1 Introduction

The infrastructure that supports modern society, including health care, education, transportation, finance, and scientific and technological research, has become inextricably tied to the continuous progress in the ability to generate, transmit, receive, and process information. While electronic devices integrated in highly complex circuits have enabled this progress, electronic devices and circuits suffer from inherent limitations, namely resistor-capacitor (RC) time delays [1]. To this end, photonic devices and circuits are envisioned to complement and perhaps, eventually, supplant electronics as the enabling technology for continuous progress in the collection, transmission, and processing of information. In all photonic systems, an essential component is the light source. Optical sources with coherent emission of light, in the form of light amplification by stimulated emission of radiation (LASER), were first demonstrated in 1960 [2]. Like the first transistors, the first lasers were macroscopic devices, with footprints on the order of centimeters to decimeters.

As the cavity size is reduced with respect to the emission wavelength, interesting physical effects, unique to electromagnetic cavities, arise. Experiments in the radio and microwave frequencies first demonstrated that the spontaneous emission rate of atoms in a cavity could be enhanced or inhibited, relative to their rate of emission in free space. The change in the spontaneous emission rate was found to depend on the geometry of the cavity as well as the orientation and spectra of the atoms [3]. In the 1990s, the first proposals and experiments were made to extend what has since become known as the Purcell effect to the optical regime. Modified spontaneous emission rates in dyes and semiconductors in optical microcavities were observed [4–7], and significant applications of the Purcell effect were reported, including diode lasers with greater modulation bandwidth [8–10], energy efficiency [11, 12], and absence of a threshold [13, 14]. While the concept of thresholdless operation continues to be a subject of debate [15, 16], the modulation and efficiency improvements enabled by wavelength-scale cavities are fairly well understood. For example, with proper design, the cavity of a subwavelength laser may be designed such that most of the spontaneous emission is channeled into the lasing mode [14, 17]. In so doing, unwanted emission into non-lasing modes is mitigated, and the below-threshold efficiency is limited only by non-radiative recombination. Since the observation of the Purcell effect in semiconductor cavities, lasing has been demonstrated in numerous wavelength and subwavelength scale structures. These structures include dielectric microdisks [18–21], photonic crystals [11, 22–25], nanowires [26, 27], nanomembranes [28–30], micro-pillars [31–33], and metal-clad nanocavities [14, 34–38].

1.1 The History of Laser Minimization

To complement micro- and nano-electronics, optical components have undergone a process of miniaturization over the past several decades. Thus, it is not an exaggeration to state that the maintenance and improvement of modern society is directly related to research advancements in photonic devices and circuits. Similar to transistors, the reduction of the size of lasers would enable a higher packing density of devices and lower power consumption per device. The first laser miniaturization came with the invention of solid-state laser diodes [39] in which the device size was reduced from meter to millimeter scale. The invention of vertical-cavity surface-emitting laser (VCSEL) [40] enabled even further miniaturization down to tens of micrometers. More recently, micro-scale whispering-gallery mode (WGM) lasers were achieved in micro-pillars/disks [18, 41] and micro-spheres [42]. In parallel efforts, advancements toward optical mode miniaturization came with the use of 2D photonic crystals in laser designs [22]. Although the optical mode sizes of these lasers may be on the order of the emission wavelength, the entire structure is quite large due to the many Bragg periods. On the other hand, micro-pillars and microdisks do not necessarily suffer from the same problem. However, until recently, packing a large number of micro-pillar/disk lasers with dielectric cavities in a small region was impractical, because the modes of these lasers are poorly confined to the active regions and may extend well beyond the physical boundaries of the cavity, leading to undesirable mode coupling of neighboring devices [19, 33].

Figure 1.1 outlines the progress in laser miniaturization over the past few decades. We see from the development time line of small lasers that the evolution of a new laser device usually takes 10–20 years: from new laser concept to first proof-of-concept optically pumped demonstration, then to electrically pumped, and in some cases to commercial applications. For practical and commercial insertion of lasers, the devices need to be continuous-wave (CW) current injected at room temperature, ideally with stable emission, reasonably long lifetime and particular properties that the already established types of lasers cannot offer.

Figure 1.2 gives a quantitative comparison of the sizes of dielectric cavities in their miniaturized form and metal-clad cavities that were introduced in late 2000 (see Figure 1.1), along with typical cavity Q factor and lasing threshold. Each type of cavity or laser design has its advantage: for example, photonic crystal dielectric lasers with few-quantum-dot gain show extremely low threshold at cryogenic temperatures; dielectric cavities generally have higher Q factors than metal-clad cavities; and metal-clad cavities have much smaller dimensions than dielectric ones. In the context of subwavelength devices, one of the most figure-of-merit is size.

The size of an optical cavity can be defined using different metrics, for example, the physical dimensions of the cavity or the size of the optical mode. If the goal of the size reduction is to increase the integration density such as in a laser array, then the cavity size should account not only for the overall physical dimensions of the resonator but also the spread of the optical mode beyond the physical boundary of the resonator. Throughout

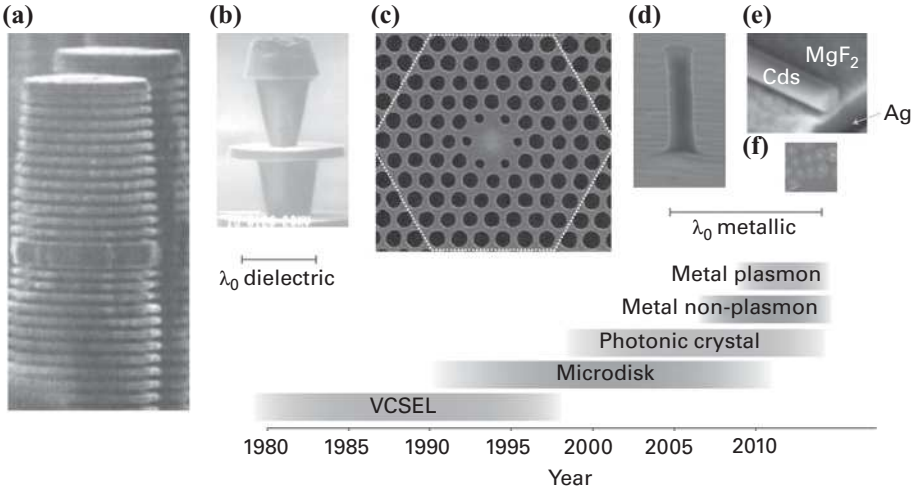

Figure 1.1 Development time line of small lasers, from first demonstration to electrical, continuous-wave, and room-temperature operation, and in some cases to commercial applications. Also shown is a size comparison of various types of small lasers. The electron microscopy pictures are scaled to the free-space emission wavelength λ_0 of each laser: (a) VCSEL, (b) microdisk laser, (c) photonic crystal laser, (d) metallic non-plasmon mode laser; (e) metallic propagating plasmon mode laser; (f) localized plasmon mode laser. The free-space wavelength scale bar of the metal-cavity–based lasers (d–f) is twice that of the dielectric lasers (a–c) to permit details to be seen. The metal-cavity lasers are typically smaller than λ_0 and dramatically smaller than corresponding dielectric-cavity lasers. Reprinted from reference [43] with permission from Macmillan Publishers Ltd.

this book, we define the subwavelength cavities following this metric. A desired nanolaser, therefore, should be smaller than the free-space emitted wavelength in all three dimensions, in terms of both the device's physical footprint and its optical mode confinement. Devices with such characteristics are essential for various practical applications including densely integrated chip-scale photonic circuits, displays, and sensors.

By this token, Figure 1.2(a), which lists laser's critical dimension and volume normalized to the free-space wavelength at which it emits, indicates that almost all dielectric laser cavities do not meet the metric because of either their large physical footprint or large mode volume. One of the smallest dielectric lasers is a microdisk laser with diameter smaller than its free-space emission wavelength [21]; however, it features a large effective mode volume V_{eff} (defined as $V_{eff} = V_a/\Gamma$, where V_a is the active region volume and Γ is the mode-gain overlap factor) because of poor mode confinement. Consequently, packing a large number of micro-pillar/disk lasers with dielectric cavities in a small region has been impractical.

On the other hand, both distributed Bragg resonators and photonic crystal cavities can be designed to have very localized energy distribution and thus very small effective mode volumes. However, tens to hundreds of Bragg layers or lattice periods are required to confine the mode and to maintain high finesse, resulting in physical footprints that are many wavelengths in size. Only since metal-clad nanolasers' inception in late 2000,

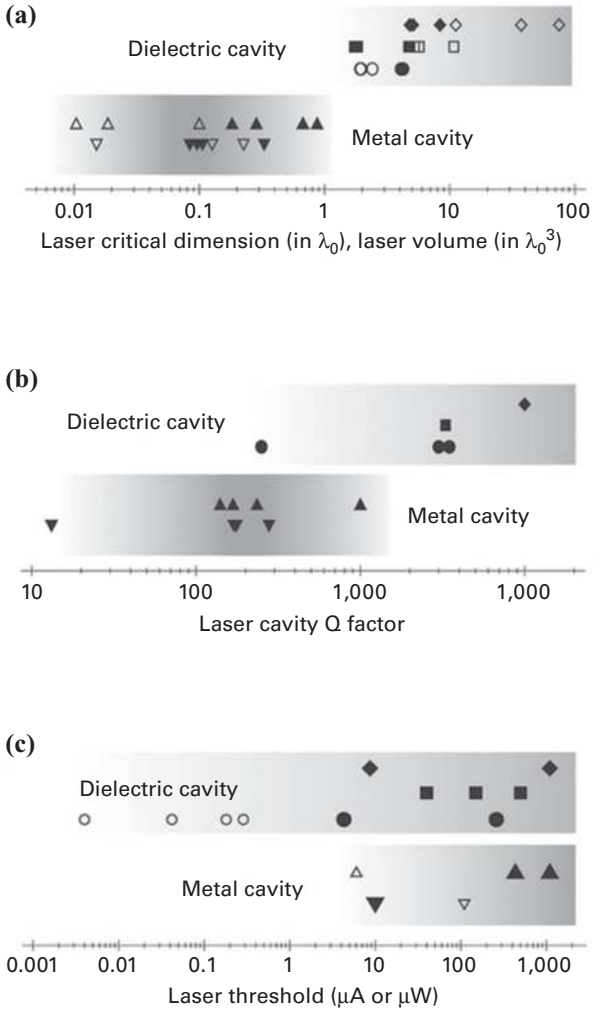

Figure 1.2 (a) Minimum extent of small lasers in one dimension (solid symbols) and minimum volume (open symbols) relative to the free-space wavelength λ_0. (b) Cavity Q factor and (c) lasing threshold in micro-watts for CW optical pumping and micro-amperes for pulsed or CW electrical pumping. In (b) and (c), open symbols denote cryogenic temperature operation; filled symbols denote room temperature operation. In (a–c), for the dielectric cavities, diamond denotes VCSEL, square denotes microdisk; circle denotes photonic crystal; for metal cavities, upward triangle denotes metallic non-plasmon mode; downward triangle denotes metallic plasmon mode. Reprinted from reference [43] with permission from Macmillan Publishers Ltd.

achieving both subwavelength physical footprint and mode volume has been possible, thanks to metal cladding's ability to strongly guide optical modes. Therefore, although we do not explicitly focus on metal-clad nanolasers in this book, we will inevitably see their reoccurrence throughout.

Figure 1.2(a) shows us that the incorporation of metal cladding has significantly reduced the size of the laser cavity. Here, we discuss in more detail how such

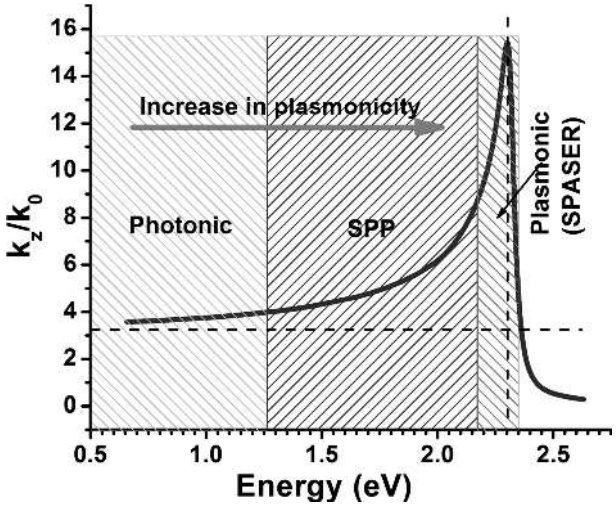

Figure 1.3 Dispersion relation at the interface of semiconductor and metal (in this case, silver), where z-axis denotes the wave propagation direction. The semiconductor's permittivity is taken to be 12, representing silicon or InGaAsP. Horizontal dashed line represents dispersion relation in the uniform semiconductor, and vertical dashed line denotes the surface plasmon resonance energy. Reprinted from reference [47] with permission from Springer Publishing.

incorporation affects the nature of optical modes in nanolasers. Optical modes in metal-clad cavities can be categorized into two regions: photonic mode and surface plasmonic polariton (SPP) mode. Furthermore, the SPP mode contains a sub-category of "spasing" mode. These modes' locations in the dispersion curve are shown in Figure 1.3. The photonic mode is the classical laser mode seen in all micro- and larger-scale lasers and some nanolasers. In this case, the mode predominantly resides in the gain medium, and the propagation constant is comparable to that in a pure dielectric cavity (horizontal dashed line in Figure 1.3). The metal acts as a prefect reflector and its role is purely to reduce mode penetration into the surroundings in order to increase mode confinement.

On the other hand, the SPP mode is a mode that propagates along the metal-dielectric interface in which the mode largely penetrates into the metal and the propagation constant is much larger than that in a dielectric cavity. Plasmonic effects have received immense attention since early 2000, being investigated as a means for nanoscale focusing and field enhancement. Even though it has always been known that the inherent challenge in plasmonic systems is the dissipation losses due to the constituent metal, the majority of studies in this area has focused on passive systems and does not account for the associated loss. A strategy to overcome this issue is to introduce optical gain in the dielectric constituent of the structures. Indeed, gain media can reduce the effective propagation loss and increase the SPP propagation length, enable transparent propagation, or even overcompensate propagation losses and lead to amplification and even stimulated generation of SPPs. Gain assisted lossless propagation in optical waveguide was first theoretically investigated by Nezhad et al. [44] and by Maier et al. [45], using semiconductor material as the gain medium. Both authors also envisioned the possibility of an SPP

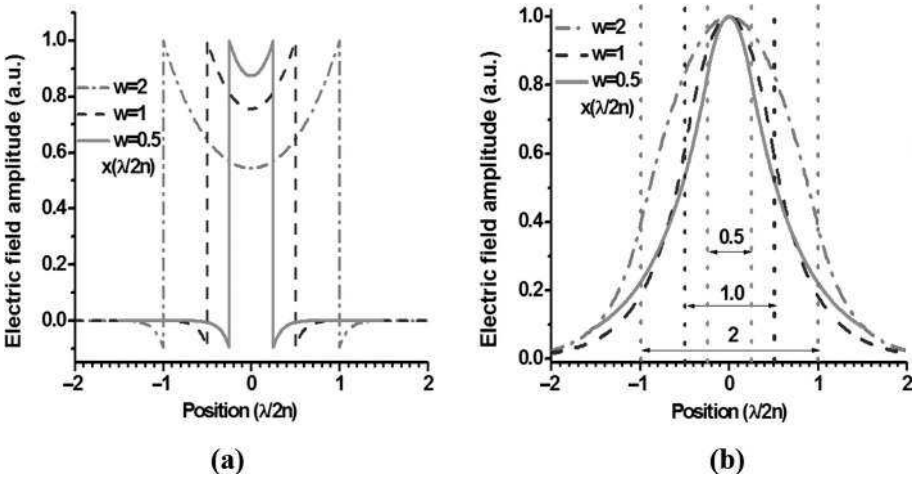

Figure 1.4 Mode profile (electric field amplitude) across the slab waveguide composed of (a) silver/semiconductor/silver and (b) air/semiconductor/air. Waveguide width w is normalized to $\lambda_0/2$ n with λ_0 = 1550 nm and n = 3.46 for III-V material. Reprinted from reference [47] with permission from Springer Publishing.

nanolaser in the case of overcompensation, if there is enough gain to compensate not only the propagation loss but also the radiation loss. The extreme case of the SPP mode is when the mode energy is close to the vicinity of the metal's SPP resonance (vertical dashed line in Figure 1.3), under which condition the mode's propagation constant as well as loss rapidly increase and approach maxima. The stimulated optical amplification under this condition is termed surface plasmon amplification by stimulated emission of radiation (SPASER), and the emission action is called "spasing" [46].

To illustrate the role metal plays in confining the optical mode, let's take a look at the fundamental mode's profile across a metal/dielectric/metal and an air/dielectric/air slab waveguide with varying dielectric core widths, shown in Figure 1.4(a) and (b), respectively. With the waveguide width w normalized to $\lambda_0/2$ n with λ_0 = 1550 nm, we see that with metal cladding, the mode is well confined in the dielectric region with all widths, even down to w = 0.5. With air cladding, however, the fundamental mode is most confined when the width exactly matches the mode wavelength at w = 1, but it becomes poorly confined when the width is further decreased. In fact, the optical confinement factor for both waveguide types is similar when the dielectric core width is above $\lambda_0/2$ n (i.e., w > 1), but as the core width decreases below $\lambda_0/2$ n (corresponding to w = 1), the confinement factor for the air-clad waveguide decreases dramatically while that for the metal-clad waveguide increases dramatically. Therefore, one can loosely categorize modes in w > 1 waveguide to be photonic modes and modes in w < 1 waveguide to be plasmonic modes, complementary to the categorization based on the propagation constant in Figure 1.3.

Because the radiative efficiency in semiconductor lasers is proportional to the product of the electronic density of states (EDOS) and the photonic density of states (PDOS), tailoring both densities of states (DOS) has played an important role in increasing laser efficiency and, consequently, reducing the device size. With electronics being the more

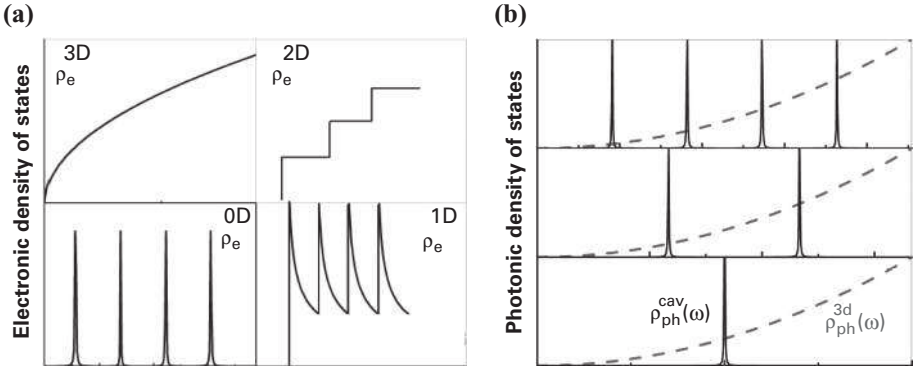

Figure 1.5 (a) Electronic density of states (EDOS) for bulk (3D), quantum well (2D), quantum wire (1D), and quantum dot (0D) semiconductors. (b) Photonic density of states (PDOS) for a 3D cavity as the cavity size reduces (from top to bottom). The PDOS of vacuum is denoted by the parabolic dashed line in each panel. Reprinted from reference [47] with permission from Springer Publishing.

mature field, engineering EDOS of materials has led the effort: the tremendous progress in material growth and fabrication technologies have enabled dimensionality reduction from bulk semiconductor (3D) to quantum wells (2D), quantum wires (1D), and eventually to quantum dots (0D), resulting in their perspective EDOS shown in Figure 1.5(a). Such EDOS engineering allows us to tailor not only the density but also the location of the electronic states in the energy space to ensure efficient use of the electronic states. From the photonics point of view, the PDOS can be engineered by changing the cavity size: generally speaking, PDOS decreases with decreasing cavity size because fewer modes are allowed within the spectral window of gain, as illustrated by Figure 1.5(b). If the electronic states in Figure 1.5(a) can be aligned with the photonic states in Figure 1.5(b) such that the peaks exactly coincide, the most energy-efficient laser device can be realized.

1.2 Active Materials for Nanolasers

As should become clear from Section 1.1, nanolasers have higher threshold gain compared to their larger-scale counterpart. This is because with typically lower Q factors, radiative loss becomes a larger part of the loss mechanism; with higher surface-to-volume ratio, surface recombination, which is a form of non-radiative recombination, also becomes a larger part of the loss mechanism. These higher losses must be compensated by material gain that is provided by the smaller-sized active material. Furthermore, self-heating − which lowers material gain − is more severe in smaller active volumes. As a result, only active medium with high material gain, namely liquid and inorganic solid-state material, can be used in nanolasers. Figure 1.6 summarizes optical gain materials that have been considered for use in nanolasers.

Inorganic semiconductors (including bulk, quantum-well, and quantum dot) are by far the most widely used gain material in small lasers, and they offer the most practical means

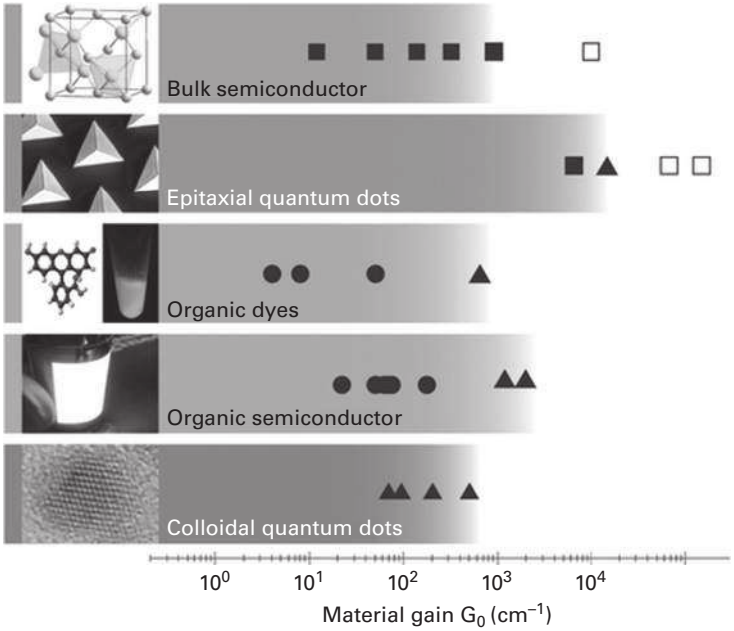

Figure 1.6 Typical material gain available from different types of active media. Square denotes material gain under electrical pumping (open square denotes cryogenic temperature material gain), circle denotes material gain measured from amplified spontaneous emission (ASE) data with nanosecond optical pumping, triangle denotes material gain measured from ASE data with sub-picosecond optical pumping. Reprinted from reference [43] with permission from Macmillan Publishers Ltd.

of achieving electrically controlled and mass-producible devices. A great advantage of inorganic semiconductors over other types of gain media is the ease with which the material may be controllably doped for the formation of heterostructures suitable for electronic injection devices. The material gain provided by semiconductors ranges from a few hundred per cm from bulk gain, to a few thousand per cm from multiple-quantum-well (MQW) gain, and to tens of thousands per cm from quantum dot (QD) gain. For MQW and QD gain, one does need to take into account the fact that the modal gain is often much smaller than the material gain, which results from the limited mode-gain overlap in these types of gain media. Nonetheless, carefully designed materials such as high-density QD or strained MQW can be utilized to increase gain further and to balance out the small mode-gain overlap. The disadvantage of using inorganic semiconductor is that device fabrication usually uses top-down fabrication procedures and requires lithography and deposition steps that can be quite complex for nanoscale devices. Additionally, integration with other material platforms, such as those based on silicon, is not trivial.

On the other hand, organic gain materials (including organic dye and semiconductor) as well as colloidal QD nanocrystals can be fabricated using the bottom-up fabrication approach that omits expensive lithography and etching steps. These gain materials can be prepared in the form of thin films with sub-micrometer precision by solution-based processes. In terms of the gain spectrum, instead of needing a priori knowledge of the desired

spectrum and choosing semiconductor compositions accordingly as in the case of inorganic semiconductors, the spectrum can be readily tuned by targeted modifications of chemical composition, dimension, and structure. Although nanolasers based on organic dye are not amenable for chip-scale integration with contemporary electronics when they are embedded in mechanically flexible polymer, they find their significance in microfluidic- and opto-fluidic-based applications [48]. A significant feature of fluidic-based lasers is that the refractive index and gain spectrum of the active medium can be tuned in real time by changing the solvent or dye that passes through the device. As for organic semiconductors, even though they have the potential for realizing electrically controlled lasers, their low charge-carrier mobility has so far prevented the realization of electrically pumped organic semiconductor lasers. Because organic light emitting diodes (OLED) have attracted significant interest in research and lighting applications, it is worth exploring nanoscale OLED and even toward organic semiconductor nanolasers. The main disadvantage of organic gain material is photo-induced bleaching of fluorescence and gain under high-excitation conditions, leading one to question their long-term stability and maximal output power.

In comparison with organic gain media, inorganic colloidal QDs retain many of the attractive properties of organic gain, such as solution-based processes and gain spectrum tenability, while offering better stability. Colloidal QDs typically have a radius between 1 and 4 nm, which is smaller than epitaxially grown QDs (whose typical radius is 3−10 nm). The material gain in thin films of colloidal QDs has been measured to be a few hundred per cm using sub-picosecond optical excitation, and lasing has been demonstrated over a wide range of wavelengths [49].

Although each type of gain material discussed earlier has its own attractiveness, in this book, we focus on nanolasers based on inorganic semiconductors, because it is the most technologically relevant gain material that is capable of current injection. Furthermore, it is compatible with the current fabrication processes for electronic integrated circuit (IC) infrastructure. Given that one of the driving forces behind semiconductor nanolaser research is the potential photonic integrated circuits (PICs) on a semiconductor chip for the fast growing information technology sector, much like the silicon-based electronic IC technology that has already infused into our everyday lives. Nano-PIC system is envisioned to supplement and expand and maybe eventually surpass the capability of electronic ICs. In current PICs, optical waveguides, splitters, couplers, and modulators can be made to subwavelength scale. However, current mainstream semiconductor lasers, while being the smallest compared to other types of lasers (e.g., dye, gas, and solid-state lasers), are still too large for PIC. Therefore, it is fitting to say that size reduction of semiconductor lasers will highly impact the development of PICs, future information technology, and other on-chip applications such as detection and sensing.

1.3 Fundamental Scale Limits of Lasers

Before diving into specific nanolaser designs and the various applications of nanolasers, we first consider the fundamental limit on the size of a laser. A typical laser has two vital components. First, it must have a resonator (or cavity) that supports optical mode(s). An

inherent size limitation associated with the resonator is that the cavity length in the mode propagation direction cannot be shorter than half of the mode wavelength in the medium. This is known as the diffraction limit. Second, a laser must have gain (or active) medium that is population inverted and supplies energy to the lasing mode(s). There needs to be enough gain medium such that population inversion can be achieved, which sets another restriction to the laser size. One exception is the exciton-polariton laser, which operates without a population inversion in the strong coupling regime. Exciton-polariton laser was first introduced in 1996 [50] and was demonstrated to operate under electrical injection in 2013 [12, 51]. While they provide an intriguing foundation to explore the quantum coherence of matter [52, 53], the sizes of these lasers are significantly larger than their emission wavelengths, typically in the 10 μm scale. We include a brief chapter (Chapter 10) to discuss the physics and development of exciton-polariton lasers, but they are not the focus of this book.

With a conceptual understanding of the two principle elements of a laser, we use a Fabry-Perot cavity as an example to derive the size limits of a laser. Figure 1.7(a) shows a typical Fabry-Perot cavity. It is an optical resonator that consists of two mirrors with reflectivity R_1 and R_2, respectively, and the space between the mirrors is filled with gain medium. The longitudinal direction is the propagation direction along cavity length L, and the transverse direction is the waveguiding plane, both of which are schematically shown in Figure 1.7(a).

For a cavity mode to get amplified, standing waves need to form in the longitudinal direction. Taking this direction to be the x-direction, physically, it means that a complex field amplitude, $E_0(x)$, at an arbitrary location x inside the cavity will maintain its original value after a round trip propagation. We denote the complex permittivity of the gain medium as $\varepsilon_g(\omega) = \varepsilon_g'(\omega) + i\varepsilon_g''(\omega)$ and the wave vector as $k(\omega) = k_0(\omega)\sqrt{\varepsilon(\omega)}$, where $k_0(\omega) = \dfrac{\omega}{c}$ is the vacuum wave vector. The standing wave requirement is mathematically expressed as

$$\sqrt{R_1 R_2} e^{i2k_z(\omega)L} E_0(x) = E_0(x) \tag{1.1}$$

Equation (1.1) can be understood by inspecting the effect of the real and imaginary parts of $k_z(\omega) = k_z'(\omega) + ik_z''(\omega)$ separately. The minimum length, $L_{\min}^{long,1}$, asserted by the real part of $k_z(\omega)$, is

$$\begin{gathered} 2k_z'(\omega)L = 2m\pi, m = 1, 2, 3, \cdots \\ L_{\min}^{long,1} = \frac{m\pi}{k_z'(\omega)} = \frac{m\pi \cdot k_0(\omega)}{\mathrm{Re}(n_{eff}(\omega))} \end{gathered} \tag{1.2}$$

with effective index defined as $n_{eff}(\omega) \equiv k_z(\omega)/k_0(\omega)$. In Equation (1.2), m is an integer, therefore the shortest possible cavity length is $\pi/k_z'(\omega) = \lambda_{eff}/2$ where λ_{eff} is the wavelength of light inside the cavity. Equation (1.2) describes the half-wavelength condition, also known as the diffraction limit. This applies to all electromagnetic waves and sets a lower limit to the cavity size. Similarly, the minimum length in the transverse direction is also set by the diffraction limit and is denoted by $L_{\min}^{trans}$.