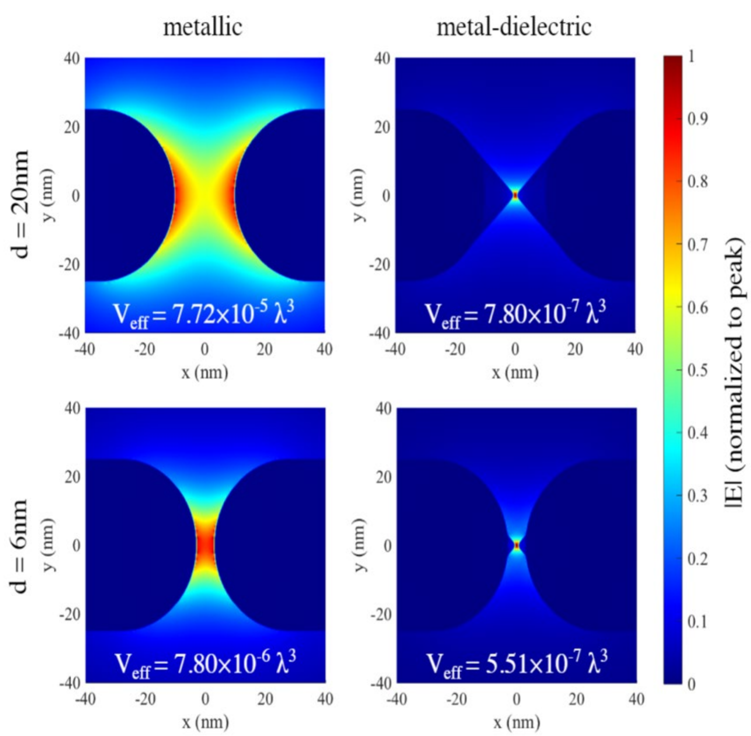

Figure 6.4. Comparison of the electric field concentration in metallic and metal-dielectric antennas for two different gaps, d=20nm and 6nm. The metal-dielectric antenna maintains strong concentration even for a large metal gap.

6.2 ELECTROMAGNETIC INVERSE DESIGN

In Chapter 2, the concept of electromagnetic simulations was introduced. Using simulations, electromagnetic devices can be designed efficiently when combined with physical intuition. This was exemplified in Chapter 3, where we showed the design of an antenna-LED coupled to a single-mode waveguide. In this section we will go into more detail about how electromagnetic devices can be designed. In doing so, we will introduce an advanced tool in electromagnetic design known as the Adjoint Method (also known as Inverse Design). This section will assume an understanding of calculus, Maxwell's equations, and geometric optics.

6.2.2 TYPES OF ELECTROMAGNETIC SIMULATIONS

There is not any single ubiquitous tool for the design and simulation of all optical devices. By making certain assumptions, one can greatly simplify a computational problem and hence increase the computational efficiency of solving it. For example, when designing macroscopic lenses (e.g. glasses or camera lenses), a full-fledged solution of Maxwell's equations is not required. In this situation, it can be assumed that incident light consists of incoherent plane waves, or rays of light, and a correct solution can be found by using known empirical laws of optics such as Snell's Law. Of course, by making assumptions, the generalizability of the simulation method decreases, and potentially the accuracy of the solution. An engineer must choose a computational method carefully to maximize the tradeoff of accuracy and computational efficiency.

Figure 6.5, shows a hierarchy of methods in optics and electrodynamics. The larger the circle, the more general but also more likely to be computationally difficult. Below each method gives typical simulation tools. Each tool will not be described in detail; the reader is encouraged to research the methods using the internet if desired.

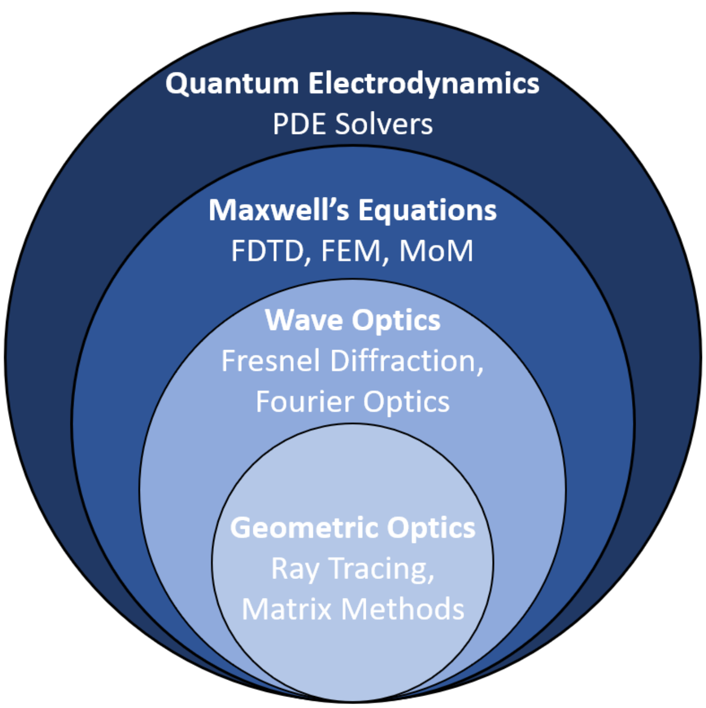

Figure 6.5. Hierarchy of solving optics and electrodynamics problems. The larger the circle, the more general, but the more difficulty of obtaining an accurate solution.

Using these tools, the solution to an electromagnetics problem can be obtained. However, these solutions assume that the geometry of an electromagnetic device that an engineer is attempting to design has already been defined. How does an engineer decide on an initial design, and subsequently tune it for the needed specifications?

6.2.3 PARAMETER SWEEP

The most conventional way to design electromagnetic devices is via parameter sweep. The steps to perform a parameter sweep consist of the following:

1. Choose a suitable simulation tool that includes the relevant physics for the problem.
2. Define a figure of merit or specification that determines how "good" the device is.
3. Choose an initial design based on physical intuition of the problem.
4. Choose a few relevant design parameters to vary, e.g. geometric features such as the length and width of a device.
5. Simulate the design for each design parameter, and evaluate the figure of merit.

Assuming that the design parameter is varied finely enough, a local optimum in the figure of merit will be found (though there is no guarantee that it will be the "best possible" device, i.e. a global optimum).

To help the reader in understanding, we will perform a toy example of a parameter sweep. In this case, we will tune the resonance of a simple dipole antenna to radiate at a wavelength of 1550nm. We outline the steps below:

1. In this case we will simulate a full solution to Maxwell's equations using the finite-difference time-domain (FDTD) method.
2. The figure of merit will be the total amount of power radiated from the antenna at a wavelength of 1550nm (transmission).
3. The dipole antenna consists of two cylindrical silver arms with radius of 25nm, and a gap of 30nm where a current source is placed. The dipole antenna will be clad with glass in all directions. The total length is initially chosen to be half the wavelength in the cladding material ($\lambda/2n = 554nm$), which corresponds to the length of an ideal standing wave for the fundamental antenna resonance. The geometry of the initial design is shown below in Fig. 6.6.
4. We vary the length of the antenna in 20nm steps.
5. A plot of the output optical power (transmission), versus antenna length is below in Fig. 6.7. The optimum length is found to be 480nm.

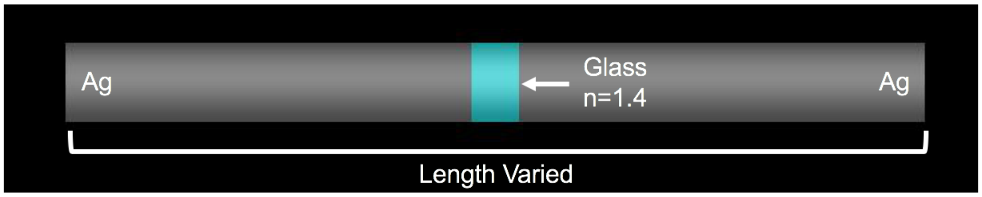

Figure 6.6. Ag antenna with variable length, and a 30nm glass gap. The whole antenna is clad in glass as well.

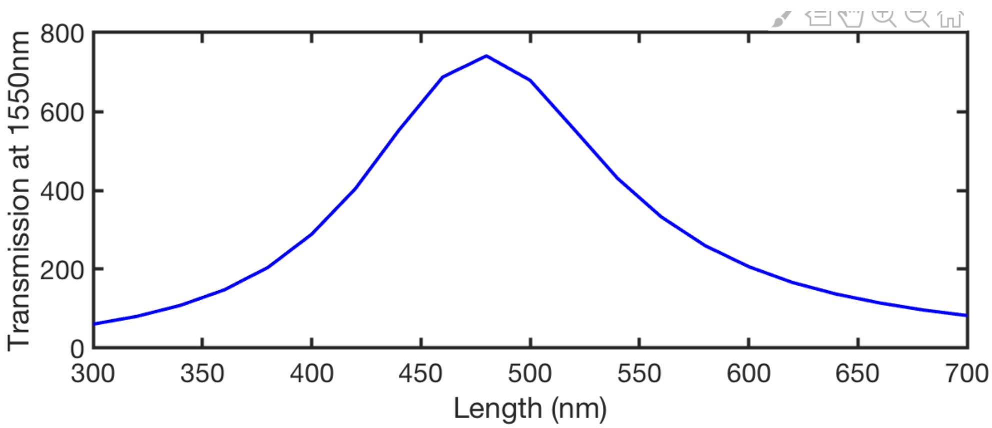

Figure 6.7. Result of the parameter sweep of the antenna length. The optical power emitting at a wavelength of 1550nm peaks at an antenna length of 480nm.

As can be seen here, the initial guess for the optimum length was incorrect. This is because (1) the derivation of that length assumes an antenna with no gap and small thickness, and (2) metals do not act ideally at optical frequencies, which causes the resonance to redshift with respect to the ideal. Thus, we see how parameter sweeps can be a useful tool for tuning and designing electromagnetic devices by applying these concepts more generally. Furthermore, parameter sweeps can be very useful when studying the behavior of experimental devices. Indeed, parameter sweeps can be compared to experimental sweeps of the same parameters, finding correlations and contrasts between expected simulated behavior and experimental behavior, which can be further explored.

6.2.4 THE ADJOINT METHOD

As shown in the previous section, electromagnetic devices can be designed by using physical intuition and smartly varying geometric parameters. But what about when devices become very complex and consist of 10s or 100s of geometric parameters? In that case, performing parameter sweeps is a less attractive or even infeasible option because the computational time required increases exponentially with the number of parameters [5]. Fortunately, there are emerging design tools which allow for the efficient optimization of electromagnetic devices. The general names for these tools are “inverse design” or “topological optimization”, but here we will discuss the Adjoint Method specifically.

The Adjoint Method is a gradient descent based algorithm. This means that the method uses information about the gradient (derivative) of the figure of merit (specification) with respect to the design parameters. This can be illustrated in Figure 6.8 where we show the difference between a Figure 6.8(a) parameter sweep based approach and a Figure 6.8(b) gradient descent based approach. In both plots, the figure of merit, f, is plotted versus a parameter that is being varied, p. This is similar to the plot from Figure 6.7 above, except we see that in general the figure of merit can have a bumpy and non-ideal behavior as the parameter changes.

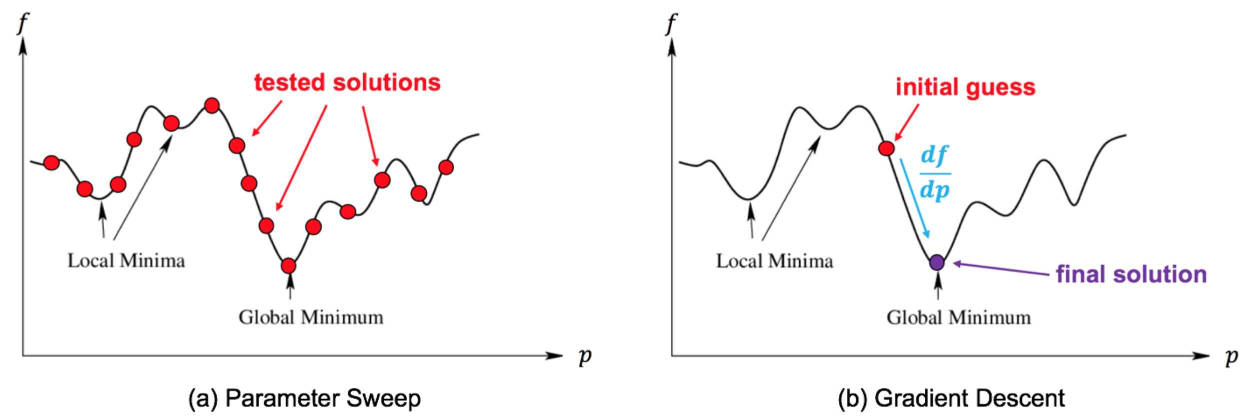

Figure 6.8. Comparison between (a) parameter sweep and (b) gradient descent. The figure of merit can have an unintuitive shape with any given parameter as shown by the black curve. A parameter sweep tries to sample the entire space of the design parameter to find the optimal solution. Gradient descent tries to find the optimal solution by using the gradient of the figure of merit with respect to the design parameters.

During a parameter sweep, we take as many samples of the parameter as possible to attempt and reconstruct the topology of the parameter space. By taking enough samples, a good solution or even the global minimum can be found. However, this requires many samples to be taken, and in general there may be many more parameters as discussed previously. The gradient descent approach is much more efficient. By choosing a good initial guess, the gradient of the figure of merit tells the user in which direction it is favorable to change the parameters. By successively applying this approach, the gradient will eventually guide the algorithm to an optimum. Since this approach uses the gradient (not just a single derivative), the algorithm can explore a large parameter space without increasing the computational time dramatically. The caveat of this approach is that it requires the choice of a good initial guess, and if this is not done correctly then the optimizer can be caught in an unfavorable local optimum. Thus, the engineer using this algorithm still needs to have a strong understanding of the problem at hand.

At this point the reader may ask how one obtains the gradient of the figure of merit with respect to the parameter space. This is really the heart of the Adjoint Method, but it is also outside of the scope of this work due to the mathematical rigor that is needed for understanding. The reader is encouraged to see the references in [6,7] for more information. Nevertheless, we will indicate the exciting conclusion here: the Adjoint Method allows one to calculate the gradient of the figure of merit with respect to any arbitrary amount of design parameters with only 2 simulations! This is a significant result, because it means that huge and complex simulations with many design parameters can be optimized very efficiently. In fact, this method was applied in the optimization of the waveguide coupler from Chapter 5. Figure 6.9 shows the design of the waveguide coupler before and after optimization via the Adjoint Method.

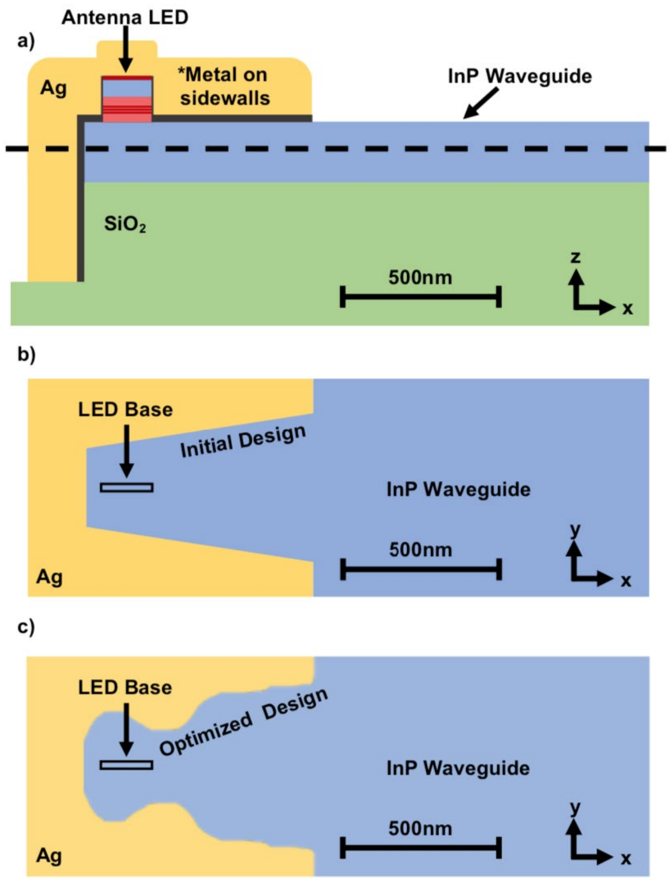

Figure 6.9. Inverse design of the waveguide coupler from Chapter X. (a) Side-view of the waveguide coupler. (b) Top view of the waveguide coupler from the indicated slice in (a) before optimization. (c) Top view of the waveguide coupler after inverse design optimization using the adjoint method. Printed with permission from [8].

Figure 6.9(a) shows the side view cross section of the coupler, while Figure 6.9(b)-(c) shows a top view cross section of the coupler before and after optimization. As can be seen here, the shape of the original tapered coupler was optimized in a somewhat unintuitive fashion to ultimately achieve a door-handle like shape. In the case of this optimization, the parameters that were changed was the interface between Ag and InP, which consists of many vertices. Such an optimization would be impractical using a conventional parameter sweep. Thus, we see how the adjoint method can be practically employed to design advanced electromagnetic devices.

REFERENCES

[1] L. Lombriser, “On the cosmological constant problem,” Phys. Lett. B 797, 134804 (2019).
[2] C. Cohen-Tannoudji, “Fluctuations in radiative processes,” Phys. Scripta T12, 19–27 (1986).
[3] H. Choi, M. Heuck, and D. Englund, “Self-Similar Nanocavity Design with Ultrasmall Mode Volume for Single-Photon Nonlinearities,” Phys. Rev. Lett. 118, 223605 (2017).
[4] S. Hu and S. M. Weiss, “Design of Photonic Crystal Cavities for Extreme Light Concentration,” ACS Photonics 3, 1647–1653 (2016).
[5] C. M. Lalau-Keraly, S. Bhargava, O. D. Miller, and E. Yablonovitch, “Adjoint shape optimization applied to electromagnetic design,” Optics Express, vol. 21, pp. 21693–21701, Sept. 2013.
[6] A. Michaels and E. Yablonovitch, “Leveraging continuous material averaging for inverse electromagnetic design,” Optics Express, vol. 26, pp. 31717–31737, Nov. 2018.
[7] S. Molesky, Z. Lin, A. Y. Piggott, W. Jin, J. Vuckovic, and A. W. ´ Rodriguez, “Inverse design in nanophotonics,” Nature Photonics, vol. 12, pp. 659–670, Nov. 2018.
[8] N. M. Andrade, S. Hooten, S. A. Fortuna, K. Han, E. Yablonovitch, and M. C. Wu, "Inverse design optimization for efficient coupling of an electrically injected optical antenna-LED to a single-mode waveguide," Optics Express, vol 27, pp. 19802-19814, 2019.