DENSELY INTEGRATED PHOTONICS STRUCTURES FOR ON-CHIP SIGNAL PROCESSING

A Thesis
Presented to
The Academic Faculty

by

Qing Li

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
August 2013

DENSELY INTEGRATED PHOTONICS STRUCTURES FOR ON-CHIP SIGNAL PROCESSING

Approved by:

Professor Ali Adibi, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Ghassan Al-Regib
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Gee-Kung Chang
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Joseph Perry
School of Chemistry and Biochemistry
Georgia Institute of Technology

Professor John A. Buck
School of Electrical and Computer Engineering
Georgia Institute of Technology

Date Approved: April 2013

To my parents,

for their endless support.

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Professor Ali Adibi, for his strong support and guidance during my PhD. As an advisor, Ali Adibi is not only intelligent and knowledgeable, but also has a nice personality towards his students, showing them true care and respect. I believe anyone that works with him long enough can tell that.

I would also like to thank Professors Gee-kung Chang, John A. Buck, Ghassan Al-Regib and Joseph Perry for reviewing my thesis and serving on my PhD defense committee. Professor Chang is always giving me a lot of advice and encouragement on my work. Professor Buck has also been very helpful during the course of my PhD, and I would like to thank him for his advice and support.

I am indebted to Mohammad Soltani for his help over the years. When I first joined the Photonic Research Group, I had the chance to work closely with him. During my three-year overlap with Mohammad, he taught me many things from theory to experiment. In addition, his diligence and devotion to his work has always been an example that I am trying to follow.

I would like to thank Siva Yegnanarayanan, our former post-doc member and a true gentleman. His broad knowledge and strong experimental skills were valuable to me. Many times when I ran into a characterization problem, Siva was first person I thought of and he would never hesitate to offer the help.

I would also like to thank other former group members, including Babak Momeni, Majid Badiei, Murtaza Askari, Chaoray Hsieh, Fengtao Wang, Saeed Mohammadi, Charles H. Camp, Ehsan S. Hosseini and Maysam Chamanzar. Ehsan has helped me develop my nanofabrication skills, and his cheerful personality makes him good

company. Maysam joined the group almost the same time as I did, and he has been very kind to offer any help he could.

My thanks also go to current group members, including Ali A. Eftekhar, Amir H. Atabaki, Payam Alipour, Farshid Ghasemmi, Reza Pourabolghasem, Majid Sodagar, Hossein Taheri, Amir H. Hosseinnia, Ye Luo, Zhixuan Xia, and many others. A lot of joyful moments have been spent together and I will remember that.

And finally, I would like to thank my friends and family. Especially, my parents have always been behind me in this almost seven-year journey. Without their support, patience and sacrifice, it is hard to imagine getting this done.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Microelectronics has enjoyed great success in the past century. As the technology node progresses, the complementary metal-oxide-semiconductor scaling has already reached a wall, and serious challenges in high-bandwidth interconnects and fast-speed signal processing arise. The incorporation of photonics to microelectronics provides potential solutions. The theme of this thesis is focused on the novel applications of travelling-wave microresonators such as microdisks and microrings for the on-chip optical interconnects and signal processing. Challenges arising from these applications including theoretical and experimental ones are addressed. On the theoretical aspect, a modified version of coupled mode theory is offered for the TM-polarization in high index contrast material systems. Through numerical comparisons, it is shown that our modified coupled mode theory is more accurate than all the existing ones. The coupling-induced phase responses are also studied, which is of critical importance to coupled-resonator structures. Different coupling structures are studied by a customized numerical code, revealing that the phase response of symmetric couplers with the symmetry about the wave propagating direction can be simply estimated while the one of asymmetric couplers is more complicated. Mode splitting and scattering loss, which are two important features commonly observed in the spectrum of high-Q microresonators, are also investigated. Our review of the existing analytical approaches shows that they have only achieved partial success. Especially, different models have been proposed for several distinct regimes and cannot be reconciled. In this thesis, a unified approach is developed for the general case to achieve a complete understanding of these two effects. On the experimental aspect, we first develop a new fabrication recipe with a focus on the accurate dimensional control and

low-loss performance. HSQ is employed as the electron-beam resist, and the lithography and plasma etching steps are both optimized to achieve vertical and smooth sidewalls. A third-order temperature-insensitive coupled-resonator filter is designed and demonstrated in the silicon-on-insulator (SOI) platform, which serves as a critical building block element in terabit/s on-chip networks. Two design challenges, i.e., a broadband flat-band response and a temperature-insensitive design, are coherently addressed by employing the redundant bandwidth of the filter channel caused by the dispersion as thermal guard band. As a result, the filter can accommodate 21 WDM channels with a data rate up to 100 gigabit/s per wavelength channel, while providing a sufficient thermal guard band to tolerate more than $\pm 15^{o}$C temperature fluctuations in the on-chip environment. In this thesis, high-Q microdisk resonators are also proposed to be used as low-loss delay lines for narrowband filters. Pulley coupling scheme is used to selectively couple to one of the radial modes of the microdisk and also to achieve a strong coupling. A first-order tunable narrowband filter based on the microdisk-based delay line is experimentally demonstrated in an SOI platform, which shows a tunable bandwidth from 4.1 GHz to 0.47 GHz with an overall size of 0.05 mm^2. Finally, to address the challenges for the resonator-based delay lines encountered in the SOI platform, we propose to vertically integrate silicon nitride to the SOI platform, which can potentially have significantly lower propagation loss and higher power handling capability. High-Q silicon nitride microresonators are demonstrated; especially, microresonators with a 16 million intrinsic Q and a moderate size of 240 μm radius are realized, which is one order of magnitude improvement compared to what can be achieved in the SOI platform using the same fabrication technology. We have also successfully grown silicon nitride on top of SOI and a good coupling has been achieved between the silicon nitride and the silicon layers.

CHAPTER I

INTRODUCTION

Microelectronics has enjoyed great success in the past century. Vast developments in this area are perhaps best described by Moore's law [1], which states that the number of transistors on an integrated circuit (IC) doubles approximately every two years. While this trend has been maintained for the past 50 years, it cannot go on indefinitely. Serious problems, such as power consumption and heat dissipation, have already arisen as the size scaling continues. In Intel's 22 nm microprocessor (Ivy-Bridge), a three-dimensional (3D) Tri-Gate technology is developed to keep Moore's law [2]. On the other hand, the equivalent physical gate oxide thickness has been predicted not to scale below 0.5 nm, indicating that the complementary metal-oxide-semiconductor (CMOS) scaling in this area has reached a wall [2]. While ingenious solutions to this problem might be possible, such efforts become more and more difficult when the fundamental limits are approached.

The integration of photonics and microelectronics provides new perspectives. For example, a major trend in computer architecture is to build multicore system-on-chip processors, which integrate many independent processing units on the same die [3]. With the increase in integration levels, communication among these units becomes bandwidth demanding. For instance, in PlayStation 3, a video game console produced by Sony Computer Entertainment, one power processing element and eight synergistic processing elements are implemented, and an interconnect bus with a peak bandwidth more than 200 gigabyte/s is implemented to facilitate the data exchange among the processors [3]. Because of the limited bandwidth capacity of the electrical wires, a high bandwidth requires a significant amount of physical links. Therefore, the

electrical interconnects can easily consume more than 50% of the total power budget, making the system design difficult [4]. In contrast, optics is capable of handling high bandwidth because of the high carrier frequency. Wavelength division multiplexing (WDM) is another option that is available for enhancing the bandwidth performance. To replace the electrical wires, a complete optical link including the optical signal generation, modulation, and detection is required. The power consumption of such optical links has been shown to be significantly lower than that of the electrical interconnects [5, 6], which makes the optical interconnects very attractive to system designers.

The incorporation of photonics to microelectronics is not only beneficial for the interconnecting problem of the digital electronic circuits. Radio-frequency (RF) photonics is another area that manifests the usefulness of such technological combination. By converting the RF signal to the optical domain, RF photonics takes advantage of low-loss waveguiding, easy reconfiguration, and high-speed responses of optical components. Traditionally, optical fiber is used because of its extreme low loss, light weight, and affordable cost [7]. One major shortcoming of this approach is the low integration level limited by the size of the fiber. Completely on-chip integration via the photonic integrated circuit (PIC) would not only make the devices compact and low-cost, but also enable applications which require a high level of integration to accomplish complex functionalities. For example, high-bandwidth analog-to-digital converter (ADC) implemented with the microelectronic circuits has difficulty in achieving a large dynamic range in the digital domain [8]. The availability of narrow-bandwidth filtering (around MHz) and fast reconfiguration (around ns) in photonics permits the realization of wideband-tunable narrowband optical filters, which can be rapidly programmed for real-time applications. By imposing the RF signal to an optical carrier and performing the signal processing in the optical domain, similar functionalities such as ADC can be attained in a cost-effective manner.

Currently silicon-on-insulator (SOI) technology seems to be the most promising candidate for the on-chip photonic and microelectronic integration [9, 10, 11]. Silicon is the host material for the majority of microelectronic devices. It also offers an excellent platform for most of photonic applications considering the following attractive properties [9, 10, 11]: (1) silicon has a large refractive index contrast to silicon dioxide, which enables tight confinement of optical waves; (2) intrinsic silicon is transparent at the telecommunication wavelength thus promising low-loss waveguides; (3) several mechanisms are available for refractive index tuning, including the thermo-optic effect and the free-carrier plasma dispersion effect; (4) the strong third-order nonlinearity of silicon can be exploited for novel applications such as wavelength conversion and on-chip amplification. Of course, no material is perfect and silicon is no exception. The biggest concern about silicon is that it has an indirect bandgap, and on-chip light sources are difficult to realize [9]. Nevertheless, silicon photonics have attracted intense research interest over the last decade and tremendous progresses have been made. Building blocks such as hybrid silicon lasers [12], high-speed modulators and switches [13, 14], hybrid detectors [15], WDM band-pass filters [16], and optical buffers [17] are reported. On the system integration side, Intel has successfully demonstrated a 50 gigabit/s photonic link with all the components on the chip scale [18], signaling that volume manufacturing of silicon photonics are within the near future.

Similar to Moore's law for microelectronics, dense integration is envisioned to be the major trend for photonic devices. Microresonators, capable of localizing and storing the optical waves both spatially and temporally at a specific wavelength range, are considered as the key building blocks for highly integrated photonic chips [19, 20]. Various types of microresonators with different geometrical and resonant properties have been proposed in the past [20]. Among all the possible microresonator architectures, microdisk and microring resonators are the ones that have been heavily

investigated [20]. Thanks to their simple structure and high flexibility in both the design and the implementation, the microdisk and microring resonators are suitable for the dense integration purpose. Various functionalities ranging from the compact modulators [14] to on-chip logics [21] have been demonstrated.

The research work presented in this thesis is focused on the novel applications of the microdisk and microring resonators for the on-chip optical interconnects and signal processing. For the interconnects, a wideband coupled-resonator filter is designed and demonstrated with a terabit/s bandwidth capacity. In addition, high-Q microdisk-based delay lines are explored for narrowband signal processing in both the silicon-on-insulator (SOI) and silicon nitride on SOI platforms. During the course of the thesis development, various challenges from the design, fabrication and characterization are overcome. Therefore, the presented work has both the theoretical and experimental flavors. For example, the initial responses from the coupled-resonator filters always show a certain level of distortions, for which two sources are identified: one is from the coupling-induced resonance frequency shift and the other from the imperfect fabrication process. For the former one, a theoretical study is carried out with the aid of numerical simulations; for the latter one, an improved fabrication process is developed to accurately control the device sizes. Another example could be found for the application of high-Q microresonators for narrowband filters. Fabrication processes are optimized to reduce the sidewall roughness of microresonators, so their propagation losses can be minimized. Moreover, theoretical studies are performed for a better understanding of the commonly observed phenomena such as mode splitting and scattering loss.

1.1 Thesis organization

This thesis consists of five chapters. Except for the first Introduction chapter, the rest of the thesis are organized as follows:

- Chapter II: We first review the basic concepts of high-Q microresonators, such as quality factors, mode volumes, mode splitting and scattering loss. Following that, a new fabrication process that is specifically developed for the thesis work is described. A brief discussion on the characterization of high-Q microresonators is also provided. These elements are all indispensible for the following works. We put them together so we do not have to revisit them in each individual work.

- Chapter III: In this chapter, theoretical efforts for the study of photonic devices are presented. Unlike electronic circuit design where sufficient knowledge is already gained and powerful designing software is available, for photonic devices it is often necessary to resort to analytical studies and customized numerical simulations. Three important problems are studied in this chapter. The first problem is on the power coupling between photonic structures, which is a common problem in the design of almost every device. We will show that for high-index-contrast material systems (e.g., SOI), the existing coupled-mode theories are not satisfactory and a modified version is provided. In the second problem, we study the phase shift caused by the coupling process, which is important for coupled-resonator devices since the phase shift will change the resonance frequency of the coupled resonator. In the third problem, mode splitting and scattering loss are studied, for which a unified model is developed.

- Chapter IV: In this chapter, experimental efforts based on the SOI platform are presented. Two devices are discussed. The first one is a third-order temperature-insensitive coupled-resonator filter, which serves as a critical building block element in terabit/s on-chip networks. The second one is on the use of high-Q microdisks for compact and low-loss delay lines, which are especially useful for narrowband filters for the RF signal processing. A tunable first-order filter is

demonstrated using such a delay line structure. In addition, several problems are identified for delay lines based on the silicon high-Q microresonators, including the relatively large propagation loss and the strong nonlinearity from silicon which limits the maximum power that can be handled by the device.

- Chapter V: To solve the problems encountered in delay lines based on silicon microresonators , in this chapter we will develop a new material platform based on the vertical integration of silicon nitride into the SOI platform. High-Q silicon nitride microresonators with compact sizes are demonstrated. Moreover, the silicon nitride has been successfully integrated with the SOI and an efficient coupling between these two layers is achieved.

- Chapter VI: The thesis work is summarized and a brief discussion on the future extensions is provided.

CHAPTER II

HIGH-Q MICRORESONATORS

Microresonators are considered as one of the key building blocks for integrated photonics because of their capability of localizing and storing optical waves both spatially and temporally at a specific wavelength range [19, 20]. Their high quality factors (Qs) and microscale mode volumes enable strong light-matter interactions, which are the essential reasons behind the wide applications of microresonators including low-threshold lasing [22], low-power optical modulation [14], single-nanoparticle sensing [23, 24], ultrasensitive micromechanical displacement detection [25], as well as fundamental studies on cavity quantum electrodynamics [26].

Over the past few decades, various microresonator architectures, including Fabry-Perot resonators, photonic crystal cavities, and whispering gallery mode (WGM) resonators have been investigated [19]. Among these different types of resonators, microrings and microdisks have received considerable interest, thanks to their simple structures and high flexibilities in both the design and implementation for integrated photonic circuits. An array of functionalities ranging from compact modulators [14] to on-chip logics [21] has been demonstrated by different groups.

In this chapter, we will first introduce the fundamental properties of microresonators which are necessary for device applications as presented in the following chapters. We will also cover the fabrication of planar microresonators such as microrings and microdisks, using the silicon-on-insulator material platform as an example. The last section is devoted to the proper characterization of microresonators, and discussions on the insertion loss, polarization control and laser-related issues are presented.

2.1 Fundamental properties of microresonators

There are a few parameters which we frequently use to characterize the properties of a microresonator. Among them are quality factor (Q), mode volume and free spectral range (FSR), whose definitions and common usage will be reviewed below. In addition, a detailed review of mode splitting and scattering loss, which are two important features of microresonators, is provided.

2.1.1 Quality factor

The quality factor is a parameter defined as

$$Q \equiv \frac{\omega \cdot U_c}{P_{\text{loss}}}, \tag{1}$$

where ω is the frequency of oscillation, U_c is the mode energy, and P_{loss} is the power dissipation rate. Using the relation $\omega = 2\pi/T$, where T is the temporal period of the oscillation, one can easily verify that Q equals 2π times the ratio of the energy stored in the resonator to the energy loss per optical cycle. If the resonator has no external excitation, Eq. 1 can be written as

$$P_{\text{loss}} = -\frac{dU_c(t)}{dt} = \frac{\omega}{Q}U_c(t), \tag{2}$$

whose solution for $U_c(t)$ is

$$U_c(t) = U_0 e^{-\frac{\omega}{Q}t}. \tag{3}$$

From Eq. 3, the so-called photon lifetime τ is defined as $\tau \equiv Q/\omega$ [1], which is often interchangeably used with Q [27] . Another parameter associated with Q or τ is the cavity decay rate γ defined as $\gamma \equiv 1/\tau = \omega/Q$.

Equation 3 indicates how the Q of a resonator can be measured in the time domain. However, in practice it is often more convenient to measure the spectrum of

[1] in the literature there is another definition of the photon lifetime as $\tau \equiv 2Q/\omega$. The difference between these two definitions is that in our case the photon lifetime is defined for the cavity energy while the other one is for the cavity amplitude.

a resonator rather than the time-domain response. By performing a Fourier transform to Eq. 3, we get a Lorentzian-shaped resonance from which an alternative expression for Q can be derived as

$$Q = \frac{\omega}{\delta\omega}, \tag{4}$$

where $\delta\omega$ denotes the full width at half maximum (FWHM) of the resonance.

According to Eq. 1, we can ascribe different Qs for different loss mechanisms. For example, in silicon microresonators, loss can arise from scattering which is caused by the surface roughness at the sidewalls. Absorption loss also occurs if there are unsaturated surface states. Moreover, when the injected power is high, free carriers are generated in silicon through the two photon absorption process, which contribute an additional loss. Generally, all these loss terms are undesired. There is one exception, however, that is, the coupling loss, which refers to the energy intentionally coupled out of the resonator for a proper device functioning. We use intrinsic Q to represent all the losses except the coupling loss, coupling Q for the coupling loss, and loaded Q for the total loss. Since the loss terms are additive, from Eq. 1, we have

$$\frac{1}{Q_{\text{loaded}}} = \frac{1}{Q_i} + \frac{1}{Q_c}, \tag{5}$$

where Q_i and Q_c represent the intrinsic and coupling Qs, respectively.

2.1.2 Mode volume

The mode volume V_m is a parameter that is used to evaluate the confinement of a resonant mode. There are certain flexibilities in defining the mode volume and the most common expression is given by [28]

$$V_m \equiv \frac{\int \varepsilon(\boldsymbol{r})|\boldsymbol{E}(\boldsymbol{r})|^2 d\boldsymbol{V}}{\max\left[\varepsilon(\boldsymbol{r})|\boldsymbol{E}(\boldsymbol{r})|^2\right]}, \tag{6}$$

where $\varepsilon(\boldsymbol{r})$ is the permittivity and $\boldsymbol{E}(\boldsymbol{r})$ is the electric field of the resonant mode. Other forms of mode volume are possible by replacing the expression in the denominator in Eq. 6 to assume different physical meanings. Most of the time, mode volume

is a concept that is useful when comparison is involved.

2.1.3 Free spectral range

Because of the wave nature of electromagnetic fields, there are a series of resonant modes for microresonators with periodic boundary conditions. For example, for microrings and microdisks, the wave propagating in the azimuthal direction has the $\exp(im\phi)$ factor (ϕ is the azimuth), where m should be an integer following the boundary condition. For each m, there is a corresponding resonance, and together they form a resonant mode family, for which the free spectral range (FSR) is defined as the frequency or wavelength spacing between adjacent resonances.

FSR is a concept that is not limited to microresonators with cylindrical symmetry. For example, for Fabry-Perot resonators, the FSR is given by [27]

$$\mathrm{FSR} = \frac{c}{n_g L}, \tag{7}$$

where c is the speed of light in vacuum, n_g is the group index of the forming waveguide, and L is the round-trip length. However, for resonators such as microrings and microdisks, only FSR is a well-defined parameter and neither n_g nor L is well defined. In some cases, we can take $L \equiv 2\pi R$ where R is the outer radius of the microresonator (other definitions for L are possible), and the group index n_g for these resonant modes can be defined as

$$n_g \equiv \frac{c}{\mathrm{FSR} \cdot 2\pi R}. \tag{8}$$

It is worth emphasizing here that the group index n_g for microresonator modes is only meaningful when comparison is involved and should be considered as an alternative parameter for the FSR.

2.1.4 Mode splitting and scattering loss

An ideal WGM resonator usually has degenerate resonant modes because of its structural symmetry. For example, clockwise (CW) and counterclockwise (CCW) WGMs

are both supported in microtoroid and microdisk resonators with identical mode properties. In particular, because they share the same resonance frequency and linewidth, only one single resonance will be observed in the transmission measurement.

The degeneracy in the resonant modes can be lifted by destroying the structural symmetry, either done purposely such as by introducing nanoparticles to the surface of microresonators [29, 30], or caused by the ubiquitous roughness from nanofabrication [28]. As a result, the two degenerate modes (i.e., the CW and CCW modes) will couple to each other and a doublet appears in the transmission spectrum, a phenomenon known as mode splitting [31]. The structural defects can also couple the confined WGMs to free-space radiation modes, generating scattering loss and thus linewidth broadening to these resonances [29].

Mode splitting and scattering loss have been investigated in many different works [28, 29, 30, 32, 33, 34]. Based on the number of scatterers and the applications, previous works can be categorized into two distinct scenarios. In the first scenario, small nanoparticles are introduced to the surface of high-Q microresonators, with focused applications such as strong light-matter interactions and nanoparticle sensing [23, 24, 26]. The number of nanoparticles is usually limited to a few, and reasonably good agreements between experimental observations and developed models have been achieved. The other scenario considers sub-wavelength scatterers that are intrinsic to microresonators, such as surface roughness caused by fabrication imperfections. In such cases, the number of scatterers is typically on the order of hundreds or even thousands, and a different approach has to be taken to study the mode splitting and scattering loss [28].

The first approach considers a coupled system consisting of the CW and CCW WGMs as wells as free space radiation modes, with their interactions assisted by each individual scatterer. The single-scatterer case has been well studied in high-Q

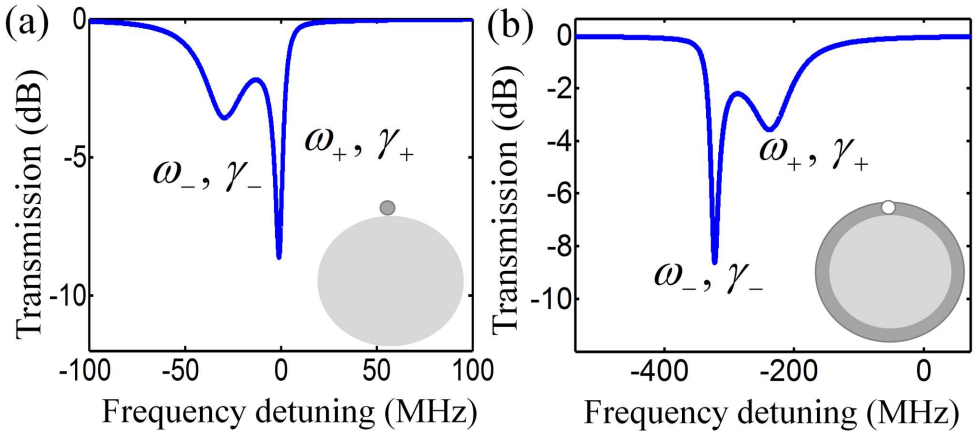

Figure 1: (a) Transmission response of a single dielectric scatterer on the surface of a microresonator. (b) Transmission response of numerous dielectric scatterers which uniformly cover the surface of a microresonator except for a vacancy.

microtoroid resonators [29, 32]. Two standing-wave modes, being symmetric and anti-symmetric combinations of the CW and CCW travelling modes, appear in the resonance spectrum. The symmetric mode has a nonzero field overlap with the scatterer, resulting in a red shift in its resonance frequency and a broadening in its linewidth (we assume positive dielectric perturbations from scatterers unless otherwise specified). On the other hand, the anti-symmetric mode has a zero field overlap with the scatterer; therefore, its resonance frequency and linewidth stay the same as those of the CW (CCW) mode without the scatterer. These two resonances are illustrated in Fig. 1(a), with ω_+ (ω_-) and γ_+ (γ_-) denoting the resonance frequency and loss rate of the eigenmode that has a higher (lower) resonance frequency of the two split modes (i.e., $\omega_+ \geq \omega_-$). Based on the above reasoning, we conclude that $\gamma_+ < \gamma_-$ for the single-scatterer case. The two-scatterer scenario has also been explored in Ref. [34], which has experimentally demonstrated that the two eigenmodes can either have no mode spitting ($\omega_+ = \omega_-$) or a symmetric ($\gamma_+ = \gamma_-$) or an asymmetric ($\gamma_+ < \gamma_-$) doublet, depending on the relative position of the two scatterers. These experimental results can be analyzed by a generalized model presented in Ref. [34], which considers multiple scatterers that are well separated apart so that the contribution from

each scatterer can be considered to be independent with each other. However, when the scatterers are closely spaced, this independent-scatterer approach fails to predict correct results. For example, for N identical scatterers, the model in Ref. [34] gives

$$\omega_{\pm} = \omega_c - NG_0 \pm G_0 \left| \sum_{n=1}^{N} e^{i2kx_n} \right|, \tag{9}$$

$$\gamma_{\pm} = N\Gamma_0 \mp \Gamma_0 \left| \sum_{n=1}^{N} e^{i2kx_n} \right|, \tag{10}$$

where ω_c is the originally degenerate resonance frequency of the WGMs before the introduction of scatterers; G_0 and Γ_0 are parameters (both positive) characterizing the resonance frequency shift and linewidth broadening caused by an individual scatterer, respectively; k is the wavenumber of the WGM; and x_n is the projection of the nth scatterer's position on the WGM's wave travelling direction. From Eq. 10, one can infer that for an arbitrary number of identical scatterers, $\gamma_+ \leq \gamma_-$. This conclusion seems valid, by arguing that the resonance corresponding to ω_- has a lower resonance frequency because of the more field overlap with the scatterers, which is also responsible for a stronger scattering loss. However, such intuitive belief is not generally true. Here we consider one extreme example in Fig. 1(b), where identical scatterers have uniformly covered the outer surface of a microresonator except for one vacancy. Assuming the scatterers and the microresonator share the same dielectric constant, the resulting structure can be regarded as a larger-radius microresonator with a negative-dielectric-constant scatterer at the vacancy point. Consequently, the mode that has a nonzero field overlap with the scatterer will incur a blue shift in the resonance frequency (relative to the resonance frequency of the larger-radius microresonator) as well as a linewidth broadening from the scattering. Hence, $\gamma_+ > \gamma_-$, contrary to the result of Eq. 10.

The second approach, developed mainly for the surface roughness present in high-Q microresonators, employs an intuitive physical model (which is essentially a phenomenological model) to describe the mode splitting and uses the volume current

method to obtain the scattering loss [28]. Here, we use the microdisk resonator as an example to give an introduction to this approach. For an isolated microdisk resonator (i.e., no external coupling), we have [27]

$$\frac{da_{\mathrm{ccw}}}{dt} = \left(-i\omega_c + i\Delta\omega_{\mathrm{ccw}} + \frac{\gamma_{\mathrm{ccw}}}{2}\right) a_{\mathrm{ccw}} + i\beta_{\mathrm{ccw}} a_{\mathrm{cw}}, \tag{11}$$

$$\frac{da_{\mathrm{cw}}}{dt} = -\left(i\omega_c + i\Delta\omega_{\mathrm{cw}} + \frac{\gamma_{\mathrm{cw}}}{2}\right) a_{\mathrm{cw}} + i\beta_{\mathrm{cw}} a_{\mathrm{ccw}}, \tag{12}$$

where a_{ccw} and a_{cw} represent the normalized energy amplitudes of the CCW and CW modes, respectively; ω_c assumes the same meaning as in Eq. 9, which denotes the unperturbed resonance frequency of the WGMs; $\Delta\omega_{\mathrm{ccw}}$ and $\Delta\omega_{\mathrm{cw}}$ are the resonance frequency shifts caused by the surface roughness to the CCW and CW modes, respectively; γ_{ccw} and γ_{cw} describe the corresponding scattering loss rates; β_{ccw} is a parameter characterizing the coupling from the CCW to the CW modes and β_{cw} is defined vice versa. From the fact that the CCW and CW modes only differ in their circulating directions, it is expected that

$$\Delta\omega_{\mathrm{ccw}} = \Delta\omega_{\mathrm{cw}},\ \gamma_{\mathrm{ccw}} = \gamma_{\mathrm{cw}},\ \text{and}\ \ \beta_{\mathrm{ccw}} = \beta_{\mathrm{cw}}^{*}. \tag{13}$$

In fact, from Maxwell's equations, $\Delta\omega_{\mathrm{ccw}}$ and β_{ccw} can be derived as [28]

$$\Delta\omega_{\mathrm{ccw}} = -\frac{\omega_c \int \Delta\varepsilon(\boldsymbol{r}) \left|\boldsymbol{E}_{\mathrm{ccw}}(\boldsymbol{r})\right|^2 d^3\boldsymbol{r}}{2\int \varepsilon(\boldsymbol{r}) \left|\boldsymbol{E}_{\mathrm{ccw}}(\boldsymbol{r})\right|^2 d^3\boldsymbol{r}}, \tag{14}$$

$$\beta_{\mathrm{ccw}} = \frac{\omega_c \int \Delta\varepsilon(\boldsymbol{r}) \boldsymbol{E}_{\mathrm{ccw}}^{*}(\boldsymbol{r}) \cdot \boldsymbol{E}_{\mathrm{cw}}(\boldsymbol{r})\, d^3\boldsymbol{r}}{2\int \varepsilon(\boldsymbol{r}) \left|\boldsymbol{E}_{\mathrm{ccw}}(\boldsymbol{r})\right|^2 d^3\boldsymbol{r}}, \tag{15}$$

where $\varepsilon(\boldsymbol{r})$ and $\Delta\varepsilon(\boldsymbol{r})$ correspond to the dielectric constant of a perfect microresonator and that of the surface roughness, respectively, and $\boldsymbol{E}_{\mathrm{ccw}}(\boldsymbol{r})$ ($\boldsymbol{E}_{\mathrm{cw}}(\boldsymbol{r})$) is the electric field of the CCW (CW) mode. Similar expressions for $\Delta\omega_{\mathrm{cw}}$ and β_{cw} exist with the exchange of the CCW and CW subscripts in Eqs. 14 and 15, and the relations in Eq. 13 become evident considering $\boldsymbol{E}_{\mathrm{ccw}}(\boldsymbol{r})$ and $\boldsymbol{E}_{\mathrm{cw}}(\boldsymbol{r})$ are conjugates of each other (assuming $\varepsilon(\boldsymbol{r})$ is real).

The scattering loss parameter γ_{ccw} (γ_{cw}) is obtained using the volume current method by computing the radiation power excited by the polarization current $\boldsymbol{J}_{\mathrm{ccw}}(\boldsymbol{r}) =$

$-i\omega_c \Delta\varepsilon(\boldsymbol{r})\boldsymbol{E}_{\mathrm{ccw}}(\boldsymbol{r})$. For example, the far-field electric field is given by [35]

$$\boldsymbol{E}^{\mathrm{far}}_{\mathrm{ccw}}(\boldsymbol{r}) = \frac{\omega_c^2 e^{ik_0 r}}{4\pi\varepsilon_0 c^2 r} \int \Delta\epsilon(\boldsymbol{r}')\boldsymbol{E}_{\mathrm{ccw}}(\boldsymbol{r}') \cdot (1 - \hat{\boldsymbol{r}}\hat{\boldsymbol{r}}) e^{-ik_0 \hat{\boldsymbol{r}}\cdot\boldsymbol{r}'}\, d^3\boldsymbol{r}', \tag{16}$$

where ε_0 and c are the permittivity and speed of light in free space, respectively; k_0 is the wavenumber; and $(r,\, \theta,\, \phi)$ are the spherical coordinates of the far-field position $\boldsymbol{r}$, with $(\hat{\boldsymbol{r}},\, \hat{\boldsymbol{\theta}},\, \hat{\boldsymbol{\phi}})$ denoting the orthogonal unit vectors in the directions of increasing $(r,\ \theta,\ \phi)$, respectively. In Eq. 16, we have adopted the $\exp(-i\omega_c t + ik_0 r)$ format for the outgoing light to be consistent with the convention used in Eqs. 11 and 12. The radiation power is calculated by integrating the Poynting vector over the sphere with radius r; and γ_{ccw}, according to its definition, is given by the power loss rate normalized by the mode energy as

$$\gamma_{\mathrm{ccw}} = \frac{\varepsilon_0 c \iint \left| r\boldsymbol{E}^{\mathrm{far}}_{\mathrm{ccw}}(\theta,\phi)\right|^2 \sin\theta\, d\theta d\phi}{2\int \varepsilon(\boldsymbol{r}) \left|\boldsymbol{E}_{\mathrm{ccw}}(\boldsymbol{r})\right|^2\, d^3\boldsymbol{r}}. \tag{17}$$

In addition, statistical information of the surface roughness can be inserted into Eq. 17, leading to an ensemble average for γ_{ccw} (γ_{cw})[28].

The eigenmodes of the coupled system are then obtained by solving the eigenvalue problem of Eqs. 11 and 12, which yields

$$a_{\pm} = \frac{1}{\sqrt{2}}\Big(a_{\mathrm{ccw}} \mp \frac{\beta_{\mathrm{ccw}}}{|\beta_{\mathrm{ccw}}|} a_{\mathrm{cw}}\Big), \tag{18}$$

where $a_{\pm}$ denote the two eigenmodes, whose resonance frequencies and scattering loss rates are given by

$$\omega_{\pm} = \omega_c + \Delta\omega_{\mathrm{ccw}} \pm |\beta_{\mathrm{ccw}}|\,, \tag{19}$$

$$\gamma_{\pm} = \gamma_{\mathrm{ccw}}. \tag{20}$$

Equation 20 predicts that $a_{\pm}$ should exhibit identical linewidths, while in experiments asymmetric lineshapes are often observed. For practical applications, $\gamma_{\pm}$ are generally assumed to be different to fit the model to the experimental data [28]. However, such an assumption contradicts Eq. 20.

To summarize, we conclude that the two major existing approaches for mode splitting and scattering loss have only achieved partial success. The first approach, which we call the independent-scatterer approach since it is based on collective contributions from each individual scatterer, works well when the number of scatterers is small and the scatterers are well separated from each other. On the other hand, the second approach, which we refer to as the intuitive physical approach since it is based on a phenomenological model for the mode splitting and the volume current method for the scattering loss, fits well for many-scatterer cases such as the surface roughness problem. However, both approaches have difficulties in providing correct scattering loss rates for the eigenmodes of the coupled system. In the independent-scatterer approach, lineshapes are predicted to be symmetric or asymmetric, but γ_+ is always no more than γ_-. In the intuitive physical approach, the model cannot predict asymmetric lineshapes (i.e., $\gamma_+ \neq \gamma_-$) in a self-consistent manner. Therefore, a unified approach which could provide a full understanding of the mode splitting and scattering loss needs to be developed. This will be presented in the next chapter.

2.2 Fabrication of planar microresonators

In this thesis we are primarily concerned with planar nanophotonic structures whose fabrication process consists of two major steps [36]. First, the pattern is defined using the lithography method. In the academic environment, electron-beam lithography (EBL) is typically employed in the patterning step because of its high accuracy and high flexibility. For nanophotonic devices, dimensional control with accuracy on the order of nanometers is often required to achieve acceptable device performance. In addition, because of the limited resolution from both the lithography system and the adopted resist, a certain level of roughness is always present in the pattern and should be minimized. In the second step, the pattern is transferred from the resist to the underneath device layer through plasma etching. An ideal etching process is desired

to add minimal roughness to the device as well as to ensure a vertical sidewall angle, which is a critical factor for dense integration.

2.2.1 Recipe using ZEP

Prior to this work, a fabrication process has already been developed within the Photonic Research Group using the JEOL JBX-9300FS EBL system available in the cleanroom for the silicon-on-insulator (SOI) material platform [36]. ZEP (ZEP520A by Zeon corporation) has been chosen as the e-beam resist, which is capable of defining fine features with a relatively good etch resistance. Individual elements such as waveguides and microdisk resonators are demonstrated with good performance. Especially, the propagation loss for microdisk resonators is among the best results ever reported ($\sim$ 0.2 dB/cm)[37]. We show some photonic structures fabricated with this recipe in Fig. 2. Figure 2(a) shows the scanning-electron micrograph (SEM) of a 20-μm-radius microdisk resonator side coupled to a waveguide, and Fig. 2(b) shows the SEM of the microdisk sidewall, whose roughness is estimated to be below 5 nm.

One problem with this recipe is the limited accuracy on controlling the device size, which is mainly caused by the tilted sidewalls of the fabricated structures. Figure 2(c) shows the cross-section of a 500-nm-wide waveguide. The sidewall angle is measured to be around 83 degrees. The fact that the sidewall is nonvertical is also evident from the top view of a straight waveguide, as shown in Fig. 2(d). At each side of the waveguide, two edge lines appear which correspond to the top and bottom surfaces of the waveguide. For a 250-nm-thick device layer, the difference between the top and bottom widths of the waveguide can be up to 50 nm. The effective waveguide width thus has to be adjusted considering this tilting effect, which is difficult since the sidewall angle can vary unpredictably by a few degrees from run to run. The error in the structure size (up to tens of nanometers) will lead to poor device performance and low yields for devices that are sensitive to size variations, and a more accurate

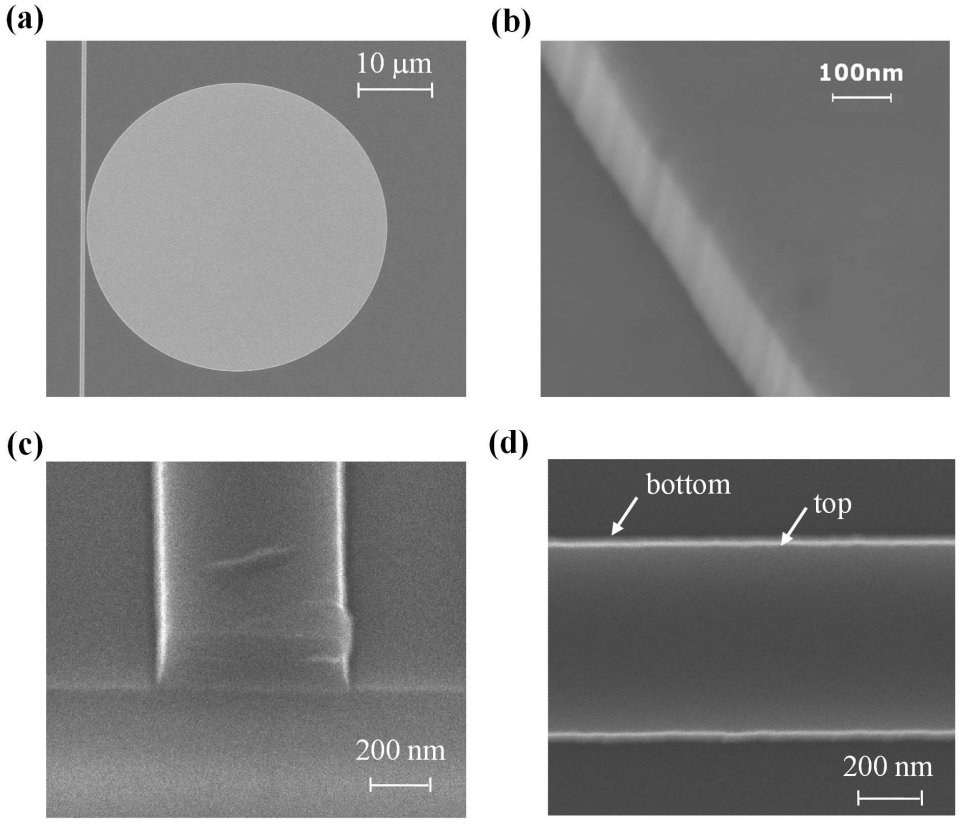

Figure 2: (a) SEM of a 20-μm-radius silicon microdisk resonator side coupled to a waveguide; (b) sidewall of the microdisk captured at a tilt angle of 30°; (c) cross-section of a 500-nm-wide waveguide at a tilt angle of 60°; (d) top view of the waveguide.

recipe has to be developed.

2.2.2 Recipe using HSQ

The starting point of developing a new recipe is to choose an appropriate e-beam resist. HSQ (XR-1541 by Dow Corning) has been reported in the literature as an accurate resist with a similar etch resistance as that of silicon dioxide. This promises a high selectivity to silicon in the subsequent plasma etching. Employing a resist with a high selectivity has several benefits. First, the sidewall of the etched structure can be more vertical even if the resist itself has a tilted sidewall. Second, only a thin layer of resist is required to mask the underneath device layer, allowing a better lithography resolution and finer feature sizes. We choose the 6% concentration of HSQ, which forms a 110-nm-thick resist layer after spin coating. TMAH (Tetramethylammonium hydroxide, by Moses Lake Industries, Inc.) is employed for the HSQ development, whose concentration is typically varied between 2% to 25% to allow different contrast (the higher the concentration, the better the contrast). The contrast of TMAH can be further enhanced by developing HSQ at high temperatures. However, as the temperature rises, TMAH starts to attack silicon as well. For example, at 80°C the 25% TMAH would etch silicon at a rate of 600 nm/min [39]. Certainly, such a situation has to be avoided because the wet etch of silicon will result in a poor sidewall angle. In our process, the developing condition has been optimized to be a 30-second dip in 25% TMAH which is heated by a hotplate set at 80°C. The actual temperature of the TMAH solution is around 43°C. After the TMAH dip, the sample is washed under running deionized water for 5 minutes to get rid of possible HSQ residues. The EBL dosage is then optimized by comparing the developed patterns for different dosage levels with the help of SEM.

We show some photonic structures fabricated with this recipe in Figure 3. Figure

3(a) shows the SEM of the HSQ pattern for a 500-nm-wide waveguide after development, using an optimized dosage of 3600 $\mu C/cm^2$ for such a dimension. The width of the HSQ pattern is monitored for various runs under the same fabrication condition, which falls in the range from 495 nm to 505 nm, indicating that the deviation from the design is less than 5 nm. This is a significant improvement compared to the ZEP case. The sidewall angle of the HSQ pattern, on the other hand, is about 75 degrees. Thanks to the high selectivity attainable in the plasma etching (~ 5), vertical sidewalls for silicon structures are still achievable, as shown by the SEM of the cross-section of the 500-nm-wide waveguide in Fig. 3(b). The HSQ is excessive for the etching of the 250 nm silicon layer and the remaining HSQ is on top of the waveguide, which can be removed by buffered oxide etcher (BOE). However, BOE also attacks the underneath oxide substrate, resulting in undercut structure, as illustrated by Fig. 3(c). For most applications, there is no need to remove the remaining HSQ mask, and its existence is experimentally verified to be minimal. Finally, Fig. 3(d) shows the top view of the waveguide. In contrast to the two edge lines shown in Fig. 2(d) for the ZEP process, here only one sharp edge is observed at each side of the waveguide, implying that the sidewall of the waveguide is indeed vertical.

2.3 Characterization of microresonators

The setup we use to characterize the fabricated samples is illustrated in Fig. 4(a), where the light coming out of a tunable laser is coupled to the waveguide through a tapered lens fiber and collected using another tapered lens fiber at the output of the waveguide before it is sent to a photo-detector. Wavelength is scanned at the laser source and transmission spectrum is measured. Figure 4(b) shows the picture of the boxed window in Fig. 4(a), where the input and output tapered lens fibers can be aligned to the on-chip waveguides through 3-D precision stages [36].

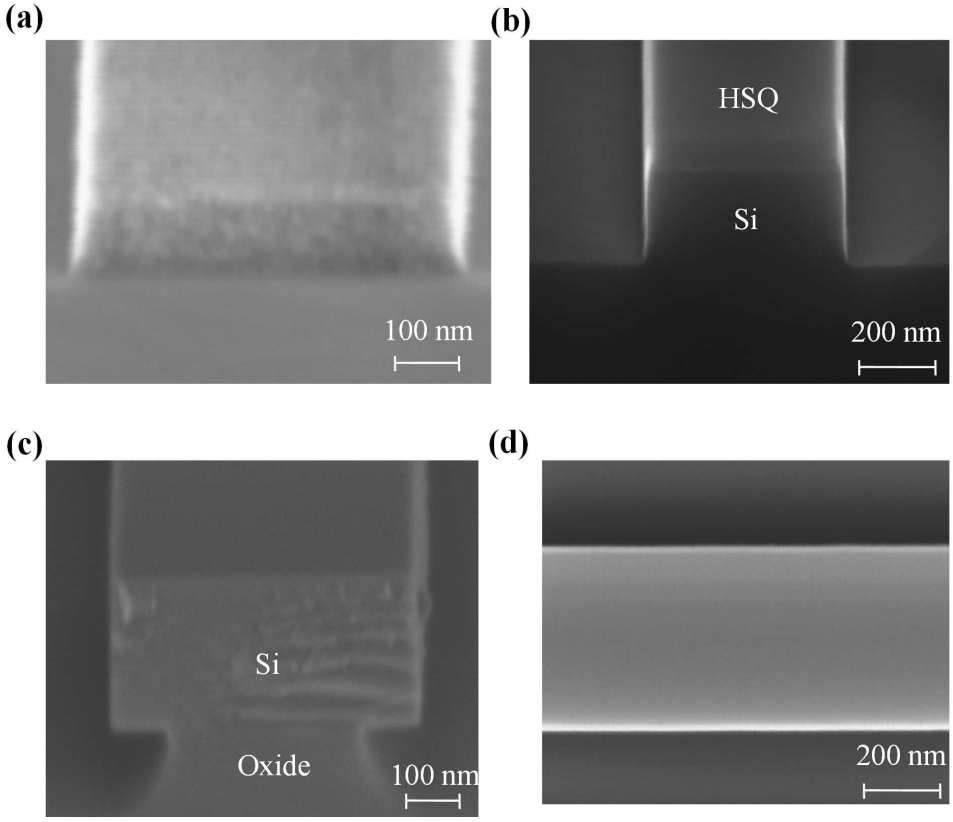

Figure 3: (a) SEM of the developed HSQ pattern for a 500-nm-wide waveguide; (b) cross-section of a silicon waveguide with 500 nm width and 250 nm height after etching, with remaining HSQ mask on top; (c) cross-section of the waveguide shown in (b) after BOE dip to strip the top HSQ; the underneath oxide substrate also gets undercut; (d) top view of the waveguide, with HSQ remaining on top.

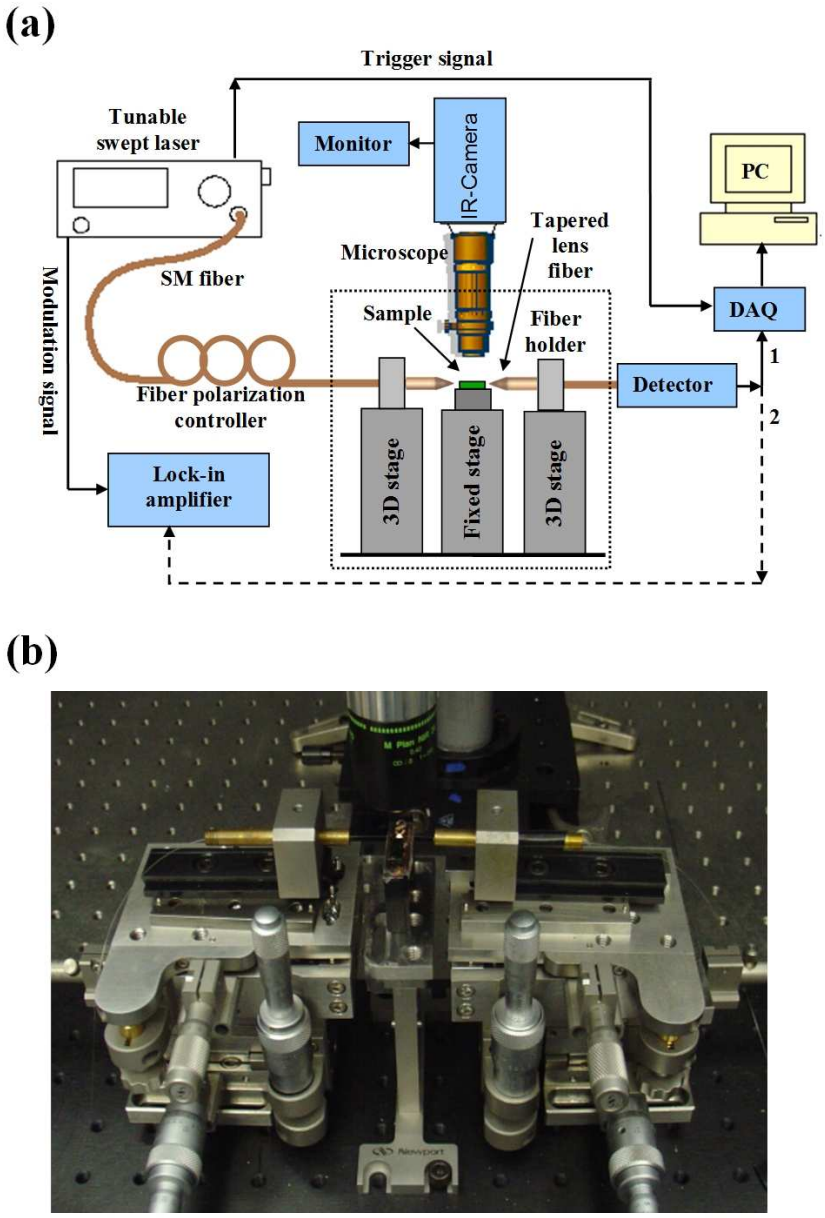

Figure 4: (a) Illustration of the tapered lens fiber setup used for microresonator characterization. (b) Portion of the setup that corresponds to the region in the dotted window in (a).

2.3.1 Insertion loss

One important parameter for the characterization process is the insertion loss (IL), which refers to the reduction of power from the laser to the photo-detector and is typically expressed in decibels (dB). Insertion loss can arise from many factors. For example, at the input and output stages, because the mode profiles between the tapered lens fiber and the on-chip waveguide are generally different, loss occurs due to an incomplete field overlap between these two modes. In addition, the difference of the propagation constants between the lens fiber mode and the waveguide mode can lead to an impedance mismatch, which also contributes some extra loss. Mathematically, these two loss terms can be expressed as [27]

$$\mathrm{IL_{in}} = \mathrm{IL_{field}}(\Delta \boldsymbol{r}) + \mathrm{IL_{imped}}, \tag{21}$$

with $\mathrm{IL_{field}}(\Delta \boldsymbol{r})$ given by

$$\mathrm{IL_{field}}(\Delta \boldsymbol{r}) = -10 \lg \left(\frac{\left| \int \boldsymbol{E}^*_{\mathrm{WG}}(\boldsymbol{r}) \times \boldsymbol{H}_{\mathrm{Fiber}}(\boldsymbol{r} + \Delta \boldsymbol{r}) \cdot d\boldsymbol{S} \right|^2}{\int \boldsymbol{E}^*_{\mathrm{Fiber}}(\boldsymbol{r}) \times \boldsymbol{H}_{\mathrm{Fiber}}(\boldsymbol{r}) \cdot d\boldsymbol{S} \int \boldsymbol{E}^*_{\mathrm{WG}}(\boldsymbol{r}) \times \boldsymbol{H}_{\mathrm{WG}}(\boldsymbol{r}) \cdot d\boldsymbol{S}} \right), \tag{22}$$

and $\mathrm{IL_{imped}}$ given by

$$\mathrm{IL_{imped}} = -10 \lg \left(1 - \left(\frac{\beta_{\mathrm{Fiber}} - \beta_{\mathrm{WG}}}{\beta_{\mathrm{Fiber}} + \beta_{\mathrm{WG}}} \right)^2 \right), \tag{23}$$

where $\boldsymbol{E}_{\mathrm{Fiber}}$ and $\boldsymbol{E}_{\mathrm{WG}}$ denote the electric fields of the tapered lens fiber and the waveguide, respectively; $\boldsymbol{H}_{\mathrm{Fiber}}$ and $\boldsymbol{H}_{\mathrm{WG}}$ are the corresponding magnetic fields; $\Delta \boldsymbol{r}$ is the displacement between the mode centers of $\boldsymbol{E}_{\mathrm{Fiber}}$ ($\boldsymbol{H}_{\mathrm{Fiber}}$) and $\boldsymbol{E}_{\mathrm{WG}}$ ($\boldsymbol{H}_{\mathrm{WG}}$) which depends on the alignment; $\boldsymbol{S}$ is the integration area that is perpendicular to the wave propagating direction; and β_{Fiber} and β_{WG} denote the propagation constants of the tapered lens fiber mode and waveguide mode, respectively. Equation 23 is derived based on the Fresnel equations under the normal incidence condition.

For samples fabricated on SOI platforms, the light propagating in the waveguide has the possibility of leaking into the underneath silicon substrate if the buried oxide

layer (BOX) is not thick enough. This loss is illustrated in Fig. 5, where $\mathrm{IL}_{\mathrm{sub}}$ denotes the substrate leakage loss. Other insertion losses include the propagation loss resulting from scattering due to rough sidewalls and the surface absorption due to unsaturated surface states. Finally, the device itself (other than the access waveguide) will introduce an additional insertion loss.

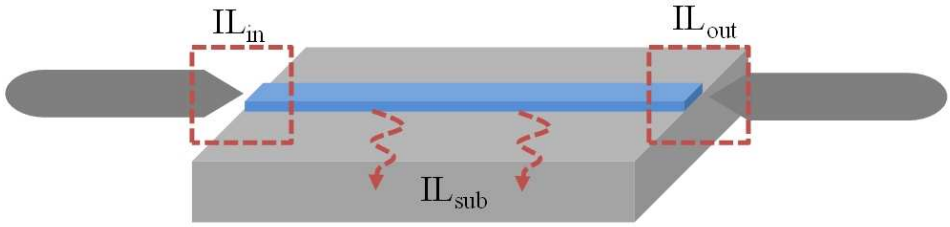

Figure 5: Illustration of the insertion loss occurring at the input ant output stages and also the substrate leakage loss.

2.3.2 Insertion loss from input and output stages

The field coming out of the tapered lens fiber can be approximated by a Gaussian beam propagating in the free space. To achieve an optimal coupling, the distance between the tapered lens fiber and the sample's facet has to be adjusted such that $\boldsymbol{H}_{\mathrm{Fiber}}$ is at its beam waist when it reaches the waveguide input, where its expression is assumed to be [27]

$$\boldsymbol{H}_{\mathrm{Fiber}} = \frac{2}{w}\frac{1}{\sqrt{2\pi}} e^{-r^2/w^2} e^{ik_0 z}, \tag{24}$$

with w being the spot size and z being the wave propagating direction. $\mathrm{IL}_{\mathrm{field}}(\Delta\boldsymbol{r})$ can then be computed from Eq. 22. One such example is given in Fig. 6 for the fundamental TE-polarized (electric field parallel to the device plane) mode of a silicon waveguide with a width of 500 nm and a height of 250 nm. We have assumed the polarization of $\boldsymbol{H}_{\mathrm{Fiber}}$ has been aligned to that of $\boldsymbol{H}_{\mathrm{WG}}$, whose mode profile is obtained through an in-house waveguide mode solver [36] and shown in Fig. 6(a). In Fig. 6(b), $\mathrm{IL}_{\mathrm{field}}(\Delta\boldsymbol{r})$ is plotted as a function of the displacement $\Delta\boldsymbol{r}$. It is observed

that $\mathrm{IL}_{\mathrm{field}}(\Delta \boldsymbol{r})$ is mostly dependent on the distance between the mode centers of the waveguide and the fiber (i.e., Δr), but not much on the angles of the displacement. In Fig. 6(c), which displays the result of $\mathrm{IL}_{\mathrm{field}}(\Delta \boldsymbol{r})$ as a function of the distance Δr, we conclude that a misalignment of 0.75 μm will lead to a degradation of coupling efficiency by 3 dB.

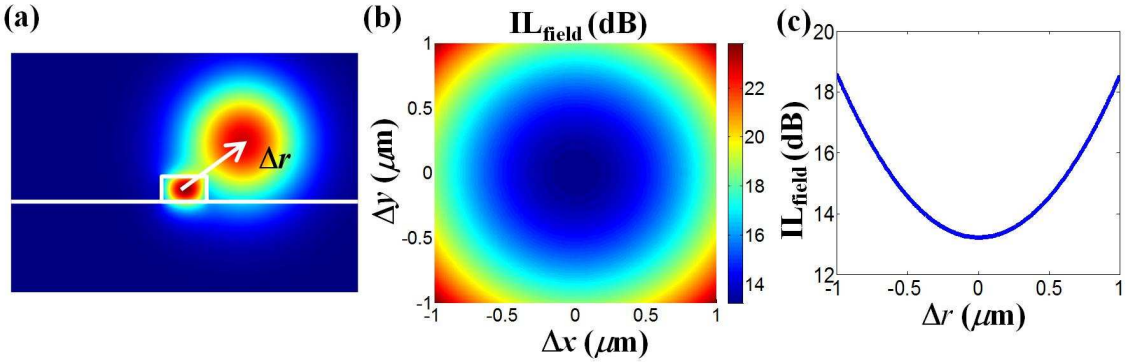

Figure 6: (a)Mode profiles of $\boldsymbol{H}_{\mathrm{Fiber}}$ and $\boldsymbol{H}_{\mathrm{WG}}$ with a displacement of $\Delta \boldsymbol{r}$ between the two mode centers; (b) $\mathrm{IL}_{\mathrm{field}}(\Delta \boldsymbol{r})$ as a function of $\Delta \boldsymbol{r}$ for a silicon waveguide with a width of 500 nm and a height of 250 nm in the air cladding. The spot size w of $\boldsymbol{H}_{\mathrm{Fiber}}$ is assumed to be 1.25 μm; (c) $\mathrm{IL}_{\mathrm{field}}(\Delta \boldsymbol{r})$ as a function of the distance Δr.

In Fig. 7(a), the effect of different waveguide geometries has been investigated, where the minimum of $\mathrm{IL}_{\mathrm{field}}$ for each waveguide geometry (i.e., $\Delta \boldsymbol{r} = 0$) in both the air and oxide cladding has been plotted. The result shown in Fig. 7(a) suggests that waveguides with smaller dimensions have a larger mode volume and better field overlap with the tapered lens fiber, resulting in a smaller intertion loss. In practice, access waveguides are usually tapered down at the input and output to form the so-called nanotapers. Most of our designs use w=250 nm waveguide width when a low insertion loss is required.

The computation of $\mathrm{IL}_{\mathrm{imped}}$ is straightforward from Eq. 23, where $\beta_{\mathrm{Fiber}} = k_0$, and $\beta_{\mathrm{WG}} = k_0 n_{\mathrm{eff}}$ with n_{eff} representing the effective index of the waveguide mode. Numerical results for the waveguide dimensions shown in Fig. 7(a) are provided in Fig. 7(b). Comparing $\mathrm{IL}_{\mathrm{imped}}$ to $\mathrm{IL}_{\mathrm{field}}$, we conclude $\mathrm{IL}_{\mathrm{imped}}$ is much smaller and is

not the dominant factor for the insertion loss at the input and output stages.

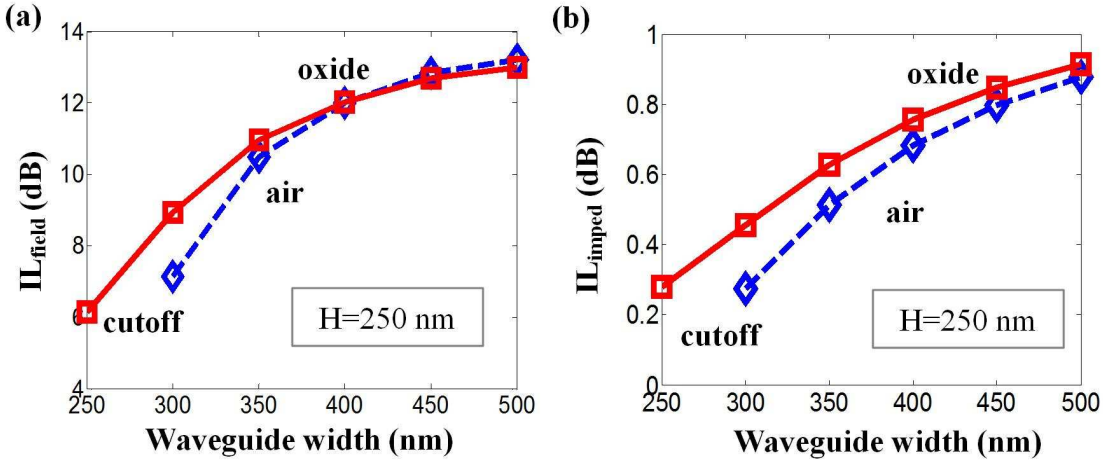

Figure 7: (a) IL_{field} under the ideal alignment condition for different silicon waveguides with a fixed height of 250 nm and different widths. The red and blue curves correspond to the oxide and air claddings, respectively. Both polarizations for the waveguide mode and the tapered lens fiber modes are assumed to be TE. (b) The corresponding IL_{imped} for waveguides shown in (a).

2.3.3 Substrate leakage loss

As we have mentioned before, the substrate leakage loss only occurs when the BOX is not thick enough to prevent optical power from leaking to the substrate. Typical BOX layer thickness ranges from 1 μm to 3 μm. For most of devices, 3 μm BOX is thick enough to eliminate the leakage while for 1 μm BOX, the leakage loss might be strong under certain circumstances. We investigate the substrate leakage loss using the RF module available in the commercial software COMSOL. The simulated structure is illustrated in Fig. 8, where perfectly matched layers (PMLs) are implemented under the substrate to truncate the simulation window. By solving the eigenvalue problem, complex propagation constants are obtained whose imaginary part corresponds to the substrate leakage loss. Figure 9(a) shows the simulation results for silicon waveguides with a height of 220 nm and various widths on a 1-μm-thick BOX. As can be observed, for the TM-polarized mode in an air cladding, the leakage loss is more than 10 dB/cm when the waveguide width is smaller than 500 nm. By contrast, the leakage loss

becomes much less if the top cladding becomes oxide. This is because the higher refractive index of the cladding material pulls up the mode profile so that the leakage through the 1-μm-thick BOX to the substrate is significantly reduced. On the other hand, for the TE-polarized mode, the leakage loss is always small (< 0.25 dB/cm) in both the air and oxide claddings, due to the fact that its mode is much more confined.

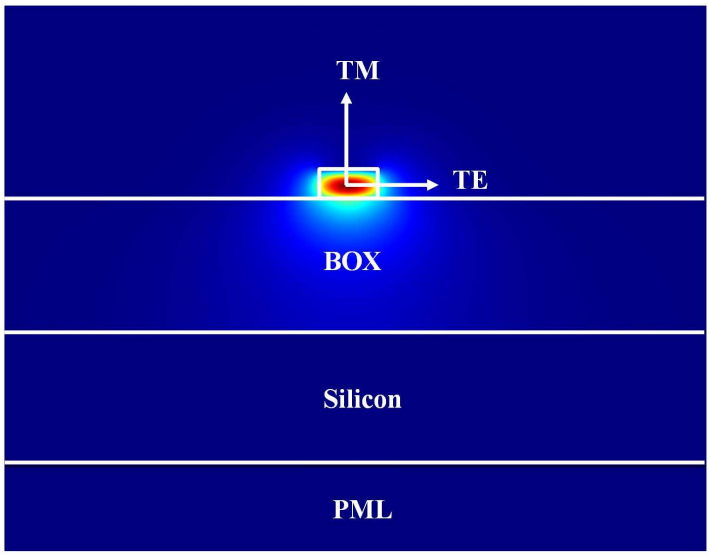

Figure 8: Cross section of the waveguide structure for the simulation of substrate leakage loss.

From the above example, we conclude that the substrate leakage loss is critically dependent on the mode confinement conditions. It is then a trivial extension that increasing the silicon thickness can also reduce the substrate leakage. In addition, since the waveguide modes typically exhibit strong dispersions, we expect that the substrate leakage loss will vary with the wavelength. In Fig. 9(b), two silicon thicknesses are studied, one with 220 nm and the other with 250 nm, for the TM-polarized mode in the air cladding. The results indicate that the substrate leakage loss increases with the wavelength, which is reasonable since the mode confinement is weaker at longer wavelengths. Also, for the TM mode, if the thickness is more than 250 nm,

the substrate leakage loss becomes small enough (< 1 dB/cm) for waveguides with width more than 450 nm.

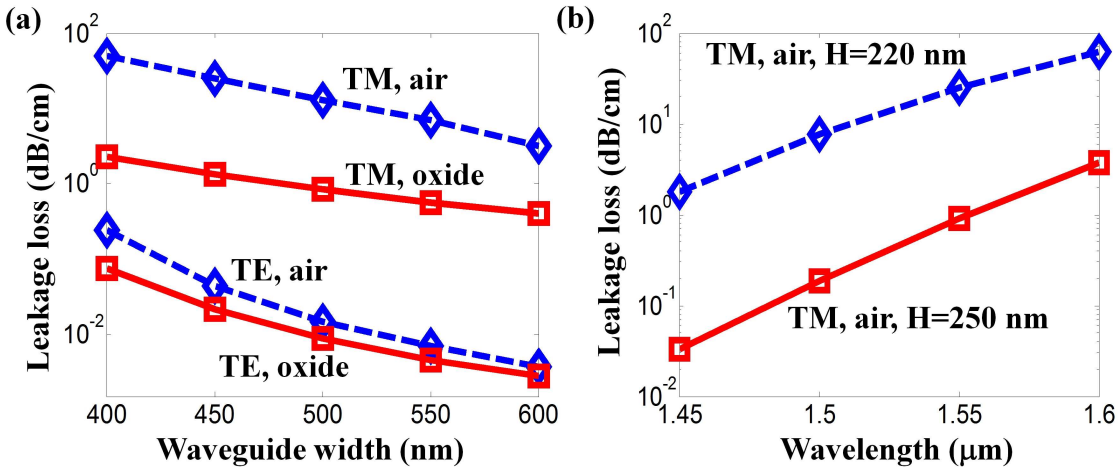

Figure 9: (a) Substrate leakage loss for silicon waveguides with height of 220 nm and 1 μm BOX at the wavelength of 1550 nm. The red and blue curves correspond to the oxide and air claddings, respectively. (b)Substrate leakage losses for the TM mode of silicon waveguides with width of 450 nm and air cladding. Two different heights (220 nm and 250 nm) are studied.

2.3.4 Polarization control

Unlike fiber-based communication links where polarization insensitivity is sought, the majority of integrated photonic circuits based on high-index-contrast material platforms are only designed for one polarization. To properly characterize the device, light with the correct polarization has to be sent into the access waveguide. However, in the setup shown in Fig. 4(a), only one polarization rotator is available in the optical path, which can rotate the polarization of light but the exact polarization state coming out of the tapered lens fiber is generally unknown.

For some waveguide geometries, it is possible to differentiate between the TE and TM polarizations using their insertion losses. For example, for the 1-μm-thick BOX and air cladding, the TM mode suffers a strong substrate leakage loss for waveguides with width less than 500 nm while the substrate leakage loss for the TE mode is generally negligible. On the other hand, as shown in Fig. 10(a), the insertion loss

at the input (or the output) stage for the TM mode is about 5-6 dB smaller than that of the TE. In Fig. 10(b), we calculate the difference of the total insertion loss between the TE and TM modes for a 6-mm-long waveguide with a 1 μm BOX and 220 nm silicon thickness. We consider the loss arising from the input and output stages as well as the substrate leakage loss; other loss terms are assumed to be either negligible or identical for the two polarizations. From Fig. 10, it is observed that for the air cladding, the insertion loss of the TM mode is much larger (mainly due to large substrate leakage loss) than that of the TE mode if the waveguide width is less than 550 nm and the measurement is done at the wavelength of 1600 nm. The TE and TM polarizations can then be found by maximizing and minimizing the transmission when adjusting the polarization rotator, respectively. The situation is different for the wavelength of 1550 nm, where the contrast between the TE and TM modes is small for waveguides with width larger than 450 nm. Consequently, such an approach based on the insertion loss to distinguish the polarization may not be applicable. In the presence of the top oxide cladding, the TM mode has about 10-12 dB less insertion loss than the TE, thanks to the reduced substrate leakage loss and better field overlap with the tapered lens fiber mode. Therefore, the TE and TM polarizations can be found by minimizing and maximizing the transmission when adjusting the polarization rotator, respectively.

2.3.5 Laser noise, scan linearity and repeatability

The tunable laser employed in the setup shown in Fig. 4 plays a central role in the spectral characterization of microresonators. Among its various properties that may affect the measurement results, here we discuss three important issues, namely, the laser noise, scan repeatability and linearity.

From elementary lasing theory, it is well known that accompanying the lasing process there is always the inevitable spontaneous emission. Consequently, a certain

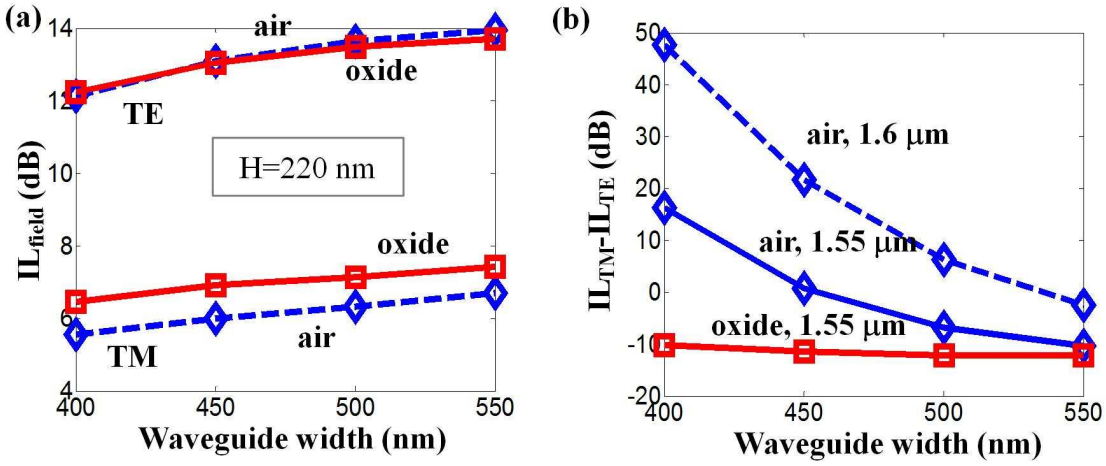

Figure 10: (a)IL_{field} for the TE and TM polarizations for silicon waveguides with height of 220 nm and 1 μm BOX, in both the air and oxide cladding. (b) The difference of the total insertion loss between the TE and TM polarizations for silicon waveguides with height of 220 nm and 1 μm BOX.

amount of incoherent light exists in the laser output that acts as noise. In Fig. 11(a), we show the transmission result for a 10-μm-radius silicon microdisk resonator under different power settings from an Agilent 81680A tunable laser. To avoid any nonlinear effects, we have misaligned the input tapered lens fiber so that the power sent into the waveguide is low enough. As can be seen from Fig. 11(a) (the resonances have been intentionally shifted for a better illustration), the extinction ratio of the same resonance increases with the laser output power and becomes saturated when the power of the laser is more than 200 μW. In addition, it has been observed in the experiment that the transmission result is independent of the actual power sent into the resonator, as long as it is not too strong to incur any nonlinear phenomenon. In other words, as we vary IL_{in} or IL_{out}, the transmission curve stays the same after power normalization (i.e., the maximum is 0 dB). This finding leads us to conclude that there is a certain amount of wideband incoherent light from the laser which arises from the spontaneous emission. Its impact to the measurement becomes smaller if the laser output increases, which agrees with the lasing theory and has been verified in Fig. 11(a). Thus, to properly characterize the device at a low injection power, the

laser is better to be set at a high working power while using external attenuations (such as having a large IL_{in} by intentional misalignment) to reduce the power coupled to the device.

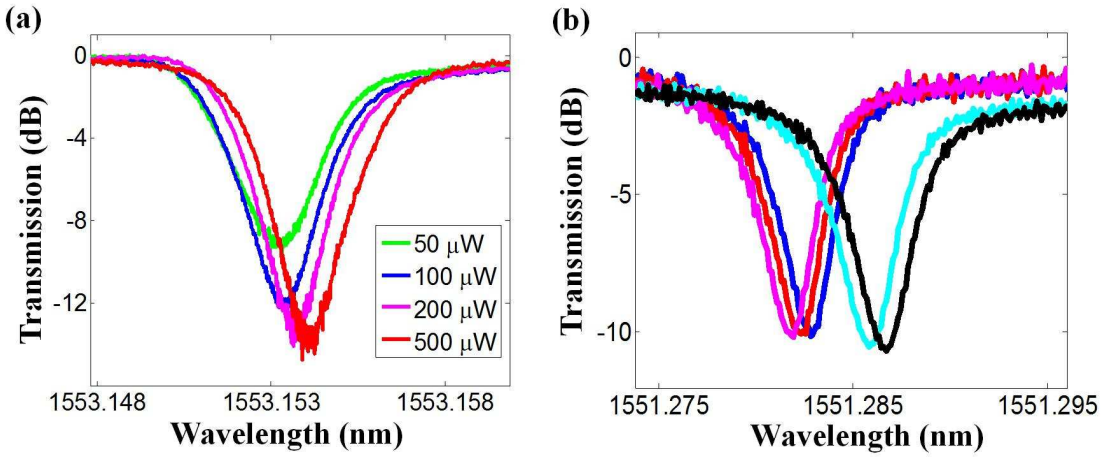

Figure 11: (a) Transmission response of a 10-μm-radius silicon microdisk resonator with height of 220 nm in the air cladding. Curves with different colors correspond to different laser setting powers, and IL_{in} has been adjusted for each case so the power coupled into the access waveguide is almost the same. (b) Five repeated transmission measurements for one resonance of a 10-μm-radius silicon microdisk resonator in a short time.

Scan repeatability refers to how well we can repeat the transmission measurement for the same device. One can understand the necessity to consider this issue from the wavelength tuning mechanism of diode lasers, which is based on piezo tuning of the cavity length. Because of the hysteresis associated with the piezo, the starting point of each scan is not exactly the same. In Fig. 11(b), we show five measurements for the same resonance of a 10-μm-radius microdisk resonator repeated in a short time. The result demonstrates that the variation of wavelength for each scan is on the order of a few picometers. Moreover, if the time interval between different scans is relatively long (typically more than a few minutes), we have also to consider the effect of the ambient environment, particularly the temperature. One degree change of temperature can shift the resonance of silicon resonators by tens of picometers, and it will also change the wavelength of the tunable laser if it has not been calibrated.

Finally, we want to discuss the scan linearity of the tunable laser, an issue that is also related to the piezo tuning of the cavity length. Usually, the laser has been designed such that the output wavelength is tuned as linearly with time as possible during the scan process. Here we use the on-chip resonator to access the scan linearity of the tunable laser. In Fig. 12(a), we plot the transmission measurement of a 10-μm-radius microdisk resonator by black curve. Starting with an arbitrary resonance and an estimated FSR (either from simulation or experience, which does not have to be very accurate), we can find its radial mode family with different azimuthal orders. Two such examples, which are marked by red and green curves, are illustrated in Fig. 12(a). Then, we measure the FSRs for each radial mode family and convert them to group indices using Eq. 8. The results are provided in Fig. 12(b), with the red and green squares corresponding to the radial mode families marked by the curves of the same color in Fig. 12(a), respectively. To identify their radial mode orders, we also plot the simulation results of the group indices for the lowest five radial modes by black lines in Fig. 12(b). It is straightforward to conclude from there that the red and green squares represent resonances of the second- and fifth-order radial modes, respectively. The good agreement between the experimental measurement and the numerical simulation demonstrates the excellent scan linearity of the tunable laser. In fact, if the error in the measured FSRs (resulting from the scan nonlinearity) is more than a few picometers, there will be a noticeable deviation of the measured result from the simulation. Especially, instead of showing a linear dispersion with the wavelength, the measured group indices will have irregular variations (spikes and dips) around their true values.

To summarize, when performing the characterization, the tunable laser has to be set at a high output power rather than a small one to reduce the impact of spontaneous emission. The laser also has a random drift from the set point which is on the order of a few picometers when starting a scan. In addition, the scan linearity of the tunable

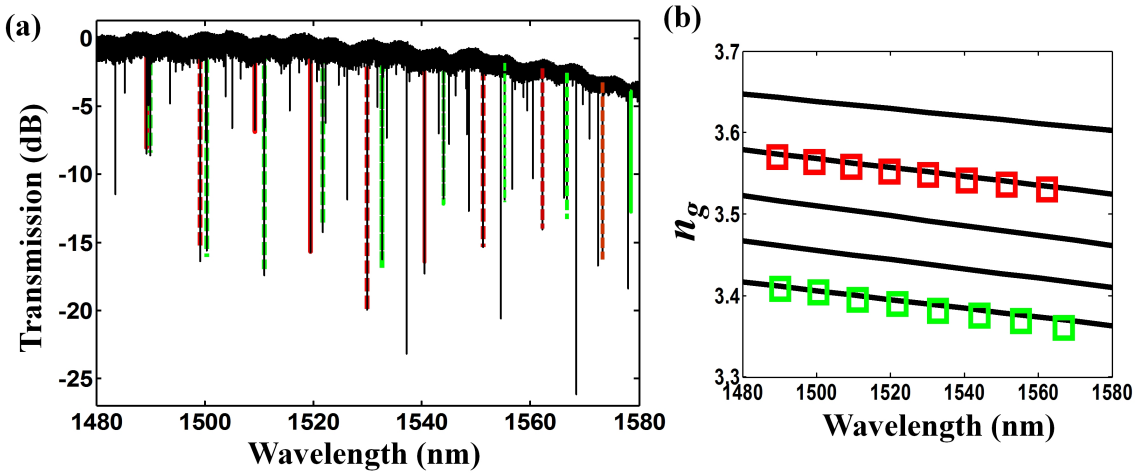

Figure 12: (a) Transmission response of a 10-μm-radius silicon microdisk resonator with height of 220 nm in the air cladding. The red and blue curves mark two different radial mode families. (b) Measured (red and blue squares) group indices for the resonances marked in (a) and also the simulation results (black lines) for the lowest five radial modes. All of them are TE polarized.

laser is generally good, and therefore the measured FSRs can be trusted to a good degree.

CHAPTER III

THEORETICAL AND NUMERICAL INVESTIGATION OF PHOTONIC DEVICES

3.1 Introduction

Theoretical modeling and numerical simulation play an important role in the fundamental studies and real applications of photonic devices. A good model will not only offer a thorough understanding on the device itself, but also may provide valuable insights on some novel applications. However, unlike electronic circuit design where standard simulation toolboxes are available, for photonic circuits it is often necessary to develop customized codes for specific tasks. Moreover, direct numerical simulations may be difficult for some structures and analytical solutions have to be sought.

In this chapter, we will focus on three important problems. The first problem is on the power coupling between photonic devices. In the second problem, we study the phase shift caused by the coupling process, which is important for coupled-resonator devices since the phase shift will change the resonance frequency of the coupled resonator. In the third problem, mode splitting and scattering loss are studied, for which a unified model is developed.

3.2 Power coupling between photonic structures

Except for a few simple structures, a direct numerical simulation of the power coupling coefficient between photonic devices is usually a formidable task due to the enormous need of computational resources. Instead, designers usually simulate the mode of each individual element first and then use the coupled mode theory (CMT) to estimate the power transfer rate between two coupled structures [36]. There are several versions

of CMT, which we will briefly review in the following. A modified CMT is then proposed for the TM-polarized (transverse magnetic) light with a focus on the high-index-contrast material systems. Numerical examinations reveal that the modified CMT provides more accurate results compared with the existing ones.

3.2.1 Existing coupled mode theory

We use the basic structure of two parallel waveguides for the explanation of the existing CMT theories. For the two waveguides shown in Fig. 13, we assume their modes are $\boldsymbol{E}_1 = \boldsymbol{E}_1(\boldsymbol{r}_T)\exp(i\beta_1 z)$ and $\boldsymbol{E}_2 = \boldsymbol{E}_2(\boldsymbol{r}_T)\exp(i\beta_2 z)$, where z is the wave propagation direction and $\boldsymbol{r}_T$ denotes the transverse spatial variables (i.e., x and y in Fig. 13(a)); $\boldsymbol{E}_1(\boldsymbol{r}_T)$ and $\boldsymbol{E}_2(\boldsymbol{r}_T)$ are the electric fields of the individual waveguide modes 1 and 2, respectively; and β_1 and β_2 are their corresponding propagation constants. From Maxwell's equations, we have

$$\nabla \times (\nabla \times \boldsymbol{E}) - k_0^2 \varepsilon_r \boldsymbol{E} = 0, \tag{25}$$

where ε_r is the relative permittivity which is related to the refractive index n as $\varepsilon_r = n^2$, and $k_0 \equiv \omega/c$ is the wavenumber of light in vacuum. After some arrangements, Eq. 25 can be rewritten as

$$\nabla(\nabla \cdot \boldsymbol{E}) - \left(\nabla_T^2 + \frac{d^2}{dz^2}\right)\boldsymbol{E} - k_0^2 \varepsilon_r \boldsymbol{E} = 0, \tag{26}$$

where ∇_T^2 denotes a Laplacian operated on the transverse parameters only.

First we consider the TE (transverse electric) polarization of slab waveguides, where only the y component of the electric field is nonzero and uniform in that direction (see Fig. 13(a)). In consequence, $\nabla \cdot \boldsymbol{E} = 0$ in Eq. 26. We assume the total electric field can be expressed as $\boldsymbol{E} = a_1(z)\boldsymbol{E}_1 + a_2(z)\boldsymbol{E}_2$ with $a_1(z)$ and $a_2(z)$ being the amplitudes. After the substitution, the so-called slowly varying envelop approximation is used to further simplify the equation: $|a''(z)| \ll k_0|a'(z)|$ so the

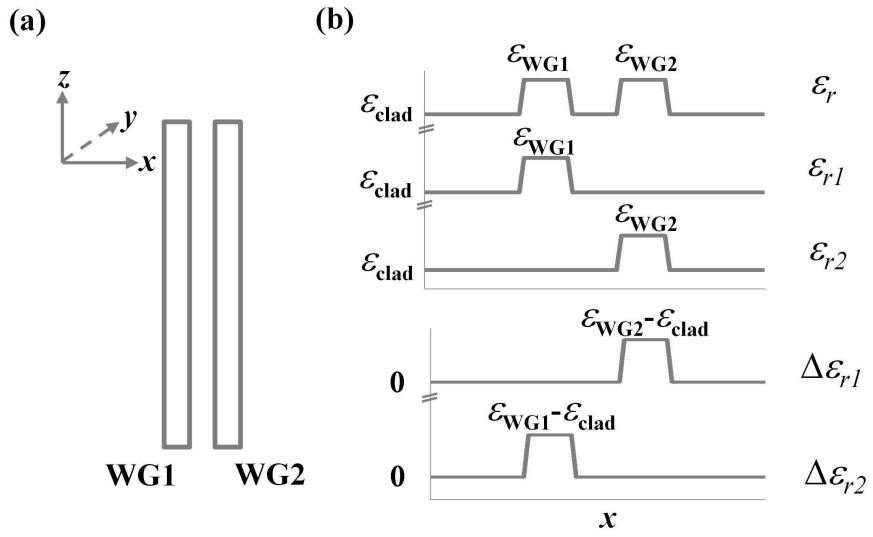

Figure 13: (a) Illustration of two coupled waveguides; (b) ε_r, ε_{r1}, ε_{r2}, $\Delta\varepsilon_{r1}$, and $\Delta\varepsilon_{r2}$ as a function of x.

second-order derivative can be neglected. As a result, we arrive at

$$2i\beta_1\frac{da_1(z)}{dz}\boldsymbol{E}_1(\boldsymbol{r}_T)e^{i\beta_1 z}+2i\beta_2\frac{da_2(z)}{dz}\boldsymbol{E}_2(\boldsymbol{r}_T)e^{i\beta_2 z}=-k_0^2(\varepsilon_r-\varepsilon_{r1})a_1(z)\boldsymbol{E}_1(\boldsymbol{r}_T)e^{i\beta_1 z} - k_0^2(\varepsilon_r-\varepsilon_{r2})a_2(z)\boldsymbol{E}_2(\boldsymbol{r}_T)e^{i\beta_2 z}, \tag{27}$$

where ε_r is the relative permittivity of the whole waveguide system while ε_{r1} and ε_{r2} are defined for the individual waveguides 1 and 2, respectively (see Fig. 13(b)); and we have used the fact that $\boldsymbol{E}_n(\boldsymbol{r}_T)(n=1,2)$ satisfies Eq. 26 with the corresponding relative permittivity ε_{rn} substituted.

By projecting Eq. 27 onto $\boldsymbol{E}_n(\boldsymbol{r}_T)(n=1,2)$ and performing an integration over the transverse area, we obtain equations for $a_1(z)$ and $a_2(z)$. For example, by projecting on $\boldsymbol{E}_1(\boldsymbol{r}_T)$, we have

$$2i\frac{da_1(z)}{dz}e^{i\beta_1 z}\int\beta_1\boldsymbol{E}_1^*(\boldsymbol{r}_T)\cdot\boldsymbol{E}_1(\boldsymbol{r}_T)\,d^2\boldsymbol{r}_T+2i\frac{da_2(z)}{dz}e^{i\beta_2 z}\int\beta_2\boldsymbol{E}_1^*(\boldsymbol{r}_T)\cdot\boldsymbol{E}_2(\boldsymbol{r}_T)\,d^2\boldsymbol{r}_T = -k_0^2a_1(z)e^{i\beta_1 z}\int\Delta\varepsilon_{r1}\boldsymbol{E}_1^*(\boldsymbol{r}_T)\cdot\boldsymbol{E}_1(\boldsymbol{r}_T)\,d^2\boldsymbol{r}_T-k_0^2a_2(z)e^{i\beta_2 z}\int\Delta\varepsilon_{r2}\boldsymbol{E}_1^*(\boldsymbol{r}_T)\cdot\boldsymbol{E}_2(\boldsymbol{r}_T)\,d^2\boldsymbol{r}_T, \tag{28}$$

where $\Delta\varepsilon_{r1}\equiv\varepsilon_r-\varepsilon_{r1}$ and $\Delta\varepsilon_{r2}\equiv\varepsilon_r-\varepsilon_{r2}$, which are illustrated in Fig. 13(b).

A further simplification can be carried out as follows. From Maxwell's equations, for the TE-polarized mode, $H_{nx} = -\beta_n E_{ny}/\omega\mu\,(n = 1, 2)$ with μ being the vacuum permeability. Therefore,

$$\begin{aligned}\int \beta_m \boldsymbol{E}_n^*(\boldsymbol{r}_T) \cdot \boldsymbol{E}_m(\boldsymbol{r}_T)\, d^2\boldsymbol{r}_T &= \int \beta_m E_{ny}^*(\boldsymbol{r}_T) E_{my}(\boldsymbol{r}_T)\, d^2\boldsymbol{r}_T \\ &= -\omega\mu \int E_{ny}^*(\boldsymbol{r}_T) H_{mx}(\boldsymbol{r}_T)\, d^2\boldsymbol{r}_T \\ &= \omega\mu \int \boldsymbol{E}_n^*(\boldsymbol{r}_T) \times \boldsymbol{H}_m(\boldsymbol{r}_T) \cdot \hat{\boldsymbol{z}}\, d^2\boldsymbol{r}_T \quad .\end{aligned} \tag{29}$$

For the case of $n = m$, the integrand in the last line of Eq. 29 is identified as the Poynting vector of the waveguide mode n, and its integration over the transverse area gives the propagating power. A convenient assumption is that each field has been normalized so it has unit power. For the case of $n \neq m$, the integration in Eq. 29 is generally nonzero, indicating the two waveguide modes are not orthogonal to each other.

Finally, we define $\bar{a}_1(z) \equiv a_1(z)\exp(i\Delta\beta z/2)$ and $\bar{a}_2(z) \equiv a_2(z)\exp(-i\Delta\beta z/2)$ with $\Delta\beta \equiv \beta_1 - \beta_2$, for which we obtain the following equation set after applying the result of Eq. 29 to Eq. 28:

$$\begin{bmatrix} 1 & X_{12} \\ X_{21} & 1 \end{bmatrix} \frac{d}{dz} \begin{bmatrix} \bar{a}_1(z) \\ \bar{a}_2(z) \end{bmatrix} = i \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix} \begin{bmatrix} \bar{a}_1(z) \\ \bar{a}_2(z) \end{bmatrix}, \tag{30}$$

where X_{nm} and k_{nm} $(n, m = 1, 2)$ are defined as

$$X_{nm} = \frac{1}{2} \int \boldsymbol{E}_n^*(\boldsymbol{r}_T) \times \boldsymbol{H}_m(\boldsymbol{r}_T) \cdot \hat{\boldsymbol{z}}\, d^2\boldsymbol{r}_T, \tag{31}$$

$$k_{nm} = \frac{\Delta\beta}{2}(-1)^{n+1}\delta_{nm} + \frac{\omega\varepsilon_0}{4} \int \Delta\varepsilon_{r_m} \boldsymbol{E}_n^*(\boldsymbol{r}_T) \cdot \boldsymbol{E}_m(\boldsymbol{r}_T)\, d^2\boldsymbol{r}_T, \tag{32}$$

where $i = \sqrt{-1}$, δ_{nm} is the Kronecker delta function, and we have assumed unit power for each waveguide mode as

$$\frac{1}{2} \int \boldsymbol{E}_n^*(\boldsymbol{r}_T) \times \boldsymbol{H}_n(\boldsymbol{r}_T) \cdot \hat{\boldsymbol{z}}\, d^2\boldsymbol{r}_T = 1, \quad (n = 1, 2). \tag{33}$$

From Eq. 31, it is easy to verify that X_{11} and X_{22} are both unity, and X_{12} and X_{21} are generally nonzero, a fact arising from the non-orthogonality of the two waveguide modes.

The above results (mainly Eqs. 31-32) are derived for the TE-polarized light. However, if we assume the electromagnetic fields of the coupled structure can be expressed as a sum of the modes from each individual waveguide, that is,

$$\boldsymbol{E}_T(\boldsymbol{r}) = a_1(z)\boldsymbol{E}_{1T}(\boldsymbol{r}) + a_2(z)\boldsymbol{E}_{2T}(\boldsymbol{r}), \tag{34}$$

$$\boldsymbol{H}_T(\boldsymbol{r}) = a_1(z)\boldsymbol{H}_{1T}(\boldsymbol{r}) + a_2(z)\boldsymbol{H}_{2T}(\boldsymbol{r}), \tag{35}$$

where the subscript $\boldsymbol{T}$ denotes the transverse part of the field [1]. Then it is shown by Refs. [27, 38] that Eqs. 31-32 are valid for the general case.

On the other hand, the old version of CMT is derived by treating the adjacent waveguide as a first-order perturbation to the waveguide mode under consideration and neglecting the fact that the two waveguide modes are generally not orthogonal. Mathematically, in Eq. 30, X_{21} and X_{12} are equal to zero in the old CMT.

Therefore, it is expected that the CMT with the X_{nm} $(n, m = 1, 2)$ factors should be more accurate, which has been verified through numerical simulations for the TE-polarized light [38]. Surprisingly, when considering the TM polarization for high-index-contrast material systems, it is found in Refs. [39, 40] that the old CMT without the X_{nm} factors still works quite well while the CMT embracing the cross power overlaping factors have significant errors. This unexpected result is explained by Ref. [39] to be caused by the assumption made in Eq. 34. In high-index-contrast material systems, the component of the electric field that is normal to the waveguide surface is discontinuous, and it is easy to verify that the total electric field given by Eq. 34 fails to obey the appropriate boundary conditions. For example, on the surface of waveguide 2, E_{2x} satisfies the correct boundary condition, but E_{1x} does not since

[1] we can only specify the relation on the transverse part because Maxwell's euqations dictate that the transverse and longitudinal components of an electromagnetic field are dependent.

it is continuous outside of waveguide 1; therefore, E_x given by Eq. 34 does not satisfy the required boundary condition. The same is true for E_x on the surface of waveguide 1.

In our work, high-index-contrast material systems such as silicon-on-insulator platforms are employed. The device layer is usually thin with a typical thickness of silicon around 220 nm. As a result, the fundamental mode of the fabricated photonic devices has its predominant electric field parallel to the device layer. When we use the effective-index method to remove the vertical dimension (i.e., y in Fig. 13(a)), this fundamental mode corresponds to the TM polarization based on its definition [2]. In practice, the old-version CMT is typically employed to design the coupling between coupled devices, which turns out to agree with the experimental results well. Nevertheless, it remains to be clarified why the old CMT without the X_{nm} factors still works despite the fact that it is a very crude model. More importantly, we want to know whether there is any better formula for the coupling calculation, a subject to be addressed in the following subsection.

3.2.2 Modified CMT for the TM polarization

As we have pointed out, the incorrect boundary assumption imposed by Eq. 34 is responsible for the problems of the CMT incorporating the X_{nm} factors. Since the magnetic field $\boldsymbol{H}$ is continuous in dielectric waveguides, Eq. 35 is still a good approximation. Furthermore, we will show that it is the only approximation needed for the TM polarized light, following which the electric field automatically satisfies the required boundary conditions.

For the TM polarized light, the only nonzero magnetic field component H_y obeys

[2]However, in the 3-D sense, it is often called TE-like polarization. We suggest readers to use the context to avoid any confusion.

the following Maxwell's equation

$$\frac{\partial}{\partial x}\left(\frac{\partial}{\varepsilon_r \partial x}H_y\right)+\frac{\partial}{\partial z}\left(\frac{\partial}{\varepsilon_r \partial z}H_y\right)+k_0^2 H_y=0, \tag{36}$$

where ε_r is only a function of x. According to Eq. 35, we assume that H_y can be approximated by the sum of the individual waveguide modes as

$$H_y(x,z)=a_1(z)H_{1y}(x)e^{i\beta_1 z}+a_2(z)H_{2y}(x)e^{i\beta_2 z}, \tag{37}$$

where H_{1y} and H_{2y} are the magnetic fields of waveguides 1 and 2, respectively, and β_1 and β_2 are their corresponding propagation constants. Note that such an approximation is good for single-mode waveguides, since all we have neglected are the radiation modes which are typically very small. Substituting Eq. 37 into Eq. 36, and using the fact that $H_{1y}(x)e^{i\beta_1 z}$ and $H_{2y}(x)e^{i\beta_2 z}$ satisfy Eq. 36 too if ε_r is replaced by ε_{r1} and ε_{r2}, respectively, we obtain

$$\begin{bmatrix} X'_{11} & X'_{12} \\ X'_{21} & X'_{22} \end{bmatrix} \frac{d}{dz} \begin{bmatrix} \bar{a}_1(z) \\ \bar{a}_2(z) \end{bmatrix} = i \begin{bmatrix} k'_{11} & k'_{12} \\ k'_{21} & k'_{22} \end{bmatrix} \begin{bmatrix} \bar{a}_1(z) \\ \bar{a}_2(z) \end{bmatrix}, \tag{38}$$

which is similar to Eq. 30 but with different coefficients of X'_{nm} and k'_{nm} $(n,m=1,2)$ as

$$X'_{nm}=\frac{1}{2}\int \frac{\varepsilon_{rm}}{\varepsilon_r}\boldsymbol{E}_m(\boldsymbol{r}_T)\times\boldsymbol{H}_n^*(\boldsymbol{r}_T)\cdot\hat{\boldsymbol{z}}\,d^2\boldsymbol{r}_T, \tag{39}$$

$$k'_{nm}=\frac{\Delta\beta}{2}(-1)^{n+1}\delta_{nm}+\frac{\omega\varepsilon_0}{4}\int \frac{\varepsilon_{rn}}{\varepsilon_r}\Delta\varepsilon_{r_m}\boldsymbol{E}_n^*(\boldsymbol{r}_T)\cdot\boldsymbol{E}_m(\boldsymbol{r}_T)\,d^2\boldsymbol{r}_T. \tag{40}$$

In the following, we will compare the modified CMT given by Eq. 38 to the existing CMTs for several different cases so the difference between these formulas can be appreciated. In addition, with the aid of numerical simulations, we can determine which formula provides the most accurate results.

3.2.3 Symmetric waveguide coupler

The first structure we study is two coupled waveguides with identical sizes. Due to the structural symmetry, by proper field normalization in Eqs. 31 and 32, we could

have $X_{12} = X_{21}$, $k_{11} = k_{22}$ and $k_{12} = k_{21}$, all of which are real numbers. There are two possible solutions for Eq. 30, which correspond to the propagation constants of the two system modes and could be solved as

$$\beta_{\pm} = \frac{\beta_1 + \beta_2}{2} + \frac{k_{11} - X_{12}k_{12}}{1 - X_{12}^2} \pm \frac{k_{12} - X_{12}k_{11}}{1 - X_{12}^2}, \tag{41}$$

where the subscript"+" ("-") denotes the mode with a higher (lower) propagation constant. It is well known that the modes of couplers which are made of identical waveguides are composed of symmetric and anti-symmetric combinations of the individual waveguide modes, and the power coupling coefficient of such couplers is given by the difference of the propagation constants of the two system modes [27]. In other words, if we launch unit power at the input of the waveguide 1, the power in the waveguide 2 is a function of the coupling length L as

$$P_{21}(L) = \sin^2(\kappa L), \tag{42}$$

where the power coupling coefficient κ is given by

$$\kappa_X = \frac{\beta_+ - \beta_-}{2} = \frac{k_{12} - X_{12}k_{11}}{1 - X_{12}^2}. \tag{43}$$

The subscript "X" in Eq. 43 suggests that this coupling coefficient is derived based on the CMT that embraces the X_{12} factor. On the other hand, the coupling coefficient derived from the old CMT can be simply obtained by setting X_{12} to be zero in Eq. 43, resulting

$$\kappa_{\text{old}} = k_{12}, \tag{44}$$

where the meaning of the subscript"old" in Eq. 44 is self-explained.

Following a similar fashion, our modified CMT predicts the propagation constants of the coupler are

$$\beta_{\text{mod},\pm} = \frac{\beta_1 + \beta_2}{2} + \frac{X'_{11}k'_{11} - X'_{12}k'_{12}}{X'^2_{11} - X'^2_{12}} \pm \frac{X'_{11}k'_{12} - X'_{12}k'_{11}}{X'^2_{11} - X'^2_{12}}, \tag{45}$$

where the subscript "mod" indicates the result is from the modified CMT. The power coupling coefficient is then given by

$$\kappa_{\text{mod}} = \frac{X'_{11}k'_{12} - X'_{12}k'_{11}}{X'^2_{11} - X'^2_{12}}, \tag{46}$$

with the parameters X'_{nm} and k'_{nm} given by Eqs. 39 and 40, respectively.

Now we have three different expressions for the power coupling coefficient of a symmetric parallel waveguide coupler, namely, κ_X (Eq. 43), κ_{old} (Eq. 44) and κ_{mod} (Eq. 46). As we have mentioned in the beginning of this section, for the high-index-contrast material systems, κ_{old} gives more accurate results compared to κ_X. From Eq. 43, we find that to the first order of X_{12},

$$\kappa_X \approx k_{12} - X_{12}k_{11} = \kappa_{\text{old}} - X_{12}k_{11}, \tag{47}$$

which shows that the difference between κ_X and κ_{old} is mainly determined by the product of X_{12} and k_{11}. In the high-index-contrast materials, k_{11} is usually comparable to k_{12}, and therefore κ_X and κ_{old} can be very different if $|X_{12}|$ is not negligible. Turning to κ_{mod}, we first have to inspect the coefficients of X'_{nm} (Eq. 39) and k'_{nm} (Eq. 40). A careful examination of Eq. 40 reveals that if $n \neq m$, $k'_{nm} = k_{nm}$, since where $\Delta\varepsilon_{rm} \neq 0$, $\varepsilon_{rn} = \varepsilon_r$ if $n \neq m$ (see Fig. 13(b)). As a result, we have $k'_{12} = k_{12}$. Similarly, when $n = m$, it is easy to verify that

$$k'_{11} = \frac{\varepsilon_{\text{clad}}}{\varepsilon_{\text{WG}}}k_{11}, \tag{48}$$

where $\varepsilon_{\text{clad}}$ and ε_{WG} refer to the dielectric constants of the cladding and waveguide core, respectively. In the high-index-contrast materials, ε_{WG} is much larger than $\varepsilon_{\text{clad}}$, and therefore k'_{11} is much smaller than k_{11}. Regarding X'_{11}, if we assume the majority of the field is confined inside the waveguide for which $\varepsilon_{r1} = \varepsilon_r$, we could approximate $X'_{11} \approx 1$ from Eq. 39. κ_{mod} shown by Eq. 46 can then be reduced to the first order of X'_{12} as

$$\kappa_{\text{mod}} \approx k_{12} - \frac{\varepsilon_{\text{clad}}}{\varepsilon_{\text{WG}}}X'_{12}k_{11}, \tag{49}$$

which indicates that κ_{mod} is close to κ_{old} (i.e., κ_{12}) when the refractive index contrast is high, since the second term in Eq. 49 can be neglected. Therefore, we have offered an explanation for the better performance of κ_{old} compared to κ_X for the TM polarization in high-index-contrast material systems.

Next, we proceed to perform a rigorous check on these three expressions with the aid of numerical simulations. Figure 14(a) shows one such example for a 2-dimensional (2-D) waveguide coupler, which is composed of waveguides with a width of 400 nm and various gaps. The refractive index of the waveguide core is 2.829, corresponding to the effective index of a 220-nm-thick silicon slab at the wavelength of 1550 nm (air cladding). We first simulate the waveguide coupler modes using an in-house numerical code implemented in the COMSOL environment. The exact power coupling coefficient is obtained from the difference between the propagation constants of the symmetric and the anti-symmetric modes. The coupling coefficients given by Eqs. 43, 44 and 46 are then computed, with the individual waveguide modes provided by the numerical mode solver too. We plot the relative error of each expression in Fig. 14(a), which is defined for any κ_{predict} as

$$\text{Relative error of } \kappa_{\text{predict}} = \frac{\kappa_{\text{predict}} - \kappa_{\text{exact}}}{\kappa_{\text{exact}}} \times 100\%. \tag{50}$$

As can be seen in Fig. 14(a), κ_{old} and κ_{mod} are almost overlapping with each other, while κ_X has much larger errors compared to κ_{old} and κ_{mod}, especially when the gap between the two waveguides is small and the power overlapping factor X_{12} is large. Though we have derived the expression of κ_{mod} for the TM-polarized light in 2-D slabs, we expect that it will also apply to the TE-like mode of 3-D waveguides [3]. To verify that, we examine 3-D silicon waveguides in Fig. 14(b) through a similar numerical code in COMSOL [41]. The waveguide structure is the same as in Fig. 14(a), with the height of 220 nm and width of 400 nm in the air cladding. As shown in Fig. 14(b),

[3]When the TE-like polarizaiton of 3-D waveguides is reduced to 2-D, it is TM-polarized.

κ_{old} and κ_{mod} are very close while κ_X has larger errors.

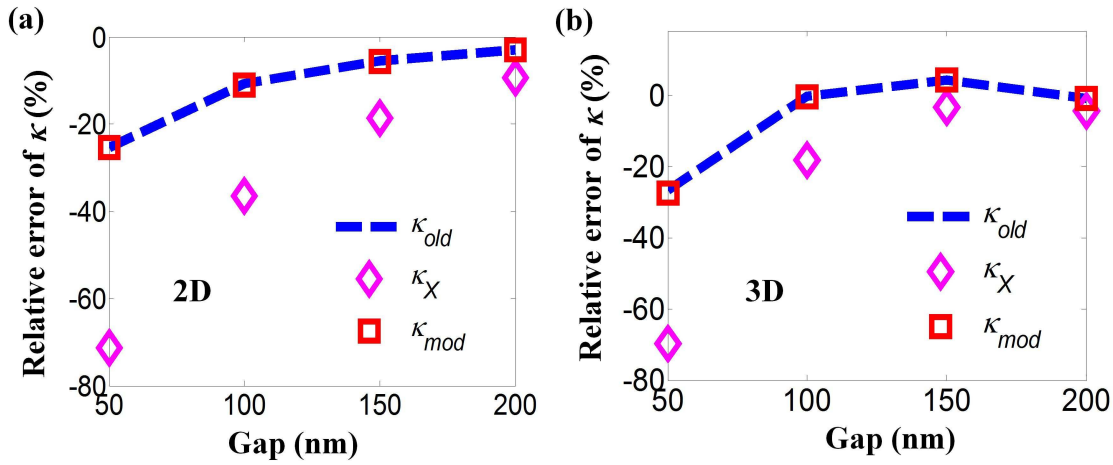

Figure 14: The relative error of κ of (a) a 2-D symmetric coupler for the TM-polarized light, and (b) a 3-D symmetric coupler for the TE-polarized light.

From the above numerical example, it seems that our modified CMT gives similar results as the old CMT without the X_{nm} factors. However, according to Eq. 43, κ is only related to the difference between β_+ and β_-, and hence the examination of κ does not provide a full picture. In fact, there is another important parameter α that should be investigated, which is defined as

$$\alpha \equiv \frac{\beta_+ + \beta_- - \beta_1 - \beta_2}{2}, \tag{51}$$

where $\beta_1 = \beta_2$ is the propagation constant of the individual waveguide. The physical significance of α can be understood by considering the response of the symmetric waveguide coupler again. Assuming a unit input at the waveguide 1, after a coupling length L the field in the waveguide 2 will be (excluding the phase factor resulting from the free propagation)

$$t_{21}(L) = ie^{i\alpha L}\sin(\kappa L), \tag{52}$$

indicating that α is connected to the phase response of t_{21} while κ is related to its amplitude response. The nonzero α is responsible for the so-called"coupling-induced

phase shift", which is of critical importance to coupled-resonator structures and will be discussed in detail in the next section [42, 43]. In Fig. 15, we plot the corresponding α for the waveguide structures studied in Fig. 14. It is easy to observe that α_{old} is similar to α_X and both are several times larger than α_{exact}. In contrast, α_{mod} is a few times smaller than α_{exact}. Therefore, the relative error of α_{old} and α_X is significantly larger than that of α_{mod}.

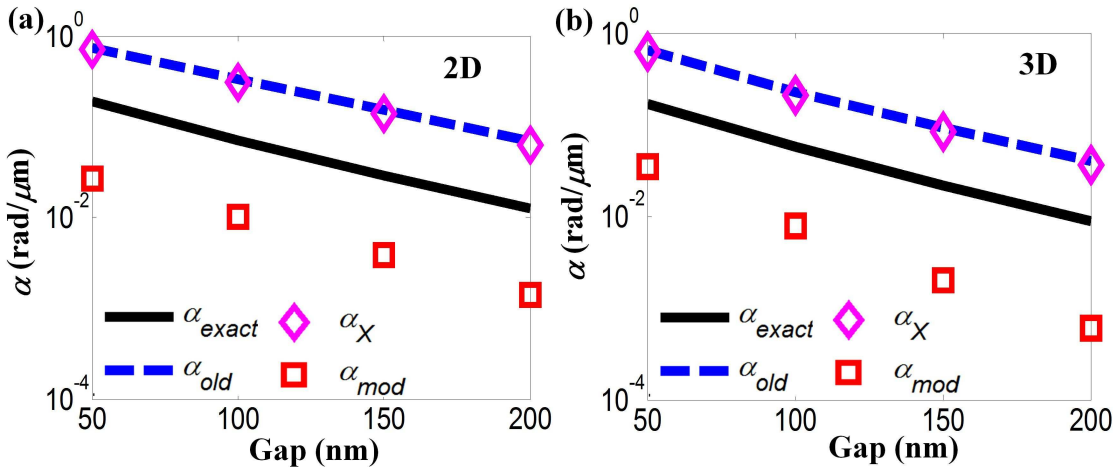

Figure 15: α of (a) a 2-D symmetric coupler for the TM-polarized light, and (b) a 3-D symmetric coupler for the TE-polarized light.

From a practical standpoint of view, however, the relative error may not a good parameter to evaluate the accuracy of these coupled mode formulations for α. For example, when α is small enough to be negligible, a large relative error is acceptable as long as its absolute error is still unimportant. Here, an appropriate measure of being small or large enough should be the power coupling coefficient κ. For this reason, we define the normalized error of α_{predict}

$$\text{Normalized error of } \alpha_{\text{predict}} = \frac{\alpha_{\text{predict}} - \alpha_{\text{exact}}}{\kappa_{\text{exact}}} \times 100\%, \tag{53}$$

and compute its value for all the cases shown in Fig. 15 in Fig. 16. For both the 2-D and 3-D cases, the normalized error of α_{old} and α_X is a few times larger than that

of α_{mod}, as expected. In addition, the normalized error of α_{mod} is reasonably small, suggesting the modified CMT offers satisfactory results.

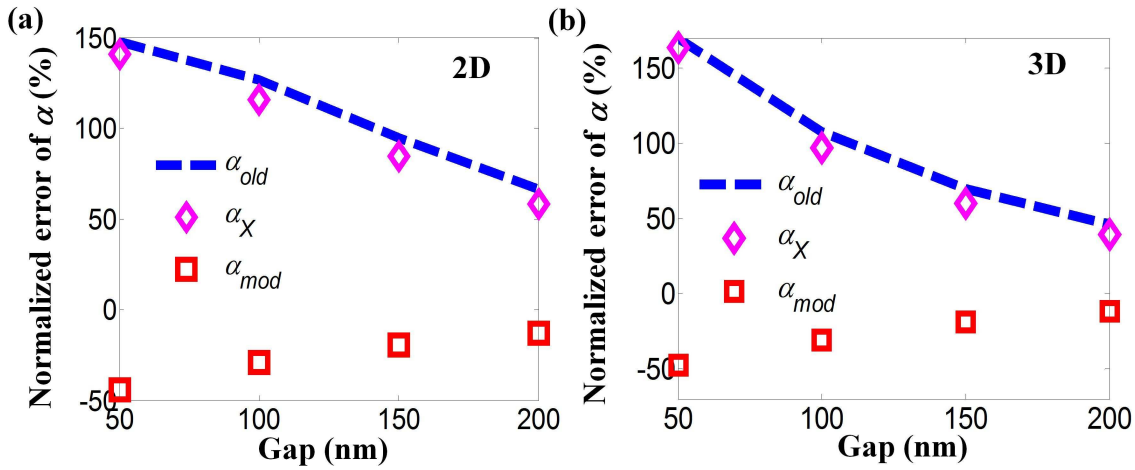

Figure 16: Normalized error of α of (a) a 2-D symmetric coupler for the TM-polarized light, and (b) a 3-D symmetric coupler for the TE-polarized light.

3.2.4 Asymmetric waveguide coupler

In this subsection, we will extend our analysis to general waveguide couplers, for which the two waveguides do not necessarily share identical properties. Without loss of generality, the coupled mode equation can be written as

$$\frac{d}{dz}\begin{bmatrix}\bar{a}_1(z)\\ \bar{a}_2(z)\end{bmatrix} = i\begin{bmatrix}\gamma_{11} & \gamma_{12}\\ \gamma_{21} & \gamma_{22}\end{bmatrix}\begin{bmatrix}\bar{a}_1(z)\\ \bar{a}_2(z)\end{bmatrix} = i\mathbf{\Gamma}\begin{bmatrix}\bar{a}_1(z)\\ \bar{a}_2(z)\end{bmatrix}, \tag{54}$$

where the matrix $\mathbf{\Gamma}$ assumes different expressions in different coupled mode formulations. For example, for the old CMT without considering the X_{nm} factors, $\gamma_{nm} = k_{nm}$ so

$$\mathbf{\Gamma}_{\text{old}} = \begin{bmatrix}k_{11} & k_{12}\\ k_{21} & k_{22}\end{bmatrix}. \tag{55}$$

For the CMT with the X_{nm} factors, from Eq. 30 we have

$$\mathbf{\Gamma}_X = \begin{bmatrix}1 & X_{12}\\ X_{21} & 1\end{bmatrix}^{-1}\begin{bmatrix}k_{11} & k_{12}\\ k_{21} & k_{22}\end{bmatrix}, \tag{56}$$

while for the modified CMT,

$$\mathbf{\Gamma}_{\text{mod}} = \begin{bmatrix} X'_{11} & X'_{12} \\ X'_{21} & X'_{22} \end{bmatrix}^{-1} \begin{bmatrix} k'_{11} & k'_{12} \\ k'_{21} & k'_{22} \end{bmatrix}. \tag{57}$$

Solving the eigenvalue problem of Eq. 54, we obtain the propagation constants of the system modes as

$$\beta_{\pm} = \frac{\gamma_{11} + \gamma_{22} \pm \sqrt{(\gamma_{11} - \gamma_{22})^2 + 4\gamma_{12}\gamma_{21}}}{2}. \tag{58}$$

Furthermore, assuming unit power launched at the waveguide 1, after a propagation length L the transmitted power at the waveguide 2 will be

$$P_{21}(L) = \frac{4|\gamma_{21}|^2}{|(\gamma_{11} - \gamma_{22})^2 + 4\gamma_{12}\gamma_{21}|} \sin^2\left(\sqrt{(\frac{\gamma_{11} - \gamma_{22}}{2})^2 + \gamma_{12}\gamma_{21}}\ L\right). \tag{59}$$

Likewise, if we send unit power at the waveguide 2, after a propagation length L the transmitted power at the waveguide 1 will be

$$P_{12}(L) = \frac{4|\gamma_{12}|^2}{|(\gamma_{11} - \gamma_{22})^2 + 4\gamma_{12}\gamma_{21}|} \sin^2\left(\sqrt{(\frac{\gamma_{11} - \gamma_{22}}{2})^2 + \gamma_{12}\gamma_{21}}\ L\right). \tag{60}$$

Equations 59 and 60 reveal several important points: (1) since $|\gamma_{12}| \neq |\gamma_{21}|$ in general case, $P_{12} \neq P_{21}$; (2) $|\gamma_{11} - \gamma_{22}|$ is generally nonzero and could be comparable or much larger than $|\gamma_{12}|(|\gamma_{21}|)$, especially when $\beta_1 \neq \beta_2$. If $|\gamma_{11} - \gamma_{22}|^2 \gg |\gamma_{12}\gamma_{21}|$, then for P_{21},

$$P_{21} \leq \frac{4|\gamma_{21}|^2}{|\gamma_{11} - \gamma_{22}|^2} \ll 1, \tag{61}$$

which indicates that the maximum power transferred from the waveguide 1 to 2 is significantly less than 100%, no matter how large the coupling length L is. The same conclusion applies to P_{12}. In consequence, to achieve an efficient coupling between the two waveguides, it is important to ensure $|\gamma_{11} - \gamma_{22}| \ll |\gamma_{12}|$ (or $|\gamma_{21}|$), which is the so-called phase matching condition. For the symmetric waveguide case, phase matching is automatically satisfied, which is one of the reasons that the symmetric waveguide couplers are more popular than the asymmetric ones in photonic circuits.

However, under certain circumstances, asymmetric waveguide couplers become necessary. For example, the two waveguides may be fabricated in two distinct material systems so the refractive indices of their waveguide cores are very different. The conventional designing rule for the phase matching is to engineer each individual waveguide geometry so that $\beta_1 \approx \beta_2$. Here, we will examine this intuitive result based on the coupled mode formulations. For example, in the old CMT without the X_{nm} factors,

$$\begin{aligned}\gamma_{11} - \gamma_{22} = k_{11} - k_{22} = \Delta\beta + \frac{\omega\varepsilon_0}{4}\int \Delta\varepsilon_{r_1}\boldsymbol{E}_1^*(\boldsymbol{r}_T)\cdot\boldsymbol{E}_1(\boldsymbol{r}_T)\,d^2\boldsymbol{r}_T \\ - \frac{\omega\varepsilon_0}{4}\int \Delta\varepsilon_{r_2}\boldsymbol{E}_2^*(\boldsymbol{r}_T)\cdot\boldsymbol{E}_2(\boldsymbol{r}_T)\,d^2\boldsymbol{r}_T,\end{aligned} \tag{62}$$

which suggests that the true phase matching condition, i.e., $\gamma_{11} = \gamma_{22}$, is not identical to $\Delta\beta = 0$, since the two integrals in Eq. 62 generally do not cancel with each other. As a result, even if we have designed the waveguides such that $\Delta\beta = 0$ (i.e., $\beta_1 = \beta_2$), the two waveguides are still phase mismatched. The same conclusion applies to the CTM with the X_{nm} factors and also our modified CMT, though γ_{11} and γ_{22} are given by significantly different expressions so the exact phase matching condition will vary (i.e., different values of $\Delta\beta$ are required to bring $\gamma_{11} = \gamma_{22}$ in these CMTs).

We examine the three different CMT formulations using the structure shown in Fig. 17. The two waveguides 1 and 2 have widths of 400 nm and 800 nm, respectively; their refractive indices have been chosen to be 2.5 and 2.02 (in the air cladding), respectively, so that their effective indices are the same at the wavelength of 1550 nm ($n_{\text{eff}} = 1.8174$). Then the two waveguides are brought closer so the power in one waveguide can be transferred to the other. To accurately measure the power at each waveguide output without ambiguity, we bend both waveguides away with perfectly matched layers (PMLs) implemented in the waveguide ends. Details of the numerical simulation will be given in the next section. According to Eq. 59, if we send unit power at the input of waveguide 1, the coupled power at the waveguide 2 can be

described as a function of the coupling length L as

$$P_{21}(L) = |T|^2 \sin^2(\kappa(L + L_0)), \tag{63}$$

where $|T|^2(\leq 1)$ and κ are functions of γ_{nm} as given by Eq. 59, and L_0 is the effective coupling length contributed by the bending part. We compute the parameters of $|T|^2$ and κ given by these CMTs, and compare the predicted power transfer (P_{21} here) to numerical simulations in Fig. 17(b). In addition, we have fitted the simulation results with Eq. 63, where $|T|^2$, κ and L_0 are extracted as 0.83, 0.185/μm and 1.4 μm, respectively. In other words, the maximum power transfer that can be achieved from waveguide 1 to 2 is 83%, even though the intuitive phase matching condition $\beta_1 = \beta_2$ has been satisfied. Moreover, as shown in Fig. 17(b), the old CMT without the X_{nm} factors and the modified CMT agree with the simulation result well when P_{21} is small (i.e., L is small). However, the maximum transferable power predicted by the old CMT is 70.8% while the one from our modified CMT is 88.7%. As a result, when L is large, the modified CMT gives more accurate result compared to the old CMT. On the other hand, the CMT with the X_{nm} factors deviates significantly from the simulation result, for which the maximum transferrable power is predicted only to be 43.6%. We list all the parameters used in Fig. 17(b) in Table 1.

Table 1: Parameters used for P_{21} in Fig. 17

P_{21} (gap=100 nm)	Simulation Fit	CMT_{mod}	CMT_{old}	CMT_X
$\lvert T\rvert^2$	0.83	0.887	0.708	0.436
κ (1/μm)	0.185	0.169	0.199	0.170

3.2.5 Conclusions

Combining all the examples examined from Fig. 14 to Fig. 17, we conclude that the old CMT without the X_{nm} factors offers good predictions for the power coupling coefficient, but generally cannot be trusted for the phase response resulting from the self-perturbation effect (usually a few times overestimate). The CMT with the X_{nm}

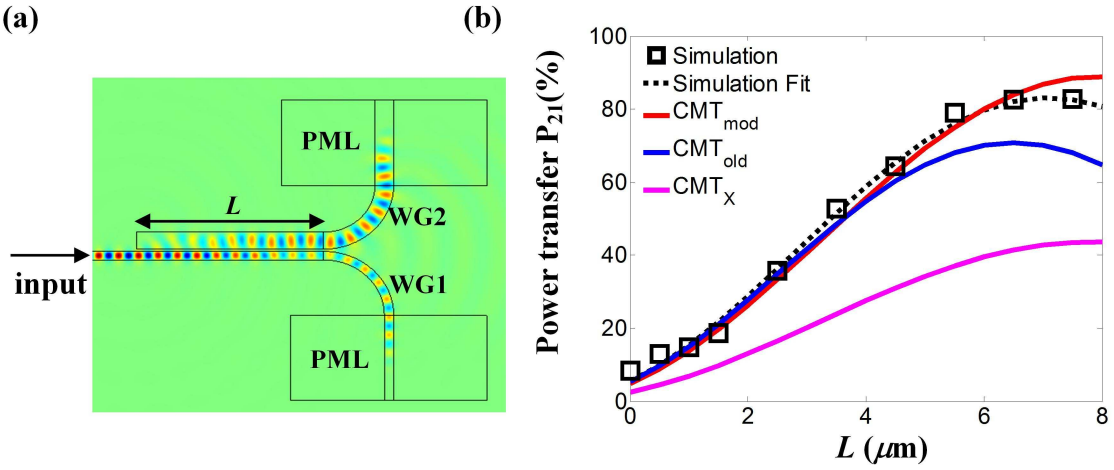

Figure 17: (a) Schematic of an asymmetric waveguide coupler with PMLs implemented in the waveguide ends. The widths of the two waveguides are 400 nm and 800 nm. (b) The numerical result and analytical predictions by various CMTs for P_{21} as a function of the coupling length L.

factors, surprisingly, is worse for the power coupling calculation (compared to the old CMT) and cannot predict an accurate phase response either (which is similar to the old CMT). Our modified CMT, overall, gives much better results on the power coupling and the phase responses. Especially, from the numerical example shown in Fig. 17, it is essential to use the modified CMT for asymmetric couplers to obtain the most accurate result.

3.3 Phase response between coupled photonic structures

In the last section, the power coupling between two closely spaced photonic structures is studied. Recently, phase responses associated with the power coupling process also attract a lot of research interest. It is shown by Refs. [42, 43] that the coupling can change the resonance frequency of each coupled resonator (the so-called "coupling-induced resonance frequency shift" (CIFS)), leading to distorted responses in coupled-resonator devices. In this section, we will first formally introduce the phase response, then we will describe the numerical method that is developed for its accurate simulation, and finally, the phase responses of a few typical structures are

studied using the developed numerical approach.

3.3.1 Phase response and its properties

Without loss of generality, the coupling region can be treated as a multi-port coupler. Figure 18 shows two such examples with different coupling geometries, where a_1 and a_2 stand for input signals, and b_1 and b_2 stand for output signals. We also assume that the reference planes for a_i and b_i $(i = 1, 2)$ are far from the coupling region, and therefore the signals can be normalized in a way that $|a_i|^2$ and $|b_i|^2$ $(i = 1, 2)$ represent powers at their respective ports. Assuming the coupling is lossless and reflection is also negligible, the output signals can be related to the input by a unitary 2×2 transfer matrix as [27]

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = U \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} u_{11}e^{i\phi_{11}} & u_{12}e^{i\phi_{12}} \\ u_{21}e^{i\phi_{21}} & u_{22}e^{i\phi_{22}} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \tag{64}$$

where u_{ij} and ϕ_{ij} $(i, j = 1, 2)$ are the amplitude and phase responses of each component of the transfer matrix U, respectively. As a convention, in Eq. 64 we have assumed that the free propagation phase term, namely, the phase contribution resulting from the wave propagation from the reference plane of a_1 (a_2) to that of b_1 (b_2) in the absence of coupling is excluded. In this sense, ϕ_{11} and ϕ_{22} are contributed by the coupling interaction and will asymptotically approach zero when the coupling is weakened to a negligible extent.

Before going into the detailed calculation of these phase terms, it is worthwhile to mention a few properties of the transfer matrix U imposed by some general considerations. For example, power conservation requires U to be unitary, which is mathematically equivalent to (using $U^\dagger = U^{-1}$)

$$U = \begin{bmatrix} u_{11}e^{i\phi_{11}} & u_{12}e^{i\phi_{12}} \\ u_{21}e^{i\phi_{21}} & u_{22}e^{i\phi_{22}} \end{bmatrix} = e^{i\theta_0} \begin{bmatrix} \sqrt{1-\kappa^2}e^{i\theta_1} & i\kappa e^{i\theta_2} \\ i\kappa e^{-i\theta_2} & \sqrt{1-\kappa^2}e^{-i\theta_1} \end{bmatrix}, \tag{65}$$

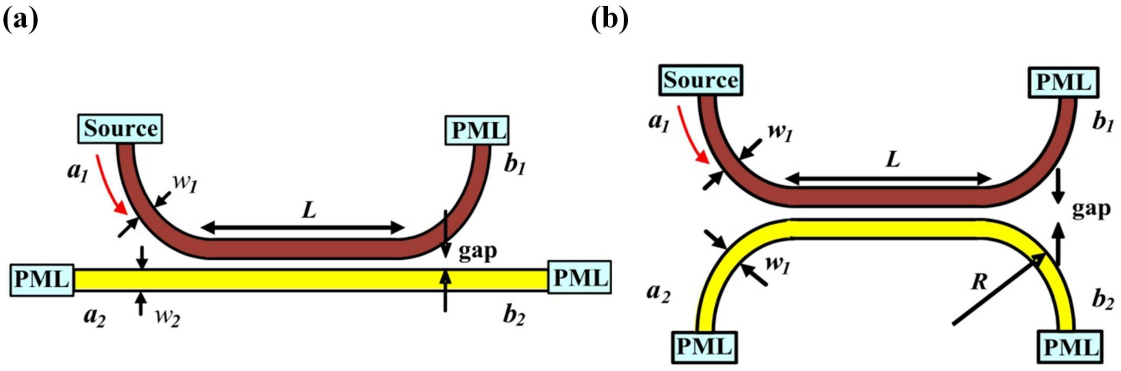

Figure 18: Schematic of the phase simulation for (a) couplers with symmetry about the vertical axis, and (b) couplers with symmetry about both the vertical and horizontal axes. Optical sources and PMLs are implemented based on the method described in the text.

where θ_0, θ_1 and θ_2 are phase parameters related to $\phi_{ij}(i, j = 1, 2)$ and κ is the power coupling coefficient between the two coupled structures (all these variables are real). From Eq. 65, it is easy to observe that

$$u_{21} = u_{12}, \tag{66}$$

$$\phi_{11} + \phi_{22} = \phi_{12} + \phi_{21} - \pi. \tag{67}$$

Note that from Eq. 66 it seems that that $P_{21} = P_{12}$ no matter what the coupling geometry is, contradicting with the result of Eqs. 59 and 60 which suggests $P_{21} \neq P_{12}$ for asymmetric waveguide couplers. This contradiction is due to different assumptions made for the input power configurations. To deduce $P_{21} = P_{12}$ from Eq. 66, we should have unit power at one input port and zero at the other. However, for the results of Eqs. 59 and 60 to be valid, we have assumed unit power for one waveguide mode and zero for the other. These two assumptions are not equivalent, since the former one is for the two input ports that are far from the coupling region while the latter one is for two waveguides that are closely spaced with a nonzero field overlap.

When the couplers have the symmetry about the middle vertical axis (see Fig. 18(a)),

reciprocity and symmetry considerations give [43]

$$\frac{\phi_{11}+\phi_{22}}{2} = \phi_{12} - \frac{\pi}{2} = \phi_{21} - \frac{\pi}{2}. \tag{68}$$

Moreover, if the coupler has the additional symmetry about the middle horizontal axis in the coupling region, as shown by Fig. 18(b), we have

$$\phi_{11} = \phi_{22} = \phi_{12} - \frac{\pi}{2} = \phi_{21} - \frac{\pi}{2}. \tag{69}$$

Note that for the coupler shown in Fig. 18(a), no such horizontal axis exists, and hence, in general $\phi_{11} \neq \phi_{22}$.

3.3.2 Numerical approach for phase simulation

A brief survey of literature reveals that the majority of numerical studies on photonic structures focus on the amplitude responses of the coupling effect, such as the power exchange between coupled structures or the radiation loss caused by the bend [27]. To the best of our knowledge, before Ref. [43], the phase responses resulting from the coupling are usually ignored since no important effects are known. Extending the previous numerical approaches to the study of phase responses, however, is not trivial. To get a reliable phase responses, numerical simulations based on either the finite-ifference time domian (FDTD) method or finite element method (FEM) require a much higher resolution compared to those for amplitude responses. To understand this, we can write the simulated field as

$$\boldsymbol{E}_{\text{simulation}} = \boldsymbol{E}_{\text{exact}}(1 + a + ib) \quad (a \text{ and } b \text{ are both real}), \tag{70}$$

where a and b represent the errors in the numerical simulation, which are always nonzero due to reflection caused by imperfect PML boundaries, large grid size, or insufficient simulation time. The amplitude responses are tolerant to these errors. For example, increasing $|b|$ from 0 to 0.1 only change the amplitude $|\boldsymbol{E}_{\text{simulation}}|^2$ by less than 1%. Similarly, increasing $|a|$ from 0 to 0.1 changes $|\boldsymbol{E}_{\text{simulation}}|^2$ by less than

20%. On the other hand, the phase responses are very sensitive to these errors. For example, the change of $|b|$ from 0 to 0.1 will add or subtract 0.1 rad from the exact phase value, which itself might be a small number. As a result, the relative error of the phase responses can be quite big.

In this work, Maxwell's equations are implemented and solved as partial differential equations in the commercial software COMSOL. Perfectly matched layers (PMLs) are implemented based on the stretched coordinate method [44, 45, 46]. For example, for the TE-polarized light in 2-D Cartesian coordinate system, the following equation is implemented

$$\frac{\partial}{\partial x}\left(\frac{S_y}{S_x}\frac{\partial E_z}{\partial x}\right)+\frac{\partial}{\partial y}\left(\frac{S_x}{S_y}\frac{\partial E_z}{\partial y}\right)+k_0^2 n^2 S_x S_y E_z=0, \tag{71}$$

where S_x and S_y are the complex coordinate stretching factors for the x and y coordinates [4], respectively . The key point of the stretched coordinate method is that S_x and S_y are equal to unity in normal regions so Eq. 71 is identical to Maxwell's equation. In contrast, for the PML region, S_x and S_y have a gradually increasing imaginary part so the light is attenuated without getting reflected. In our code, S_x is chosen to be

$$S_x=1-2i(\frac{\rho}{d_x})^2, \tag{72}$$

where ρ is the distance from the beginning of the PML and d_x is the thickness of the PML layer for the x coordinate. Similar expressions apply to S_y for the y coordinate. On the other hand, for the TM-polarized light in 2-D Cartesian coordinate system, the following equation can be implemented

$$\frac{\partial}{\partial x}\left(\frac{S_y}{S_x}\frac{1}{n^2}\frac{\partial H_z}{\partial x}\right)+\frac{\partial}{\partial y}\left(\frac{S_x}{S_y}\frac{1}{n^2}\frac{\partial H_z}{\partial y}\right)+k_0^2 S_x S_y H_z=0, \tag{73}$$

where the $1/n^2$ factor has to stay inside the differential bracket since it is a function of x and y. The last component we need for a complete wave propagation simulation

[4]From hereon the coordinate system is different from what's shown in Fig. 13(a): x, y, and z in Fig. 19 correspond to z, x, y in Fig. 13.

is the optical source, which is achieved in COMSOL by setting proper boundary conditions. As illustrated in Fig. 19(a), the waveguide mode that we want to launch is specified at the source line. In addition, the source is surrounded by PMLs to absorb any radiated wave, which can either arise from the source or reflection from other components.

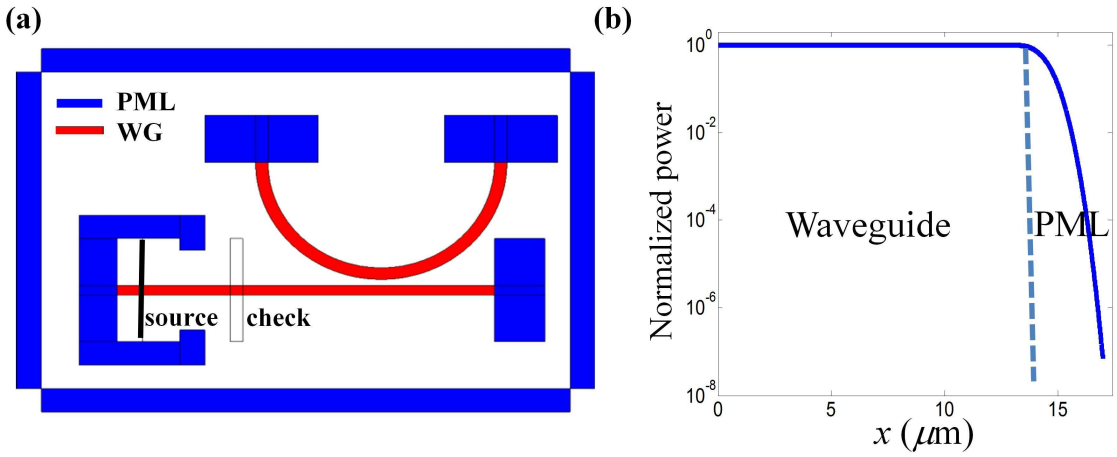

Figure 19: (a) Simulation structure to obtain the phase responses in COMSOL. The blue regions are PMLs and the red regions are waveguides. The rest of the area is cladding (air here). The optical source is implemented at the source line by specifying the appropriate boundary condition. (b) Optical power measured along the waveguide till the end of the PML in the case of a simple waveguide.

For simulations based on FEM, mesh size has to be small enough to avoid any artificial effects. In our case, cubic meshes with a grid size less than 50 nm are employed. To further minimize the impact of the mesh, we focus on the net effect of the coupling by comparing two situations when the neighboring waveguide is absent and present, which can be achieved by changing the refractive index of the neighboring waveguide while keeping the same mesh. Before performing any real simulations, it is essential to verify that a simple waveguide mode can be launched and absorbed by the PML with negligible reflection. Figure 19(b) shows the normalized power along a waveguide till the end of the PML, and good PML performance is observed.

3.3.3 Simulation results and discussions

We start with the simple coupler shown in Fig. 18(b). Similar to the numerical simulations carried out in the last section, we use the 2-D version of the SOI material system which has an effective index of 2.829 at the wavelength of 1578 nm [42]. The polarization is chosen to be TM [5]. For the simulation, the waveguide width w_1 is chosen to be 400 nm, the radius of the bend R is 6 μm, and the gap is fixed to be 100 nm while the coupling length L is varied. Figure 20(a) shows the simulated ϕ_{11} and ϕ_{12} as a function of the coupling length L, both of which exhibit a linear response with the same slope. Such responses can be easily understood from Eqs. 52 and 69, from which we conclude

$$\phi_{11} = \phi_{12} - \frac{\pi}{2} = \alpha(L + L_0), \tag{74}$$

where α is defined in Eq. 51 and L_0 is the effective length of the bending part. If we compute α based on the definition, its value agrees with the one extracted from Fig. 20(a) well (0.077 rad/μm) [42]. We want to point out that α is not always positive, as examples with a negative α have been examined in Ref. [43].

Next, we proceed to examine the coupler shown in Fig. 18(a), and the results are plotted in Fig. 20(b). Because of the lack of symmetry about the horizontal axis, $\phi_{11} \neq \phi_{22}$. In addition, we observe that as L increases, ϕ_{11} first increases then decreases to a negative value. A negative ϕ_{11} means the net contribution of the coupling to the waveguide mode 1 is to reduce its effective propagation constant (remember that phase is a product of the propagation constant and the propagation length), which is counterintuitive since the conventional wisdom considers the neighboring waveguide as a positive index perturbation and hence the propagation constant of the perturbed mode will always increase. On the other hand, the relation of Eq. 68 is still satisfied,

[5] Most of the results here are taken from Ref. [42]. It should be noted that the polarization there (TE) is defined for a 3-D waveguide,i.e., the predominant electric field is in the silicon layer. When reduced to a 2-D waveguide, the same mode corresponds to the TM-polarized light based on the definition.

and the slope of ϕ_{12}, ϕ_{21} and $(\phi_{11} + \phi_{22})/2$ with L is the same as the one shown in Fig. 20(a).

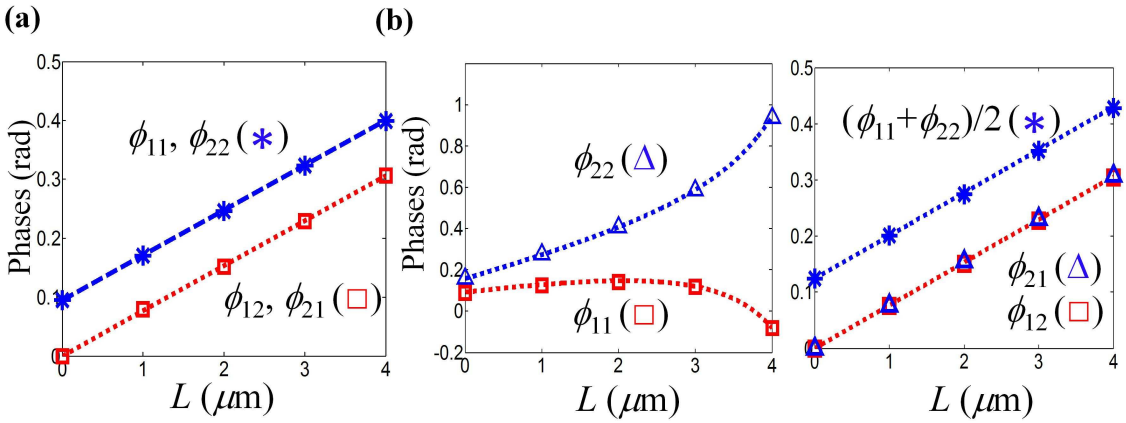

Figure 20: Simulated phase responses for (a) couplers shown in Fig. 18(a), and (b) couplers shown in Fig. 18(b). The detailed geometric parameters are provided in the text.

The major difference between the couplers shown in Figs. 18(a) and (b) are their the input and output bend parts, which are asymmetric about the horizontal axis in the former and symmetric in the latter. This reveals the importance of the input and output bends to the whole coupling structure. A semi-analytical model has been developed in Ref. [42], which explains the behavior of each coupler shown in Fig. 20. We will not go into the details here, but only point out the implications of this study to the practical applications of coupling structures. First of all, the coupler shown in Fig. 18(a) is equivalent to an asymmetric waveguide coupler even the two waveguides have the same width, and consequently, the maximum power transfer is always less than unity. Second, for structures that are sensitive to such coupling-induced phase responses, symmetric couplers as shown in Fig. 18(b) are more preferred because their phase responses are a simple function of the coupling length as given by Eq. 74. This is especially advantageous for 3-D devices, for which the direct simulation of wave propagation is difficult. However, we can still use Eq. 74

to estimate the phase responses by computing α for the 3-D case and substituting an approximate effective length for the bend parts. In contrast, the phase responses for the asymmetric couplers as shown in Fig. 18(a) are comprised of two terms. The first term is contributed by the linear response given by Eq. 74, which is shared by both ϕ_{11} and ϕ_{22}. The second term is contributed by the input and output bends and is nonlinear with the coupling length L. What's more, its sign is different in ϕ_{11} and ϕ_{22}, so both ϕ_{11} and ϕ_{22} are nonlinear with L but their sum is still linear, agreeing with the simulation results shown in Fig. 20(b). Because of the presence of this nonlinear term, calculation of the phase responses for 3-D asymmetric couplers is challenging, and therefore we recommend symmetric couplers for devices such as coupled-resonator filters so the coupling-induced phase shift can be easily estimated and corrected.

3.4 A unified approach to mode splitting and scattering loss

The basics of the mode splitting and scattering loss are reviewed in Chapter II, and we conclude there that the existing approaches are only applicable to some specific situations. Here, we will develop a unified approach to the mode splitting and scattering loss, which applies to an arbitrary number of scatterers. In addition, we will reveal the conditions under which our model can be reduced to the results of the two existing approaches discussed in chapter II in their respective regimes. Numerical evidences and experimental results are provided to support the validity of the developed model.

3.4.1 Model derivation

Our approach considers the interactions among the CW and CCW modes and the free space continuum, with the coupling provided by scatterers. This is similar to the independent-scatterer approach as in Refs. [29, 34], but with a key modification that leads to distinct results.

For clarity, we use the microdisk resonator as an example for the derivation, while the result is generally applicable to any microresonator with a two-fold degeneracy in its resonance spectrum. For an isolated microdisk resonator (i.e., no external coupling), we can intuitively write down the following equation set

$$\frac{da_m}{dt} = -\left(i\omega_c + \frac{\kappa_0}{2}\right) a_m + i\sum_{n=1}^{N}\Big(\sum_{m'=\text{cw,ccw}} g_{n,m,m'}a_{m'} + \sum_j g_{n,m,j}b_j\Big), \tag{75}$$

$$\frac{db_j}{dt} = -i\omega_j b_j + i\sum_{n=1}^{N}\sum_{m=\text{cw,ccw}} g_{n,j,m}a_m, \tag{76}$$

where a_m and b_j are the normalized energy amplitudes of the m (m = cw or ccw) WGM and the jth free space mode, respectively (ω_c and ω_j are their corresponding original resonance frequencies); κ_0 is the intrinsic cavity loss without including the scattering loss; $g_{n,m,m'}$ is a parameter describing the scattering of the m WGM to the same ($m = m'$) or the counterpropagating ($m \neq m'$) WGM induced by the nth scatterer; $g_{n,j,m}$ is a similar parameter characterizing the nth-scatterer-induced scattering of the m WGM to the jth free space mode and $g_{n,m,j}$ is defined vice versa. For now we have used a discrete set of eigenmodes $[b_j]$ normalized in a finite but large enough volume to represent the free space continuum, and this restriction will be removed later.

In this model, each scatterer is treated as a dipole. The electric field $\boldsymbol{E}$ excites the polarization of the nth scatterer as $\boldsymbol{P} = \varepsilon_0\alpha_n\boldsymbol{E}_n$, where α_n and $\boldsymbol{E}_n$ are the polarizability and the electric field at the location of the nth scatterer, respectively. The interaction of the polarization $\boldsymbol{P}$ with the electric field $\boldsymbol{E}$ is given by $-\boldsymbol{P}\cdot\boldsymbol{E}^*$ [47], with both the electric fields in $\boldsymbol{P}$ and $\boldsymbol{E}$ normalized to their corresponding mode energies. For example, the jth free space mode is expressed as

$$\boldsymbol{E}_j(\boldsymbol{r}) = \frac{1}{\sqrt{\varepsilon_0 V_c}} e^{i\boldsymbol{k}_j\cdot\boldsymbol{r}}\hat{\boldsymbol{n}}_j, \tag{77}$$

where V_c is the normalization volume of the free space modes; $\boldsymbol{k}_j$ is the wave vector of the jth mode; and $\hat{\boldsymbol{n}}_j$ is the unit polarization vector. Similarly, the energy-normalized

electric field of the m WGM has the following form

$$\boldsymbol{E}_m(\boldsymbol{r}) = \frac{f(\boldsymbol{r})}{\sqrt{\int \varepsilon(\boldsymbol{r}) \left|f(\boldsymbol{r})\right|^2 \, d^3\boldsymbol{r}}} e^{ik_m x} \hat{\boldsymbol{n}}_m, \tag{78}$$

where we have explicitly written out the phase term $\exp(ik_m x)$ with k_m being the wavenumber of the m WGM along the mode circulating direction and x being the projection of $\boldsymbol{r}$ on that; $f(\boldsymbol{r})$ accounts for the amplitude as well as the phase variation other than $\exp(ik_m x)$; $\varepsilon(\boldsymbol{r})$ is the dielectric constant; and $\hat{\boldsymbol{n}}_m$ is the unit vector describing the polarization of the m WGM. To simplify the above expression, we can define a parameter V_m as

$$V_m \equiv \frac{\int \varepsilon(\boldsymbol{r}) \left|f(\boldsymbol{r})\right|^2 \, d^3\boldsymbol{r}}{\varepsilon_0}, \tag{79}$$

and $\boldsymbol{E}_m(\boldsymbol{r})$ can be alternatively expressed as

$$\boldsymbol{E}_m(\boldsymbol{r}) = \frac{f(\boldsymbol{r})}{\sqrt{\varepsilon_0 V_m}} e^{ik_m x} \hat{\boldsymbol{n}}_m. \tag{80}$$

Note that V_m defined in Eq. 79 generally does not bear the unit of volume. However, we notice that $f(\boldsymbol{r})$ is scalable in Eq. 78. If we normalize the electric field of the WGM to that of a reference point, for instance, where the amplitude of the electric field is the maximum, $f(\boldsymbol{r})$ can be interpreted as the relative field strength and V_m defined above has the unit of volume.

We now proceed to calculate the coupling coefficients $g_{n,m,m'}$, $g_{n,m,j}$, and $g_{n,j,m}$ based on their definitions in Eqs. 75 and 76. Starting with Maxwell's equations, we have [35]

$$\nabla \times (\nabla \times \boldsymbol{E}(\boldsymbol{r},t)) + \mu\varepsilon(\boldsymbol{r}) \frac{\partial^2 \boldsymbol{E}(\boldsymbol{r},t)}{\partial t^2} = -\mu \frac{\partial^2 \boldsymbol{P}(\boldsymbol{r},t)}{\partial t^2}, \tag{81}$$

where μ is the permeability of free space. For the m WGM,

$$\boldsymbol{E}(\boldsymbol{r},t) = a_m(t)\boldsymbol{E}_m(\boldsymbol{r}) = e^{-i\omega_c t}(a_m(t)e^{i\omega_c t})\boldsymbol{E}_m(\boldsymbol{r}), \tag{82}$$

where we have separated the fast oscillating term $\exp(-i\omega_c t)$ with the slowly varying term $a_m(t)\exp(i\omega_c t)$. Treating $\boldsymbol{P}(\boldsymbol{r},t)$ as a first-order perturbation, Eq. 81 can be

approximated as [35, 48]

$$2\frac{d}{dt}(a_m(t)e^{i\omega_c t}) \approx i\omega_c e^{i\omega_c t}\int \boldsymbol{P}(\boldsymbol{r},t)\cdot\boldsymbol{E}_m^*(\boldsymbol{r})\,d^3\boldsymbol{r}. \tag{83}$$

$\boldsymbol{P}(\boldsymbol{r},t)$ consists of contributions from each scatterer as

$$\boldsymbol{P}(\boldsymbol{r},t) = \varepsilon_0\sum_{n=1}^{N}\alpha_n\Big(\sum_m a_m(t)\boldsymbol{E}_m(\boldsymbol{r}) + \sum_j b_j(t)\boldsymbol{E}_j(\boldsymbol{r})\Big)\delta(\boldsymbol{r}-\boldsymbol{r}_n), \tag{84}$$

where $\boldsymbol{r}_n$ stands for the position of the nth scatterer. Substituting the detailed expression of $\boldsymbol{P}(\boldsymbol{r},t)$ into Eq. 84 and comparing it to Eq. 75, we arrive at

$$g_{n,m,m'} = \frac{\alpha_n\omega_c\left|f(\boldsymbol{r}_n)\right|^2}{2V_m}e^{i(k_{m'}-k_m)x_n}, \tag{85}$$

$$g_{n,m,j} = \frac{\alpha_n\omega_c f^*(\boldsymbol{r}_n)}{2\sqrt{V_mV_c}}e^{i(\boldsymbol{k}_j\cdot\boldsymbol{r}_n-k_mx_n)}\Big(\hat{\boldsymbol{n}}_j\cdot\hat{\boldsymbol{n}}_m(\boldsymbol{r}_n)\Big), \tag{86}$$

where x_n is the projection of the nth-scatterer's position $\boldsymbol{r}_n$ along the WGM circulating direction. In Eq. 86, the dependence of the polarization of the m WGM $\hat{\boldsymbol{n}}_m$ on the position of the nth scatterer $\boldsymbol{r}_n$ has been explicitly expressed, given that $\hat{\boldsymbol{n}}_m$ is not necessarily a constant vector (for example, the dominant electric field for the TM-polarized WGM is E_ϕ, which is in the direction of $\hat{\boldsymbol{\phi}}$). In deriving $g_{n,m,j}$, we have used the fact that only those free space modes that can resonate with the WGMs (i.e., $\omega_j \approx \omega_c$) need to be considered. Taking a similar procedure for Eq. 76 leads us to

$$g_{n,j,m} = \frac{\alpha_n\omega_c f(\boldsymbol{r}_n)}{2\sqrt{V_mV_c}}e^{i(-\boldsymbol{k}_j\cdot\boldsymbol{r}_n+k_mx_n)}\Big(\hat{\boldsymbol{n}}_j\cdot\hat{\boldsymbol{n}}_m(\boldsymbol{r}_n)\Big). \tag{87}$$

Note that $g_{n,m,j}$ and $g_{n,j,m}$ obtained here are different from those obtained in Refs. [29, 34], where the derivation is based on the interaction among quantized fields and the $\exp(i\boldsymbol{k}_j\cdot\boldsymbol{r}_n)$ ($\exp(-i\boldsymbol{k}_j\cdot\boldsymbol{r}_n)$) factor is missing in $g_{n,m,j}$ ($g_{n,j,m}$).

With the knowledge of the coupling coefficients $g_{n,m,m'}$, $g_{n,m,j}$, and $g_{n,j,m}$, we are ready to solve Eqs. 75 and 76. Instead of studying the fast oscillating terms $a_m(t)$ and $b_j(t)$, it is more convenient to work with the slowly changing variables $\bar{a}_m(t) \equiv a_m(t)\exp(i\omega_c t)$ and $\bar{b}_j(t) \equiv b_j(t)\exp(i\omega_j t)$. We first solve for Eq. 76 for

$\bar{b}_j(t)$ which is then substituted back to Eq. 75 to obtain $\bar{a}_m(t)$. In addition, we have used the standard procedure to change the summation over the mode number j by an integral over the wave vector $\boldsymbol{k}$ space, which gives us [49]

$$\frac{d\bar{a}_m(t)}{dt} = -\frac{\kappa_0}{2}\bar{a}_m(t) + \sum_{m'=\text{cw,ccw}} (iG_{m,m'} - \frac{\Gamma_{m,m'}}{2})\bar{a}_{m'}(t), \tag{88}$$

with

$$G_{m,m'} \equiv \frac{\alpha_n \omega_c \left|f(\boldsymbol{r}_n)\right|^2}{2V_m} e^{i(k_{m'}-k_m)x_n}, \tag{89}$$

$$\begin{aligned}\Gamma_{m,m'} \equiv & \sum_{n=1}^{N}\sum_{n'=1}^{N} \frac{\alpha_n \alpha_{n'} \omega_c^4 f^*(\boldsymbol{r}_n) f(\boldsymbol{r}_{n'})}{(4\pi)^2 V_m c^3} e^{i(-k_m x_n + k_{m'} x_{n'})} \times \\ & \iint \left[(1-\hat{\boldsymbol{k}}\hat{\boldsymbol{k}})\cdot\hat{\boldsymbol{n}}_m(\boldsymbol{r}_n)\right]\cdot\left[(1-\hat{\boldsymbol{k}}\hat{\boldsymbol{k}})\cdot\hat{\boldsymbol{n}}_{m'}(\boldsymbol{r}_{n'})\right] \times e^{ik_0\hat{\boldsymbol{k}}\cdot(\boldsymbol{r}_n-\boldsymbol{r}_{n'})}\sin\theta\, d\theta d\phi,\end{aligned} \tag{90}$$

where $k_0 = \omega_c/c$ is the wavenumber of light with angular frequency ω_c in free space; $(k,\, \theta,\, \phi)$ are the spherical coordinates of the wave vector $\boldsymbol{k}$, with $(\hat{\boldsymbol{k}},\, \hat{\boldsymbol{\theta}},\, \hat{\boldsymbol{\phi}})$ denoting the orthogonal unit vectors in the directions of increasing $(k,\, \theta,\, \phi)$, respectively. The integral in $\Gamma_{m,m'}$ involves an integration over the spherical surface in the wave vector space, and its value depends on the polarization of the WGMs as well as the relative positions of scatterers. Hence, it is a geometric factor.

3.4.2 Comparison with the independent-scatterer approach

The geometric integral in $\Gamma_{m,m'}$ given by Eq. 90 can be computed for any given $\boldsymbol{r}_n$ and $\boldsymbol{r}_{n'}$. For example, if the WGM is TE-polarized, $\hat{\boldsymbol{n}}_m = \hat{\boldsymbol{z}}$ (see Fig. 21(a)). In addition, we can choose the x axis to be in the direction of $\boldsymbol{r}_n - \boldsymbol{r}_{n'}$. The geometric integral in Eq. 90 can then be simplified as

$$\begin{aligned}&\int_0^{\pi} \sin^3\theta\, d\theta \int_{-\pi}^{\pi} e^{ik_0|\boldsymbol{r}_n-\boldsymbol{r}_{n'}|\sin\theta\cos\phi}\, d\phi \\ &= 2\pi \int_0^{\pi} \sin^3\theta J_0(k_0 d_{n,n'}\sin\theta)\, d\theta = \frac{8\pi}{3} p(k_0 d_{n,n'}),\end{aligned} \tag{91}$$

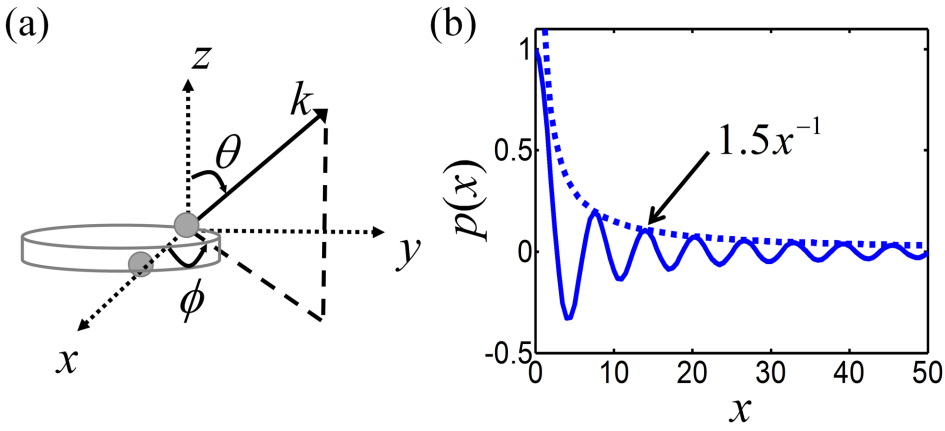

Figure 21: (a) Illustration of the adopted coordinate system for the calculation of the geometric integral in Eq. 91: the z axis is perpendicular to the microresonator, and the x axis is chosen to be along the relative position of the two scatterers under consideration. (b) The solid line is the numerical result of $p(x)$ defined in Eq. 92, and the dotted line is $3/2x$, which corresponds to the envelop of $p(x)$ when x is large.

where $d_{n,n'} \equiv |\boldsymbol{r}_n - \boldsymbol{r}_{n'}|$; $J_0(x)$ is the Bessel function of the first kind of order zero; and $p(x)$ is defined as

$$p(x) \equiv \frac{3}{4} \int_0^\pi \sin^3\theta J_0(x \sin\theta)\, d\theta. \tag{92}$$

In deriving Eq. 91, integral representations of the Bessel functions are used and mathematical details are left to Ref. [49]. From Eq. 91, we find that the geometric integral in $\Gamma_{m,m'}$ is only a function of the separation distance between scatterers; consequently,

$$\Gamma_{m,m'} = \sum_{n=1}^{N} \sum_{n'=1}^{N} \frac{\alpha_n \alpha_{n'} \omega_c^4 f^*(\boldsymbol{r}_n) f(\boldsymbol{r}_{n'})}{6\pi V_m c^3} e^{i(-k_m x_n + k_{m'} x_{n'})} p(k_0 d_{n,n'}). \tag{93}$$

In Fig. 21(b), numerical values of $p(x)$ are evaluated. For $x = 0$, $p(0) = 1$; when the argument x is large, the envelope of $p(x)$ decreases at a rate of x^{-1}, which can be proven by a rigorous calculation [49]. In the independent-scatterer approach discussed in chapter II, only the $n = n'$ terms (i.e., $d_{n,n'} = 0$ and $p(0) = 1$) are considered in the double summation over n and n' for $\Gamma_{m,m'}$, and the $n \neq n'$ terms are neglected, with the hope that the contribution from these terms is small if scatterers are well separated from each other [34]. From Fig. 21(b), we estimate that a reasonably large

separation should be on the order of $d_{n,n'}/\lambda_0 > 2.4$ ($|p(k_0 d_{n,n'})| < 0.1$), with λ_0 corresponding to the free space wavelength ($\lambda_0 = 2\pi c/\omega_c$). Otherwise, the omission of $n \neq n'$ terms can introduce significant errors and even lead to erroneous conclusions (such as $\gamma_+ \leq \gamma_-$). In the following, we will examine one numerical example for two scatterers attached to the surface of a microresonator, where we show it is essential to include $p(x)$ for a complete understanding of the simulation results.

3.4.3 Comparison with the intuitive physical approach

When the number of scatterers is large, the forms of $\Gamma_{m,m'}$ given by Eq. 90 (or Eq. 93) is not that convenient to work with. From Eq. 90, we notice that if we define

$$\boldsymbol{S}_m(\theta,\phi) \equiv \frac{\omega_c^2}{4\pi\sqrt{V_m c^3}} \sum_{n=1}^{N} \alpha_n f(\boldsymbol{r}_n) e^{ik_m x_n} \hat{\boldsymbol{n}}_m(\boldsymbol{r}_n) \cdot (1 - \hat{\boldsymbol{k}}\hat{\boldsymbol{k}}) e^{-ik_0 \hat{\boldsymbol{k}}\cdot\boldsymbol{r}_n}, \tag{94}$$

$\Gamma_{m,m'}$ can be rewritten as

$$\Gamma_{m,m'} = \iint \boldsymbol{S}_m^*(\theta,\phi) \cdot \boldsymbol{S}_{m'}(\theta,\phi) \sin\theta \, d\theta d\phi. \tag{95}$$

The expression of $\boldsymbol{S}_m(\theta,\phi)$ in Eq. 94 is invariant if we scale $f(\boldsymbol{r})$; hence, we can remove the restriction of $\boldsymbol{E}_m(\boldsymbol{r})$ defined in Eq. 78 (or Eq. 80) which requires it to be energy normalized, and extend the definition of $\boldsymbol{S}_m(\theta,\phi)$ to arbitrary $\boldsymbol{E}_m(\boldsymbol{r})$ as

$$\boldsymbol{S}_m(\theta,\phi) = \frac{\omega_c^2}{4\pi}\sqrt{\frac{\varepsilon_0}{U_m c^3}} \sum_{n=1}^{N} \alpha_n \boldsymbol{E}_m(\boldsymbol{r}_n) \cdot (1 - \hat{\boldsymbol{k}}\hat{\boldsymbol{k}}) e^{-ik_0 \hat{\boldsymbol{k}}\cdot\boldsymbol{r}_n}, \tag{96}$$

with U_m defined as

$$U_m \equiv \int \varepsilon(\boldsymbol{r}) \left|\boldsymbol{E}_m(\boldsymbol{r})\right|^2 d^3\boldsymbol{r}, \tag{97}$$

which corresponds to the energy of the m WGM. Also, any distribution of scatterers on the surface of a microresonator can be treated as a special case of surface roughness, with the dielectric perturbation given by

$$\Delta\varepsilon(\boldsymbol{r}) = \sum_{n=1}^{N} \varepsilon_0 \alpha_n \delta(\boldsymbol{r} - \boldsymbol{r}_n). \tag{98}$$

Combining Eqs. 96 and 98, a general form for $\boldsymbol{S}_m(\theta,\phi)$ is found as

$$\boldsymbol{S}_m(\theta,\phi)=\frac{\omega_c^2}{4\pi}\sqrt{\frac{1}{\varepsilon_0 U_m c^3}}\int \Delta\varepsilon(\boldsymbol{r})\boldsymbol{E}_m(\boldsymbol{r})\cdot(1-\hat{\boldsymbol{k}}\hat{\boldsymbol{k}})e^{-ik_0\hat{\boldsymbol{k}}\cdot\boldsymbol{r}}\,d^3\boldsymbol{r}. \tag{99}$$

A direct comparison shows that, except a constant, $\boldsymbol{S}_m(\theta,\phi)$ given by Eq. 99 in our model is equal to $r\boldsymbol{E}_{\rm ccw}^{\rm far}$ with $\boldsymbol{E}_{\rm ccw}^{\rm far}$ given by Eq. 16 in chapter II, provided that we can identify ($\hat{\boldsymbol{k}}$, $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\phi}}$) (which are the coordinates of the wave vector $\boldsymbol{k}$) in Eq. 99 with ($\hat{\boldsymbol{r}}$, $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\phi}}$)(which are the coordinates of the far-field position $\boldsymbol{r}$) in Eq. 16. Furthermore, according to Eq. 88, $\Gamma_{m,m}$ describes the scattering loss rate of the CCW (CW) WGM, similar to $\gamma_{\rm ccw}$ in the phenomenological model described in the last chapter. Their difference is that $\Gamma_{m,m}$, given by Eq. 95, is a spherical integration of $|\boldsymbol{S}_m(\theta,\phi)|^2$ in the wave vector space, and $\gamma_{\rm ccw}$, given by Eq. 17(of the last chapter), is a spherical integration of $|r\boldsymbol{E}_{\rm ccw}^{\rm far}|^2$ in the coordinate space. From the mathematical point of view, however, there is no difference in their detailed expressions and one can verify $\gamma_{\rm ccw}=\Gamma_{m,m}$.

3.4.4 Applications: two scatterers

We start with two identical submicron scatterers ($\alpha_1=\alpha_2=\alpha$) which are placed on the surface of a microdisk resonator with an angular separation of $2\phi_0$. The mode of the resonator is assumed to be TE polarized. From Eqs. 89 and 93, we obtain

$$G_{m,m}=2G_0, \tag{100}$$

$$G_{m,-m}=2G_0\cos(2m\phi_0), \tag{101}$$

$$\Gamma_{m,m}=2\Gamma_0\left(1+p(2k_0R\sin\phi_0)\cos 2m\phi_0\right), \tag{102}$$

$$\Gamma_{m,-m}=2\Gamma_0\left(p(2k_0R\sin\phi_0)+\cos 2m\phi_0\right). \tag{103}$$

where $G_0=\alpha\omega_c|f(\boldsymbol{r}_1)|^2/2V_m$ and $\Gamma_0=\alpha^2\omega_c^4|f(\boldsymbol{r}_1)|^2/6\pi V_m c^3$. Substituting these expressions into Eq. 88, we obtain two solutions which correspond to the split modes. Following the same convention as in the last chapter, we use ω_+ to indicate the mode

with a higher resonance frequency, with γ_+ being its loss rate. The solutions can be found as

$$\omega_\pm = \omega_c - 2G_0 \pm 2G_0|\cos 2m\phi_0| \quad , \tag{104}$$

$$\begin{aligned}\gamma_\pm =& 2\Gamma_0 \left(1 + p(2k_0 R\sin\phi_0)\cos 2m\phi_0\right) \\ & \mp 2\Gamma_0 \left(|\cos 2m\phi_0| + \mathrm{sign}(\cos 2m\phi_0) p(2k_0 R\sin\phi_0)\right).\end{aligned} \tag{105}$$

As has been discussed in the comparison with the independent-scatterer approach, when the two scatterers are separated at large distances, $p(x)$ can be neglected in Eq. 105 and we have

$$\gamma_\pm \approx 2\Gamma_0 \mp 2\Gamma_0|\cos 2m\phi_0|, \tag{106}$$

which is identical with Eq. 10 of the previous chapter from the independent-scatterer model ($N = 2$). However, when the two scatterers are close to each other, $p(x)$ has to be considered in Eq. 105. One special example is that the two scatterers overlap with each other (i.e., $\phi_0 = 0$ and $p(0) = 1$), which can be treated as a single-scatterer case. Equation 105 then predicts γ_+ and γ_- to be 0 and $8\Gamma_0$, respectively. In contrast, Eq. 10(of the previous chapter) provides γ_+ and γ_- to be 0 and $4\Gamma_0$, respectively. Using the result for the single scatterer ($\gamma_+ = 0$ and $\gamma_- = 2\Gamma_0$), we find that Eq. 105 is accurate and Eq. 10 only predicts half of the exact number for γ_- [6]. Another interesting observation is that when approaching the zeros of $\cos 2m\phi_0$, $p(2k_0 R\sin\phi_0)$ is generally nonzero, and

$$\gamma_\pm \approx 2\Gamma_0 \mp 2\Gamma_0 \mathrm{sign}(\cos 2m\phi_0) p(2k_0 R\sin\phi_0). \tag{107}$$

If we sweep ϕ_0 continuously, each time $\cos 2m\phi_0$ crosses its zero points, its sign will change and there will be abrupt changes in γ_+ and γ_-, indicating the zeros of $G_{m,-m}(\propto \cos 2m\phi_0)$ are singular points of γ_+ and γ_-. Moreover, since the relation between γ_+

[6] The scattering loss is proportional to the square of the dielectric perturbation so should be four times bigger if the dielectric perturbation doubles

and γ_- are reversed when passing the zeros of $\cos 2m\phi_0$, it is always possible to observe $\gamma_+ > \gamma_-$ in the neighborhood of $G_{m,-m} = 0$ (as long as $p(x)$ is not negligible).

3.4.5 Numerical approach for eigenfrequency simulation

To verify the derived theoretical results, we perform a numerical investigation for a two-scatterer example using an in-house two-dimensional (2-D) microresonator mode solver implemented in the COMSOL environment [41]. To obtain the scattering loss, perfectly matched layers (PMLs) are implemented based on the stretched coordinate method [46]. For example, in the cylindrical coordinate system, for the TE polarization, we have [44]

$$\left[\frac{\partial}{\rho\,\partial\rho}\left(\rho\frac{\partial}{\partial\rho}\right)+\frac{1}{\rho^2}\frac{\partial^2}{\partial\phi^2}\right]E_z = -k_0^2 n^2 s_\rho^2 E_z, \tag{108}$$

where n is the refractive index at each region; s_ρ is the complex coordinate stretching factor for the PML (light is only attenuated in the increasing ρ direction); and k_0 is the eigenvalue we try to obtain, which is related to the complex eigenfrequency ω as $\omega = k_0 c$.

Equation 108 is implemented and solved as a partial differential equation (PDE) in COMSOL [41]. Because COMSOL does not provide the cylindrical coordinate system for structures without axial symmetry, Eq. 108 is converted back to the Cartesian coordinate system as

$$\left[\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right]E_z = -k_0^2 n^2 s_\rho^2 E_z. \tag{109}$$

Figure 22 shows the structure we simulate, where the microdisk is centered at the origin. s_ρ is chosen to be the following form for the PML region

$$s_\rho = 1 + ia\frac{(\sqrt{x^2+y^2}-\rho_0)^2}{d^2}, \tag{110}$$

where ρ_0 and d are the starting radius and the thickness of the PML region, respectively, and a is a parameter that can be adjusted for the optimum performance of PML (we take $a = 3$ in our simulation). s_ρ is 1 for other regions.

Similarly, for the TM polarization, the equation can be implemented as

$$\left[\frac{\partial}{\partial x}\left(\frac{\partial}{n^2\partial x}\right)+\frac{\partial}{\partial y}\left(\frac{\partial}{n^2\partial y}\right)\right]H_z=-k_0^2 s_\rho^2 H_z. \tag{111}$$

The placement of n^2 inside the first-order derivative is to ensure correct boundary conditions when the PDE is solved (i.e., E_ϕ to be continuous) [45].

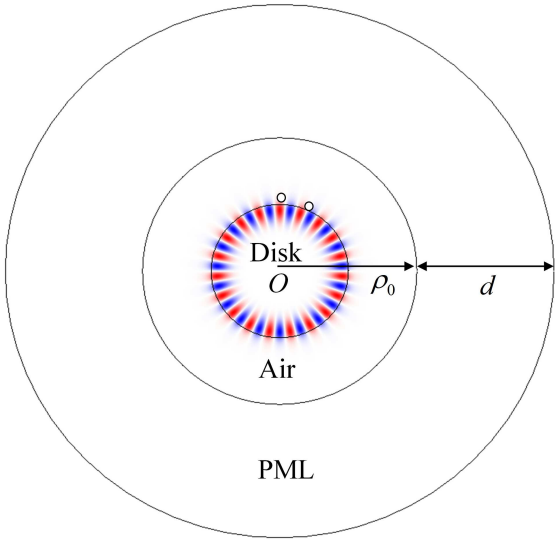

Figure 22: Simulated structure in COMSOL. The two small scatterers have been exaggerated in size for the illustration purpose.

3.4.6 Simulation results and discussions

The inset of Fig. 23 illustrates the studied structure, which consists of two 10-nm-radius scatterers attached to the surface of a 2-μm-radius microdisk resonator. We fix the position of one scatterer and sweep the position of the other. The complex eigenfrequencies of the coupled system are computed by the mode solver, offering both the resonance frequencies and the scattering loss rates for the two eigenmodes. In Fig. 23, two normalized parameters ω_{diff} and γ_{diff} are plotted, which are defined as

$$\omega_{\text{diff}} \equiv \frac{\omega_+ - \omega_-}{4G_0}, \tag{112}$$

$$\gamma_{\text{diff}} \equiv \frac{\gamma_+ - \gamma_-}{4\Gamma_0}. \tag{113}$$

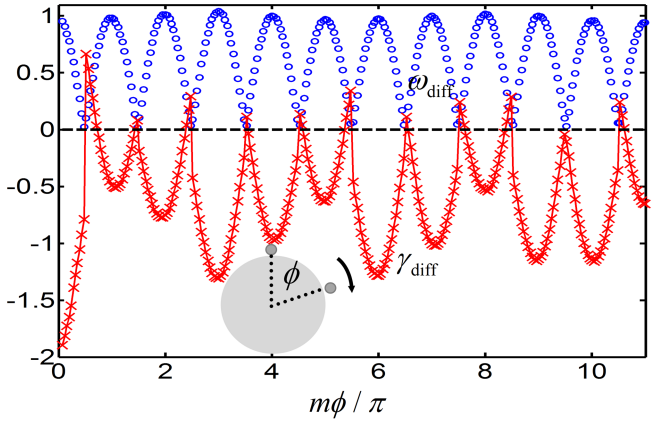

Figure 23: Simulation results of ω_{diff} and γ_{diff}, which are defined by Eqs. 112 and 113, respectively, for two 10-nm-radius scatterers attached to the surface of a 2-μm-radius microdisk resonator as illustrated by the inset. The refractive index of the microdisk is 2.829 (obtained using the effective index method for a 220-nm-thick silicon layer), and the refractive index of the scatterers is twice as big (i.e., $n_{\text{scatterer}} = 5.658$) to make the scattering effect significant. m is the azimuthal order of the WGM (TE polarized), which is 19. ω_c is obtained from the simulation for an ideal microdisk resonator without any scatterers as 1.2282723e15 rad/s, and G_0 and Γ_0 are obtained from the single-scatterer simulation as 5.88e10 rad/s and 9.3e8 rad/s, respectively.

From Eqs. 104 and 105, our theoretical model predicts

$$\omega_{\text{diff}} = |\cos m\phi|, \tag{114}$$

$$\gamma_{\text{diff}} = -|\cos m\phi| - \text{sign}(\cos m\phi)p(\phi), \tag{115}$$

where the angular separation between the two scatterers are $\phi = 2\phi_0$ as shown in the inset of Fig. 23. Comparing Fig. 23 to Eqs. 114 and 115, we find ω_{diff} agrees with Eq. 114 well; and γ_{diff} indeed changes sign when passing the zeros of ω_{diff}. We also notice that the magnitude of γ_{diff} can be less than -1; especially, it approaches to -2 when the two scatterers are close to each other. In Fig. 24, we plot two additional normalized parameters ω_{sum} and γ_{sum} defined as

$$\omega_{\text{sum}} \equiv \frac{\omega_+ + \omega_- - 2\omega_c}{4G_0}, \tag{116}$$

$$\gamma_{\text{sum}} \equiv \frac{\gamma_+ + \gamma_-}{4\Gamma_0}, \tag{117}$$

which are shown by blue solid line and red triangle marks, respectively. According to Eqs. 104 and 105,

$$\omega_{\text{sum}} = -1, \tag{118}$$

$$\gamma_{\text{sum}} = 1+p(\phi)\cos m\phi. \tag{119}$$

As observed from Fig. 24, ω_{sum} has a few percent fluctuations around the theoretical

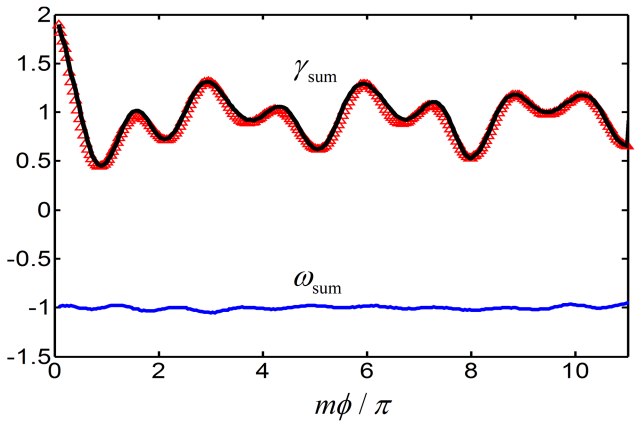

Figure 24: Numerical results of ω_{sum} and γ_{sum}, which are defined by Eqs. 116 and 117, respectively. The red triangles corresponds to γ_{sum} directly from simulation, while the black line corresponds to γ_{sum} obtain by extracting $p(\phi)$ from the simulation result of γ_{diff} first (using Eq. 115) and then computing the numerical values of Eq. 119.

value (i.e., -1), largely arising from the limited positioning resolution of the moving scatterer when we sweep it along the perimeter of the microdisk (1 nm in the COMSOL environment). With the help of Eq. 115, we could extract $p(\phi)$ from the numerical result of γ_{diff} shown in Fig. 23, and the result is depicted by the red triangles in Fig. 25. Moreover, using the obtained $p(\phi)$, γ_{sum} could be computed based on Eq. 119. The result, which is shown by black solid line in Fig. 24, agrees with the one from direct simulation (red triangle marks) well, implying that our theoretical model is self-consistent. One may notice that $p(\phi)$ shown in Fig. 25 is different from the one plotted in Fig. 21(b). This is because $p(\phi)$ shown in Fig. 21(b) is for the three-dimensional (3-D) case, while our simulation considers a 2-D model. The essential difference can be traced back to the difference in the free-space Green's function [35].

Employing the 2-D free-space Green's function and following a similar procedure as in the 3-D case, we obtain

$$p(\phi) = J_0(k_0 d) = J_0\left(2k_0 R \sin\frac{\phi}{2}\right), \tag{120}$$

where d is the distance between the two scatterers. This result can also be expected

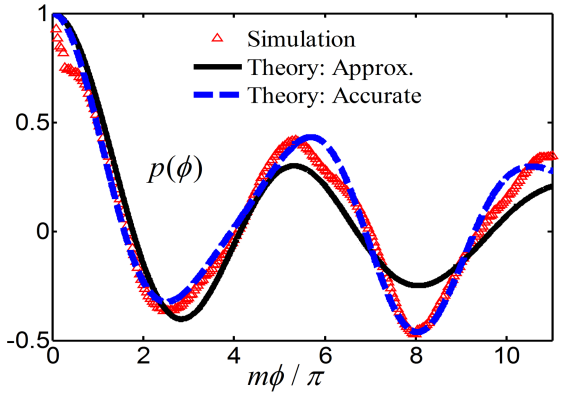

Figure 25: The red triangles corresponds to $p(\phi)$ extracted from the simulation result of γ_{diff} shown in Fig. 23 based on Eq. 115. The black solid line is the numerical result of Eq. 120, which is the theoretical prediction of $p(\phi)$ using the approximate 2-D Green's function. The blue dashed line is the theoretical prediction of $p(\phi)$ using the accurate 2-D Green's function.

from Eq. 92, by taking the inclination coordinate $\theta = \pi/2$ and skipping the integration over θ. In Fig. 25, we have plotted the predicted $p(\phi)$ given by Eq. 120 by black solid line, which agrees with the one extracted from the numerical simulation (shown by red triangles) reasonably well. The deviation there arises from two facts. First, we have certain positioning error when sweeping the scatterer in the simulation, as already mentioned for ω_{sum} in Fig. 24. Second, an accurate $p(\phi)$ requires taking the effect of the microdisk resonator to the free-space Green's functions into account, which has been omitted in Eq. 120 (or Eq. 92). In Ref. [49], a derivation is provided for the accurate calculation of $p(\phi)$ in the 2-D space, and the result is shown by blue dashed line in Fig. 25, which agrees with the simulation result well. From the asymptotic behavior of $J_0(x)$, one notice that in the 2-D case, the magnitude of $p(x)$ decreases

with the separation distance d of the two scatterers as $1/\sqrt{k_0 d}$, instead of $1/k_0 d$ as in the 3-D case (Fig. 21(b)). Therefore, a larger separation distance is required for 2-D models to neglect the effect of $p(\phi)$ ($d/\lambda_0 > 8$ for $|p(\phi)| < 0.1$).

3.4.7 Conclusions

In summary, we have developed a unified model that applies to an arbitrary number of scatterers, which provides a comprehensive understanding on the mode splitting and scattering loss in high-Q WGM microresonators. Compared with the independent-scatterer approach which is commonly used for the a-few-scatterer scenario, our work reveals that the independent-scatterer model has neglected the interference terms from different scatterers, whose effect decreases with the separation distance d as $1/k_0 d$ for 3-D cases and as $1/\sqrt{k_0 d}$ for 2-D cases. Thus, the independent-scatterer model only works when scatterers are well separated. Compared with the intuitive physical approach which is developed for the many-scatterer scenario, we have derived an additional coupling term between the CW and CCW modes (i.e., $\Gamma_{m,-m}$) that has been missing in the phenomenological model used by the intuitive physical approach. This modification leads to the prediction of asymmetric lineshapes in a self-consistent manner. Moreover, combined with numerical studies and experimental results, the unified model has provided many new understandings on the mode splitting and scattering loss in high-Q WGM microresonators. For example, we prove that the intuitive belief that $\gamma_+ \leq \gamma_-$ is not generally true, and counter examples can even be found for two scatterers attached to the surface of WGM microresonators. Our work also unveils that when mode splitting disappears, the scattering loss rates of the two eigenmodes are generally different, and γ_+ and γ_- become singular at these points. In Ref. [49], the model has also applied to the fabrication-induced roughness case, which corresponds to thousands of scatterers, and a good agreement between our model and experimental observations has been achieved. We believe such a unified

approach does not only fill the gap for the existing theoretical works on the mode splitting and scattering loss in high-Q WGM microresonators, but also will play an indispensable role for the practical applications of these phenomenons to produce the most accurate results.

CHAPTER IV

DEVICES ON SILICON-ON-INSULATOR PLATFORM

In this chapter, we are going to present two pieces of experimental work demonstrated on the silicon-on-insulator (SOI) platform. The first one is a third-order temperature-insensitive coupled-resonator filter, which serves as a critical building block element in on-chip terabit/s networks. The second one is on the use of high-Q microdisks for compact and low-loss delay lines, which are especially useful for narrowband filters for RF signal processing.

4.1 A temperature-insensitive wideband coupled-resonator filter

Recently, bandwidth capacity on the order of terabit/s has been in real demand for on-chip interconnects for switching and routing signals among multicore microprocessors [5, 6]. Integrated optical filters are one of the promising solutions to achieve on-chip interconnects with a reasonable power budget. In such approaches, the large bandwidth is provided by the aggregate bandwidth contributed by each WDM channel, which itself is accommodated by each filter channel. For this reason, filters with many identical channels are required. In addition, the passband of these filter channels is desired to be flat to minimize the signal distortion. However, because of the strong material dispersion of silicon, achieving a flat-band response over a wide bandwidth is difficult. Another challenge for such filters is related to the existence of random hot spots in the on-chip environment, where the local temperature variation can be up to tens of degrees. The large temperature fluctuation will make the filter channels drift with temperature, causing the wavelength-registration problem [17].

In previous studies, both single- and coupled-resonator filter architectures implemented in the SOI platform have been reported [16, 17, 50, 51]. In Ref. [50], 20 wavelength channels were simultaneously switched using a single microring resonator. Because of the narrowband filter channels (3-dB bandwidth around 0.1 nm), small temperature shifts can easily cause a frequency mismatch between the WDM signal and the filter channel. In Refs. [16, 17, 51], coupled-resonator structures with large bandwidth filter channels were utilized to provide extra thermal guard bands. However, these coupled-resonator filters suffered from a large free spectral range (FSR), which makes scalability to large WDM channel counts challenging. In addition, though a multimode interferometer coupler has been employed for a wavelength-insensitive coupling between the access waveguide and the end resonators in Ref. [17], the coupling between resonators is evanescent and exhibits a strong dispersion over the entire wavelength range. Consequently, the overall filter performance shows a strong wavelength dependence, and large distortions are observed in Ref. [17].

In this work, we show that using an evanescent coupling scheme for both the waveguide-to-resonator and resonator-to-resonator couplings can ensure the flat-band filter response over a large wavelength range. One drawback of the evanescent coupling scheme is that the bandwidth of the filter channel increases with wavelength, resulting in a redundant channel bandwidth at high wavelengths. We show that by overlaying a polymer cladding with a negative thermo-optic coefficient on top of the silicon device, we can significantly reduce the sensitivity of the filter performance to the ambient temperature variations. Moreover, through careful balancing between the dispersion of the bandwidth and the thermal property of the filter, the redundant bandwidth of the filter channel caused by the dispersion can be employed as a thermal guard band. As a result, the filter can accommodate 21 WDM channels with a data rate up to 100 gigabit/s (Gb/s) per wavelength channel, while providing a sufficient thermal guard band to tolerate more than $\pm 15^{o}$C temperature fluctuations in the

on-chip environment [17, 51].

4.1.1 Device design

In response to the design challenges mentioned at the beginning for the multi-channel filters, the design approach adopted in this work consists of two aspects, the first being a broadband coupling engineering and the second being a temperature-insensitive design. Key results for the first aspect of the design are summarized in Fig. 26. The structure investigated in this work is a third-order filter as shown by Fig. 26(a), where κ_0^2 and κ_1^2 denote the power-coupling ratios between the end resonator and the access waveguide and that between the adjacent resonators, respectively. Because such coupled-resonator filters are sensitive to the coupling-induced resonance frequency shifts (CIFS), the coupling between the access waveguide and the end resonator is achieved through a symmetric coupler which has the symmetry along the wave propagation direction. Using the approximation method for the coupling-induced phase shift of the symmetric couplers (see discussions in Chapter III), the CIFS could be calculated and corrected by pre-distorting the resonator [51]. To minimize the impact of the strong material dispersion of silicon as wells as the waveguide geometric dispersion, the FSR of the filter has to be as small as possible. Since the targeted 3-dB bandwidth of the filter channel is on the order of 1 nm to accommodate a wideband wavelength channel (e.g., 100 Gb/s), an FSR around 5 nm is empirically found to be appropriate to provide the necessary bandwidth and adequate channel isolation. To obtain a flat-band response, κ_0^4 and κ_1^2 have to satisfy certain relation. For example, in the narrowband-filter design, $\kappa_1^2/\kappa_0^4 \approx 1/8$ is required, while in the wideband-filter design the exact ratio of κ_1^2/κ_0^4 may be different [51]. Nevertheless, to preserve this ratio for the whole wavelength range, κ_1^2 and κ_0^4 must follow similar wavelength-dispersion curves.

In this work, resonators and waveguides are evanescently coupled to each other.

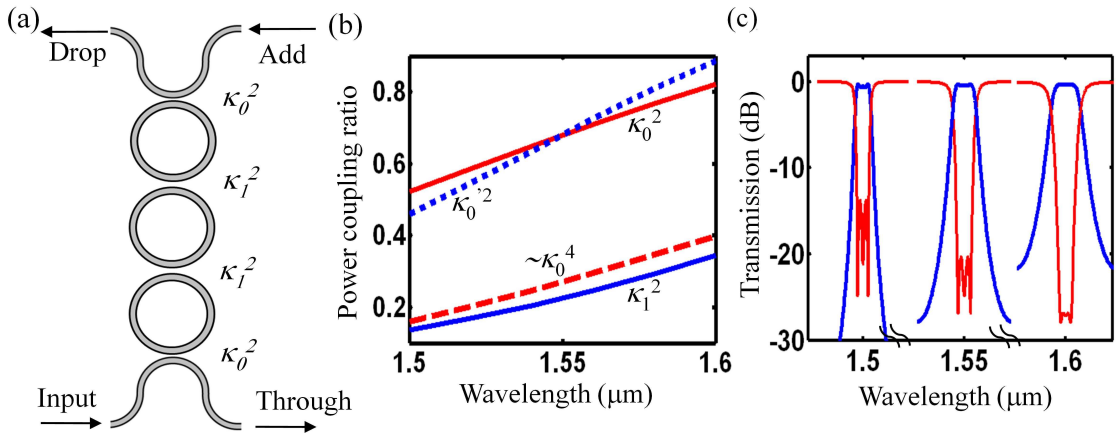

Figure 26: (a) Schematic of the third-order filter; (b) power-coupling ratios of the synchronous parallel waveguide couplers with different gaps and coupling lengths: κ_0^2 (gap 160 nm, length 4.4 μm), $\kappa_0^{'2}$ (gap 300 nm, length 10.2 μm), and κ_1^2 (gap 300 nm, length 5.2 μm); (c) three simulated filter channel responses with power-coupling ratios κ_0^2 and κ_1^2 provided in (b).

The waveguide dimension is first optimized to fine tune the thermal property of the filter. We use a waveguide with a height of 144 nm and a width of 425 nm covered with a top polymer cladding to achieve an athermal response exactly at the wavelength of 1.5 μm for the fundamental transverse electric mode (i.e., the predominant electric field is oriented in the silicon layer). The choice of this wavelength for the athermal operation will be explained later. Since we are targeting many filter channels ($\sim$ 20) over a wide wavelength range ($\sim$ 100 nm), the resonator's coupling to the access waveguides and the adjacent resonators needs to be optimized to achieve a wideband performance. In this regard, a microring performs better than a racetrack because the microring has a shorter coupling length under the same FSR and power coupling ratio. While a detailed comparison could be performed through three-dimensional finite-difference time-domain (FDTD) simulations, here we adopt a simpler approach where the coupling region of the two coupled structures is modeled by a parallel waveguide coupler with the same gap and an equivalent coupling length [42]. Figure 26(b) shows the simulation results for the dispersion of the power coupling rate of a

synchronous parallel waveguide coupler (with the mentioned waveguide dimension) with different gaps and coupling lengths. The upper two curves correspond to two couplers with identical power coupling ratios at 1.55 μm. As can be seen, the one with a shorter coupling length exhibits a weaker dispersion over the wavelength range of $1.5 - 1.6$ μm. This observation leads us to use microrings (instead of racetracks) with a radius of 21 μm, which follows the choice of the FSR of 5 nm.

The gaps between the waveguide-resonator and resonator-resonator are found using the following procedure. First we employ the modified coupled mode theory which has been discussed in detail in Chapter III to calculate κ_0^2 and κ_1^2 for a range of gaps, and a table could be drawn. Then based on the required bandwidth of the filter channel at a specific wavelength (typically at 1.55 μm), we determine the appropriate κ_0^2 and κ_1^2. The gaps can be simply looked up from the computed coupling table, which are found to be 160 nm for the resonator and waveguide coupling and 300 nm for the coupling between resonators. Finally, κ_0^2 and κ_1^2 are computed for all the wavelengths using the modified CMT, and the results are shown in Fig. 26(b), where the upper solid and the lower solid lines correspond to κ_0^2 and κ_1^2, respectively. The condition for a flat-band response is also verified in Fig. 26(b), where κ_0^4 and κ_1^2 exhibit similar dispersion curves. Using these parameters for the device, the overall filter response is obtained through a rigorous matrix analysis [52]. In Fig. 26(c), we show three such filter channels around 1.5 μm, 1.55 μm, and 1.6 μm. From Fig. 26(c), we observe that: (1) filter responses are reasonably good, and flat-band filter channels exhibit more than 15-dB extinctions in the through port and more than 20-dB out-of-band rejections in the drop port; (2) the 3-dB bandwidth increases from 0.75 nm at 1.5 μm to 1.5 nm at 1.6 μm, primarily because of the dispersion; (3) the bandwidth of the filter channels increases with wavelength, and the finesse of the filter decreases; consequently, the out-of-band rejection becomes smaller (more than 30 dB at 1.5 μm and only 20 dB at 1.6 μm); and (4) the through-port responses are improved at high

wavelengths, indicating a better impedance match among the coupled resonators.

Following the broadband coupling engineering, we proceed to the second aspect of the design for the temperature-insensitive operation. Major results of this part are shown in Fig. 27. Figure 27(a) shows the concept of a thermal guard band. The solid curve shows the pass-band response of the optical filter, and the solid filled rectangle shows the information bandwidth of the optical signal. The information bandwidth can be as large as the 3-dB bandwidth of the filter channel without incurring strong distortions [36]. The remaining spectral portion of the filter bandwidth, shown by the hatched region, illustrates the available thermal guard band. Our design for the temperature-insensitive filter chooses the complete athermal operating point at the filter channel with the smallest bandwidth (in our case this is at $\lambda = 1.5$ μm). This ensures that throughout the entire working wavelength range, the overall available thermal guard band is optimally utilized.

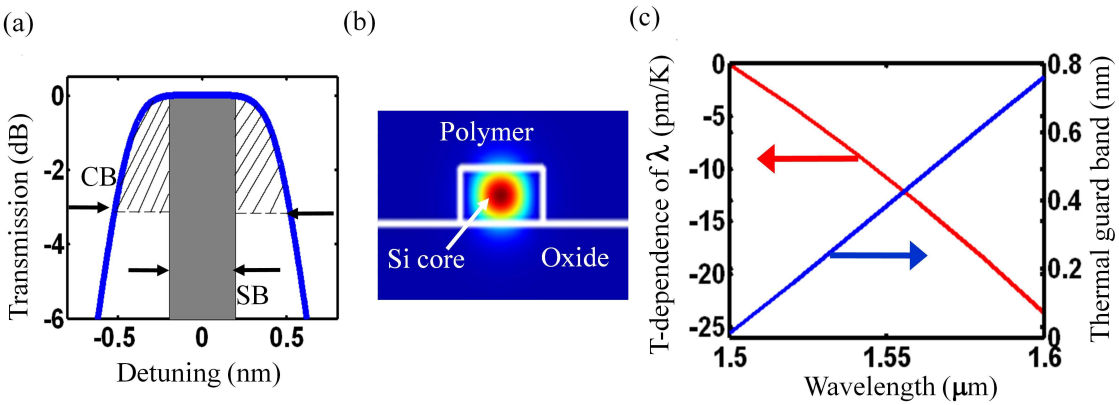

Figure 27: (a) Illustration of extra channel bandwidth as thermal guard band (hatched region); CB: filter channel bandwidth; SB: signal information bandwidth; (b) cross section of the waveguide structure on an SOI substrate with a polymer cladding; (c) the left axis shows the wavelength temperature dependence of the filter, and the right axis shows the thermal guard band for 100 Gb/s wavelength channels.

To achieve the temperature-insensitive operation, we overlay the silicon device with an acrylate polymer (see Fig. 27(b)) with a refractive index close to oxide (1.45)

and a large negative thermo-optic coefficient of -4.5×10^{-4}/K [53]. To increase the overlap between the optical field and the polymer cladding, the silicon device layer is thinned to 144 nm, and the required waveguide width is calculated to be 425 nm for the athermal operation at 1.5 μm. Figure 27(c) shows the simulated temperature dependence of the filter over the entire working wavelength range, which increases from 0 pm/K at 1.5 μm to -23 pm/K at 1.6 μm.

To illustrate the operation of our filter, we consider the following scenario: if the filter is used for incoming signals modulated at a 100 Gb/s data rate per wavelength channel, at 1.5 μm the available guard bandwidth is small (3-dB bandwidth 0.75 nm), but because of the centering of the athermal operation to this wavelength, the filter channel shifts little with temperature change. At the other end of the filter operation, i.e., at 1.6 μm, the filter channel has a relatively large temperature dependence of -23 pm/K, but it also has a large available thermal guard band of 0.75 nm (with 3-dB bandwidth of 1.5 nm), thus allowing for $\pm$15°C emperature fluctuations. For lower-data-rate signals, even larger temperature fluctuations can be tolerated.

4.1.2 Device fabrication and characterization

The device is fabricated on an SOI wafer with a 144-nm-thick silicon layer on top of a 3-μm-thick oxide substrate. The device is patterned using a JBX-9300FS e-beam lithography system based on the HSQ process, followed by plasma etching with the Cl_2 gas. Details of these fabrication steps could be found in Chapter II. For this device, the remaining HSQ has to be removed so the acrylate polymer can be applied directly on top of the silicon waveguide, otherwise the thermal compensation by the polymer will be discounted. To do that, we dip the sample in the buffered oxide etcher, which also attacks the oxide underneath the silicon waveguide. Fortunately, after the spin-coating, the polymer will fill the undercut region; and since the refractive index of the polymer is close to that of oxide, such undercutting has a negligible impact on

the device performance.

Figure 28(a) shows the scanning-electron micrograph (SEM) of the fabricated device, and Fig. 28(b) provides the measured transmission result, where 21 filter channels with more than 10-dB through-port extinctions and 20-dB drop-port out-of-band rejections are obtained. The insertion loss of the filter can be estimated by the difference between the maxima of the through and the drop ports to be less than 1 dB. In Fig. 28(c), we show three representative filter channels at wavelengths 1.5 μm, 1.55 μm, and 1.6 μm, respectively. Compared with the simulation results shown in Fig. 27(c), the 3-dB bandwidths agree well with the design, being 0.75 nm, 1.12nm, and 1.5 nm for the corresponding channels. The out-of-band rejections in the drop port at low wavelengths are limited to 22 dB, while the ones predicted by the simulation shown in Fig. 27(c) are higher ($>$ 25 dB). This deviation is believed to arise from the limited dynamic range of our measurement setup, which has a relatively large noise floor. The reduction of the out-of-band rejections at high wavelengths, as seen for the filter channels near 1.6 μm, is expected and is in agreement with the theory. The through-port responses are a little degraded from the design, as the extinction drops to 10 dB at 1.5 μm instead of 15 dB, and the shape is also deformed. This is because the through-port responses are very sensitive to the resonance-frequency mismatches between the coupled resonators [42], which can be easily caused by the imperfect fabrication process.

The temperature behavior of the filter is also tested by heating the device using a thermal stage and measuring the transmission subsequently. In Fig.28(c), the corresponding through-port responses after a 10^{o}C temperature rise are shown by the dotted lines (deliberately lowered down by 3 dB for a better illustration). We can verify that the filter is indeed athermal at 1.5 μm and has wavelength temperature dependencies of -12 pm/K and -25 pm/K at wavelengths 1.55 μm and 1.6 μm, respectively. This is a significant improvement compared to the conventional unclad

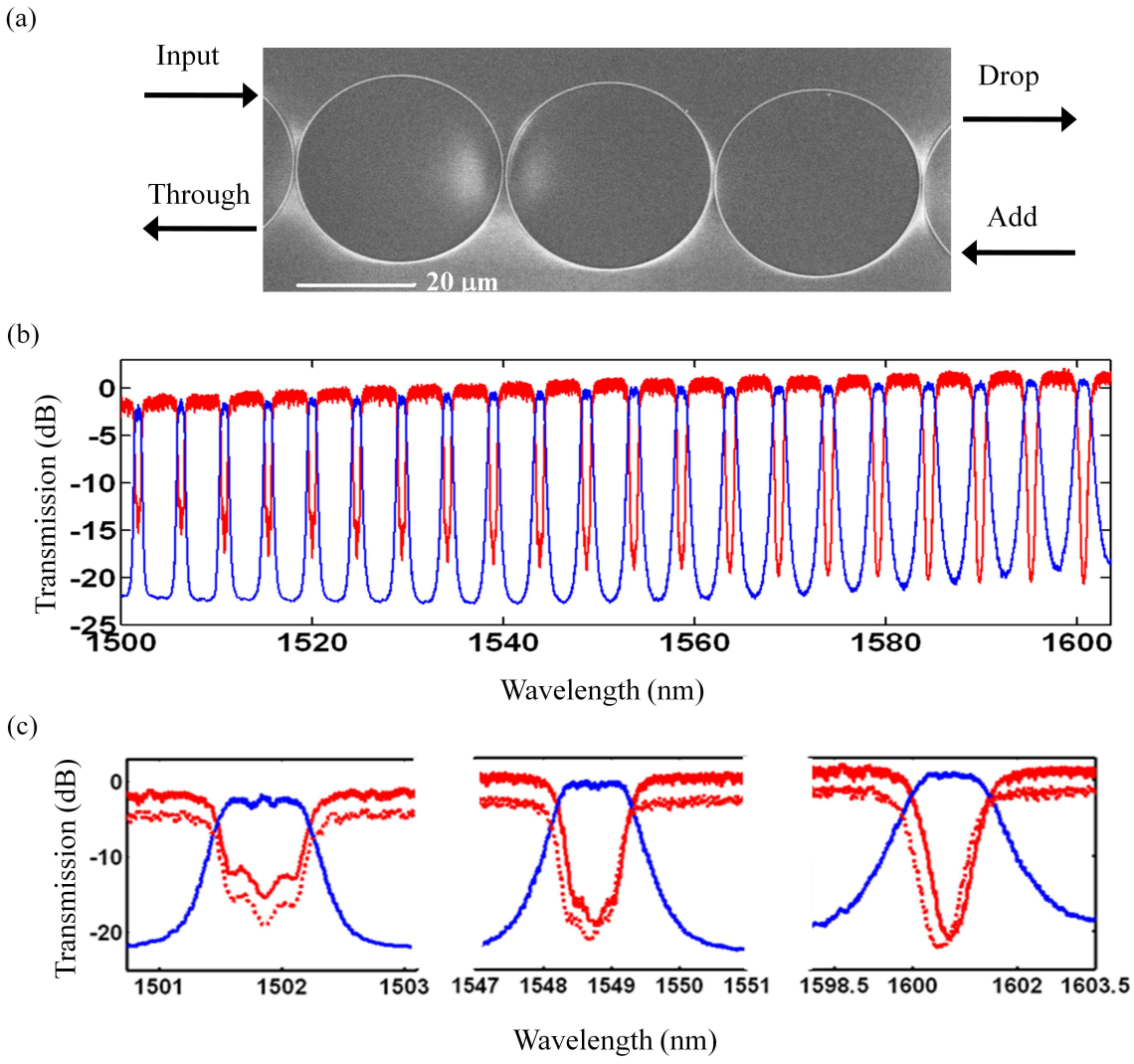

Figure 28: (a) SEM of the fabricated filter; (b) measured transmission responses for the fabricated filter; (c) detailed measured spectra of three representative channels. The dotted lines show the through-port responses after a 10°C temperature rise.

filters ($\sim$ 70 pm/K) [17].

In conclusion, we have designed and demonstrated a third-order flat-band filter that can accommodate 21 WDM channels with a data rate up to 100 Gb/s per wavelength channel, totaling up to a 2.1 terabit/s aggregate capacity. The filter can tolerate more than thirty-degree temperature fluctuations in the on-chip environment. The footprint of the filter is less than 0.005 mm^2. In the future, we envision the addition of active switching functionality to the filter through carrier injection (either all-optically or electrically), and a high-performance switch with a switching speed on the order of nanoseconds is expected. Such a compact, wideband, and temperature-insensitive optical switch would be a critical building block in future on-chip optical networks.

4.2 Tunable and low-loss microdisk-based delay lines

Integrated optical filters have numerous important applications. Recently, wideband-tunable narrowband filters have attracted a lot of research interest for radio-frequency (RF) signal processing, which can potentially outperform current RF technologies in terms of bandwidth and dynamic range [54]. Developments in this area are fueled by progress made in both the design and implementation of narrowband optical filters. For example, a systematic design methodology has been introduced in Ref. [54], which requires only three basic building elements, i.e., couplers, phase shifters and all-pass filters (APFs) for the design of almost arbitrary filters. Figure 29(a) shows one such example, where the filter is a cascade of many unit cells. Each unit cell is composed of a 3-dB coupler, a phase shifter and an APF, and therefore its transmission response has one pole and one zero [54]. Using a similar approach as for digital filter design, any filtering function can be decomposed to poles and zeros that can be implemented by these unit cells. For example, in Fig. 29(b), a Butterworth filter is demonstrated with

the use of eight stages of unit cell. Regarding the physical realization, the silicon-on-insulator (SOI) platform emerges as a suitable choice due to the possibility of forming low-loss waveguides, accurate couplers, and tunable devices. Moreover, being compatible with the standard complementary metal-oxide-semiconductor (CMOS) technology, it promises large-scale integration and manufacturing at a low cost. This is important since in practice many stages of unit cell (> 16) are required to achieve desired filtering responses.

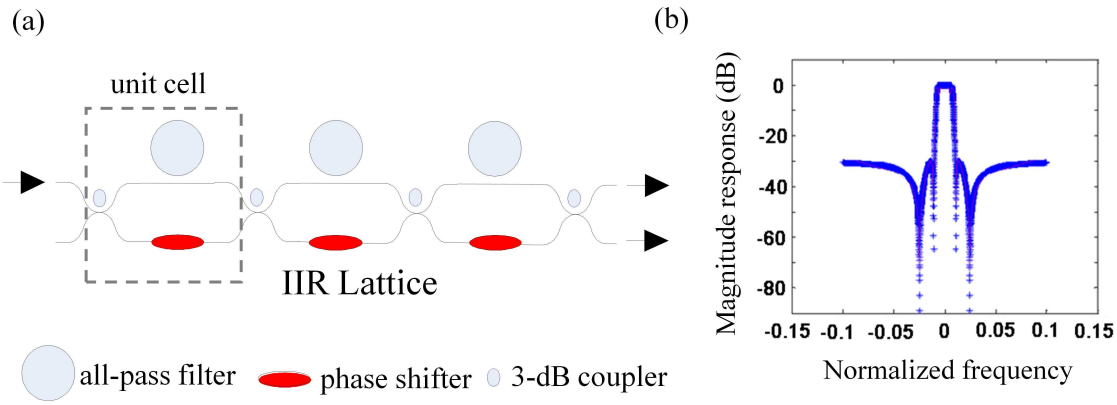

Figure 29: (a) Schematic of a IIR lattice filter which is a cascade of unit cells; (b) the amplitude response of a Butterworth filter made of eight stages of unit cell.

For narrowband filters, the implementation of 3-dB couplers and phase shifters are straightforward (although, fast-response phase shifters are quite challenging). The APF, in its simplest form, can also be constructed using a closed waveguide loop. However, such a loop has to be very big to access the RF domain, since the time delay we want to get from the loop is on the order of 100 ps (so the bandwidth of the filter can be tuned from a few GHz to tens of MHz), which translates to a propagation length of 7 mm for typical silicon waveguides. More importantly, the insertion loss from such a loop has to be negligible, which is challenging for silicon ridge waveguides with typical propagation loss rates (per unit length) on the order of 2-3 dB/cm [55]. In Ref. [56], rib waveguide is used for the low-loss delay line.

Though the propagation loss rate is reduced to 0.5 dB/cm, the bending radius of the rib waveguides has to be large enough to avoid radiation loss. As a result, the size of the delay line is on the order of millimeters, making dense integration difficult. On the other hand, delay lines based on microresonators, such as microrings and microdisks, can be more compact. Microrings are single mode and easy to work with, but their propagation loss is similar to ridge waveguide's (2-3 dB/cm). Microdisks usually show lower loss, but due to the multimode nature, its implementation is not straightforward. In this work, we will focus on the design of delay lines using over-coupled microdisk resonators. In addition, a first-order tunable narrowband filter is experimentally demonstrated based on such a delay line structure.

4.2.1 Device design

To design a microdisk-based delay line, several challenges should be overcome. First of all, microdisks are multimode; one usually has to work with one particular mode while avoiding interference from others. To make this task easier, a modified microdisk structure is employed here. As shown by the scanning-electron micrograph (SEM) in Fig. 30(a), the microdisk has an outer radius of 20 μm with a 230-nm-thick silicon device layer and a 1-μm-thick oxide cladding. A hole is etched in the center of the microdisk to minimize the number of modes yet not perturbing the first few radial modes (i.e., the fields of the first few radial modes do not overlap with the inner wall). The geometry of the structure is optimized using a three-dimensional finite-element-method (FEM) microresonator mode solver implemented in COMSOL (details could be found in Ref. [36]), and the resulting structure is a microdonut with a width of 4 μm. The mode profiles and effective indices for the lowest four radial TE modes (electric field parallel to the microdisk plane) of the microdonut are shown in Figs. 30(b) and (c), respectively. As can be seen, the optimized microdisk resonator is still multimode though the number of modes has been reduced. The

second design challenge is the required strong coupling between the waveguide and the microresonator to achieve delays in the order of 100 ps. This is difficult because the microdisk modes are highly confined inside the disk, and their field overlaps with the mode of the access waveguide are small.

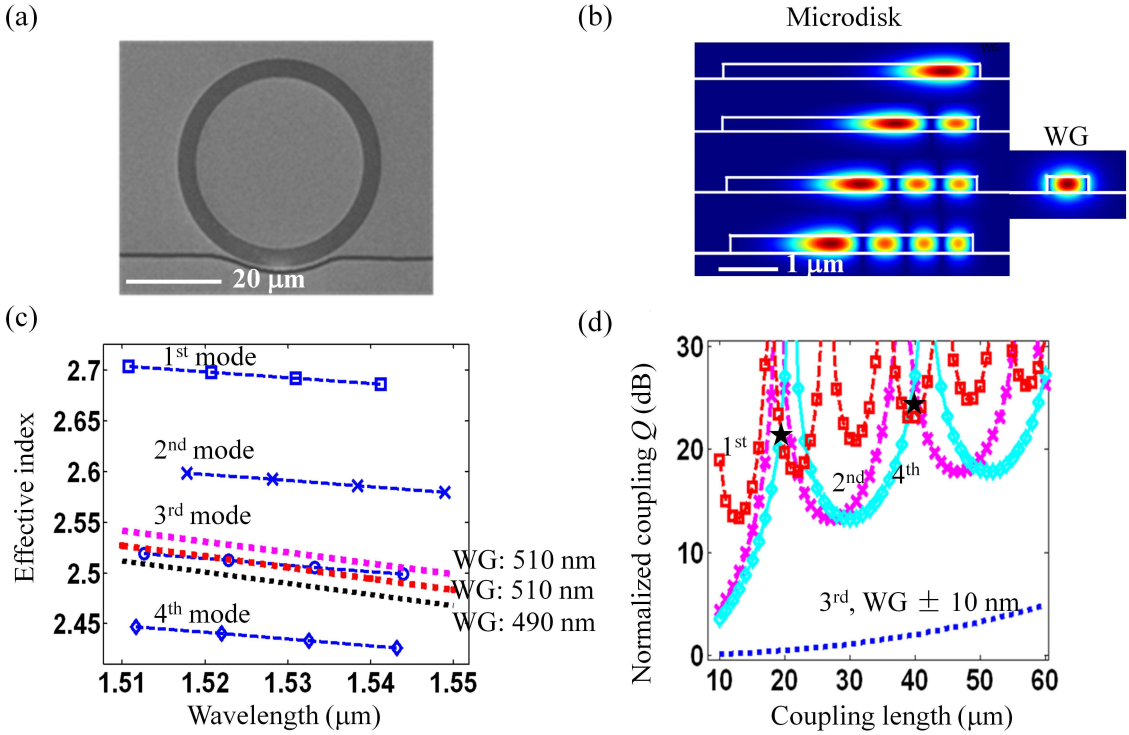

Figure 30: (a) SEM of a 20-μm-radius microdisk resonator pulley coupled to an access waveguide with a width of 500 nm. (b) Mode profiles of the lowest four radial TE modes of the microdisk resonator (left) and the access waveguide (right). (c) Effective indices for the microdisk modes (dashed line) and the waveguides (dotted line) with different widths. (d) Coupling Qs (Q_c) of the 1st-(square), 2nd-(cross), and 4th-order (diamond) microdisk modes relative to that of the phase-matched 3rd-order mode as a function of the coupling length (the coupling gap is 250 nm). The dotted line corresponds to the case that the 3rd-order mode has certain phase mismatch caused by a 10-nm deviation in the access waveguide width from the phase-matching point (i.e., 500 nm). The two stars mark the two optimum coupling lengths which give us more than 20 dB suppression of coupling for the undesired modes.

To resolve these two problems mentioned above, we use the pulley coupling scheme, as shown by the SEM in Fig. 30(a), where the modes of the microdisk and the waveguide weakly interact for a long coupling length to achieve the desired coupling strength

[57]. Furthermore, the pulley coupling scheme enables us to selectively couple to the desired mode, since for long coupling lengths the coupling process becomes very sensitive to the phase-matching condition. To see this, we use the result from Chapter III for the asymmetric couplers,

$$\begin{aligned} P_{21}(L) &= \frac{4|\gamma_{21}|^2}{|(\gamma_{11}-\gamma_{22})^2+4\gamma_{12}\gamma_{21}|}\sin^2\left(\sqrt{(\frac{\gamma_{11}-\gamma_{22}}{2})^2+\gamma_{12}\gamma_{21}}\ L\right) \\ &= (|\gamma_{21}|L)^2\ \mathrm{sinc}^2\left(\sqrt{(\frac{\gamma_{11}-\gamma_{22}}{2})^2+\gamma_{12}\gamma_{21}}\ L\right), \end{aligned} \tag{121}$$

where the $\mathrm{sinc}(x)$ function is defined as $\mathrm{sinc}(x) = \sin(x)/x$. Let us assume the "1" represents the access waveguide, and "2" represents the microdisk mode with a specific radial mode order. First we consider the phase matched radial mode, for which $\gamma_{11} = \gamma_{22}$, and $P_{\mathrm{match},21} = \sin^2(|\gamma_{21}|L) \approx (|\gamma_{21}|L)^2$ if $|\gamma_{21}|L \ll 1$ [1]. Then, for the phase-mismatched radial modes, we assume their γ_{12} and γ_{21} are still similar to those of the matched mode (which are given by field overlaps and in practice are a few tens of percent different). But for those radial modes $|\gamma_{11}-\gamma_{22}|$ could be much bigger than $|\gamma_{21}|$ (or $|\gamma_{12}|$), due to different effective indices as shown in Figs. 30(c). Consequently,

$$\frac{P_{n,21}(L)}{P_{\mathrm{match},21}(L)} \approx \mathrm{sinc}^2\left(\Delta\gamma_n L\right), \tag{122}$$

where n stands for the radial mode order and $\Delta\gamma_n \equiv (\gamma_{n,11}-\gamma_{n,22})/2$. From Eq. 122, we find that if the radial mode n is phase mismatched, its coupling relative to that of the matched one will generally decrease as L increases [2]. Therefore, if we make one mode of the microdisk resonator phase-matched to the access waveguide and other modes phase-mismatched, in principle by using a large L we can ensure only one radial mode is excited with others being suppressed.

In this work, we choose to work with the third-order radial mode of the microdisk resonator due to its ease of phase matching to the typical waveguide widths around

[1] We assume $\gamma_{12} = \gamma_{21}^*$, which is true for symmetric waveguide couplers but may not be the case for asymmetric waveguide couplers. However, mathematical rigor is not our focus here.

[2] The sinc function is rapidly oscillating. Here we are talking about the envelop of the curve.

500 nm, whose mode profile and effective indices are also provided in Figs. 30(b) and (c), respectively. As already pointed out in the discussions for asymmetric waveguide couplers in Chapter III, the conventional phase-matching condition, i.e., $\beta_1 = \beta_2$, is not exact. But here we will use it as a good approximation for the phase-matching condition, which corresponds to the intersection point in Fig. 30(c) where the dispersion curves of the microdisk and the waveguide meet. At a specific wavelength, the actual coupling strength may deviate from its designed value due to the limited fabrication accuracy on controlling the waveguide width (which is on the order of 10 nm in our fabrication process). Therefore, one has to optimize the coupling length to effectively suppress the undesired resonant modes, but still to tolerate the fabrication-imperfection-induced phase mismatch for the desired mode. This optimization is carried out based on Eq. 121, where all the parameters are calculated using the modified coupled mode theory discussed in Chapter III. In Fig. 30(d) , the coupling quality factors (coupling Qs) for the undesired modes as well as that for the desired mode with a 10-nm deviation in the waveguide width from the phase-matching point (as discussed in Chapter II, the coupling Q of each radial mode is inversely proportional to its coupling strength to the access waveguide). Note that all the values in Fig. 30(d) are normalized to the coupling Q of the phase-matched mode. From Fig. 30(d), there are two optimum lengths, i.e., 19 μm and 37 μm, with which the coupling Qs of the adjacent modes are more than 20 dB higher than that of third-order microdisk mode. We choose to use the 37 μm so the gap between the microdisk and waveguide can be relaxed to 250 nm, and the fabrication error ($\pm$10 nm deviation in waveguide width) will only cause a small increase (2 dB) in the coupling Q of the desired mode.

4.2.2 A tunable first-order filter: design

To illustrate the application of the microdisk-based delay line, we apply it to a first-order filter architecture as shown in Fig. 31, where an APF is placed at the upper arm of a Mach-Zehnder interferometer (MZI). The APF itself consists of a racetrack resonator with a microdisk-based delay line inserted in the loop. The coupling between the MZI arm to the APF is achieved through another MZI with a tunable phase shifter (ϕ_{MZI} as shown in Fig. 31) implemented in its lower arm. Two additional phase shifters are employed to tune the resonance frequencies of the microdisk (ϕ_{disk}) and the racetrack resonators (ϕ_{FB}). In the lower arm of the MZI, a similar APF without the microdisk delay line is installed, which is used to balance the MZI response, as will be explained later. Using the transfer matrix analysis [52], the transfer function of the upper APF is obtained as

$$T_{\mathrm{APF}} = i\exp(i\frac{\phi_{\mathrm{MZI}}}{2})\frac{\sin(\phi_{\mathrm{MZI}}/2) - T_{\mathrm{disk}}\exp\left(i(\phi_{\mathrm{MZI}}+2\phi_{\mathrm{FB}}-\pi)/2 + i\beta_r L_{\mathrm{APF}}\right)}{1-\sin(\phi_{\mathrm{MZI}}/2)\,T_{\mathrm{disk}}\exp\left(i(\phi_{\mathrm{MZI}}+2\phi_{\mathrm{FB}}-\pi)/2 + i\beta_r L_{\mathrm{APF}}\right)}, \tag{123}$$

where $i=\sqrt{-1}$; β_r is the complex propagation constant of the forming waveguide of the APF with the imaginary part corresponding to the propagation loss; L_{APF} is the perimeter of the APF; and T_{disk} is the transfer function of the microdisk resonator given by

$$T_{\mathrm{disk}} = \frac{r-\exp(i\phi_{\mathrm{disk}}+i\beta_d L_{\mathrm{disk}})}{1-r\exp(i\phi_{\mathrm{disk}}+i\beta_d L_{\mathrm{disk}})}, \tag{124}$$

where β_d is the complex propagation constant of the microdisk resonator; L_{disk} is its perimeter; and $r=\sqrt{1-\kappa^2}$ with κ being the power coupling coefficient between the microdisk resonator and the access waveguide.

If the loss of the microdisk is neglected, T_{disk} can be approximated as

$$T_{\mathrm{disk}} \approx -\exp\left(2i\tan^{-1}\left(\frac{\delta\omega_d\tau_d}{2}\right)\right), \tag{125}$$

where $\delta\omega_d$ is the angular frequency detuning relative to the resonance frequency ω_d of the microdisk, and τ_d is the group delay on resonance. In Eq. 123, if we also neglect

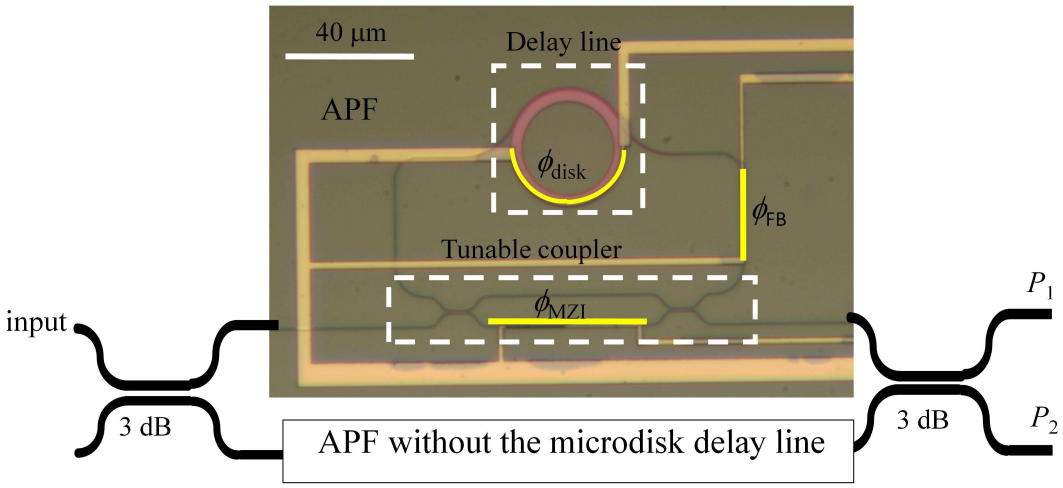

Figure 31: Schematic of a first-order tunable filter. The APF is shown by the optical micrograph. The highlighted areas show the metallic microheaters that work as phase shifters. The three phase shifters share the same ground for current injection.

the loss term in β_r and expand it around the resonance frequency of the microdisk as

$$\beta_r(\omega) \approx \beta_r^0(\omega_d) + \frac{1}{v_g}\delta\omega_d, \tag{126}$$

where v_g s the group velocity of the waveguide mode. By controlling the phase shift ϕ_{FB}, the following condition can be satisfied

$$\frac{(\phi_{\mathrm{MZI}} + 2\phi_{\mathrm{FB}} + \pi)}{2} + \beta_r^0(\omega_d)L_{\mathrm{APF}} = 2m\pi, \tag{127}$$

where m is an integer. Under this condition, and with a further approximation of $\tan^{-1}(\delta\omega_d\tau_d/2) \approx \delta\omega_d\tau_d/2$ for $|\delta\omega_d\tau_d/2| < 1$, Eq. 123 can be simplified as

$$T_{\mathrm{APF}} \approx i\exp(i\frac{\phi_{\mathrm{MZI}}}{2})\frac{\sin(\phi_{\mathrm{MZI}}/2) - \exp\left(i\delta\omega_d\left(L_{\mathrm{APF}} + \tau_d v_g\right)/v_g\right)}{1 - \sin(\phi_{\mathrm{MZI}}/2)\exp\left(i\delta\omega_d\left(L_{\mathrm{APF}} + \tau_d v_g\right)/v_g\right)}. \tag{128}$$

Equation 128 can now be interpreted as the response of an APF with a tunable power coupling coefficient of $\cos(\phi_{\mathrm{MZI}}/2)$ and an effective perimeter of $L_{\mathrm{APF}} + \tau_d v_g$. Compared to the physical length of L_{APF}, it suggests much more compact filters can be designed using this APF as a unit cell. To give an example, the APF shown in

Fig. 30 has a perimeter of 300 μm, and the microdisk resonator is designed to provide a 100-ps delay on resonance. As a result, the effective perimeter is about 7.8 mm, a size that the waveguide-based delay lines have to end up with to achieve similar delay responses [56]. Finally, the first phase term in Eq. 128 can be neglected, since we have a similar APF implemented in the lower arm of the MZI without the microdisk in its loop. In the off-resonance regime of the microdisk resonator, the responses from the upper and the lower racetrack resonators cancel with each other, and the MZI is balanced. In this sense, the filter is first-order because its response is mainly contributed by the microdisk resonator in the upper APF.

4.2.3 A tunable first-order filter: demonstration

The first-order filter shown in Fig. 31 is fabricated on an SOI wafer with a 230-nm-thick silicon device layer and a 1-μm-thick buried-oxide layer. The devices are patterned using a JBX-9300FS e-beam lithography system followed by plasma etching with a Cl_2 chemistry. Next, a 1-μm-thick flowable-oxide (FOx) layer is spin coated on top of the silicon devices, and metallic microheaters are fabricated on top of the FOx layer to implement the phase shifters using the thermo-optic effect [58]. We have to configure the phase shifters to generate desired filtering responses. For example, ϕ_{MZI} controls the bandwidth of the filter; ϕ_{disk} is to shift the resonance frequency of the microdisk and thus the center frequency of the filter; and whenever ϕ_{MZI} or ϕ_{disk} changes, ϕ_{FB} has to be adjusted according to Eq. 127. The APF at the lower arm of the MZI should follow the same operation as that of the upper APF for ϕ_{MZI} and ϕ_{FB} to keep a balanced MZI in the off-resonance regime of the microdisk resonator.

We show some configuration results of the output port P_2 (see Fig. 31) in Fig. 32, where a notch filter with a tunable bandwidth and center frequency has been realized. In Fig. 32(a), only one resonance appears in the spectrum and no racetrack modes are observed due to our balanced operation; other radial modes of the microdisk

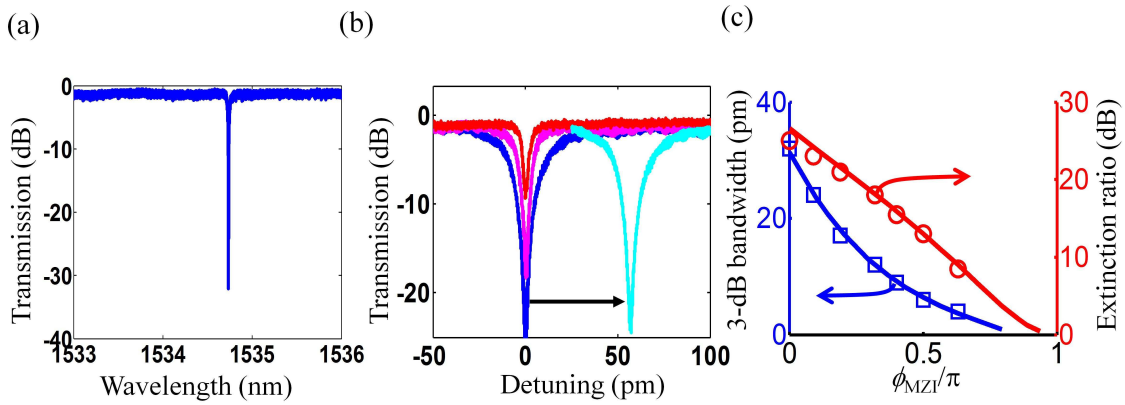

Figure 32: (a) Notch-filter responses of the P_2 port of the first-order filter shown in Fig. 31 measured around wavelength 1530 nm. (b) The 3-dB bandwidth has been tuned from 32 pm (blue solid line) to 10 pm (magenta dash-dotted line) and to 3.6 pm (black dotted line). The center frequency can also be tuned (blue to cyan solid lines). (c) Summary of the experimental data of the 3-dB bandwidths (squares) and extinction ratios (circles) under different coupling conditions controlled by ϕ_{MZI}. The solid curves are the fitted results based on Eqs. 123 and 124. Loss parameters and on-resonance delay offered by the microdisk resonator are extracted.

are also absent, suggesting the pulley coupling scheme has successfully achieved a single-mode operation. In Fig. 32(b), we show several representative resonances with different tuning parameters, whose 3-dB bandwidths are varied from 32 pm (4.1 GHz) to 10 pm (1.3 GHz) and to 3.6 pm (0.47 GHz). The resonance frequency of the filter can also be tuned. In addition, we observe that the extinction ratio of the resonance decreases as the filtering bandwidth becomes smaller, which is caused by the nonzero propagation loss from the racetrack as well as the microdisk resonator. Most of the losses come from the microdisk, since that is where the optical signal gets most of its delay. By fitting Eqs. 123 and 124 to the experimental data (assuming complex β_r and β_d), as shown in Fig. 32(c), the propagation losses of the racetrack waveguide and the microdisk resonators are extracted to be 5 dB/cm and 0.5 dB/cm, respectively. This reveals another advantage of using the microdisk-based delay lines instead of the waveguide-based ones (besides compactness): a reduction of the propagation loss by one order of magnitude under the same fabrication conditions for typical single-mode

ridge waveguides. In our case, the loss of the microdisk is still high, which limits the minimum bandwidth that can be attained (0.47 GHz here). On the other hand, the maximum bandwidth (4.1 GHz here) is limited by the on-resonance delay offered by the microdisk, which is inferred from Fig. 32(c) to be 160 ps. The deviation from the design (100 ps) is largely caused by the fabrication error in the access waveguide width (2 dB increase in coupling Q). The power consumption of each phase shifter is measured to be 20-30 mW per π phase shift, and we estimate the total power consumption of this first-order filter is less than 100 mW [59].

4.2.4 Power handing capability: nonlinear analysis

From the above example, it seems that the silicon microdisk resonator works well as a low-loss delay line. The implicit assumption is that the power is low enough so only linear responses need to be considered. However, silicon is a strong nonlinear material with a large χ^3 coefficient. The real part of χ^3 is responsible for the self-phase modulation and also cross-phase modulation if more than one frequencies are involved, and phenomena such as four-wave mixing result. On the other hand, the imaginary part of χ^3 is responsible for the two photon absorption (TPA), which refers to the excitation of electron-hole pairs through the absorption of two photons consecutively. As a result, silicon is no longer transparent at the telecommunication wavelength range (~ 1.55 μm), despite the fact that its bandgap (1.24 eV) is much larger than the single photon energy (~ 0.8 eV around 1.55 μm). In addition, the generated free carriers (i.e., electrons and holes) will also modify the refractive index of silicon and absorb the light. In typical silicon ridge waveguides (i.e., around 500×200 nm^2), such complicated nonlinear processes become important when the travelling power is more than 50 mW. Unfortunately, the power range of RF signals can be up to 100 mW. The situation is even worse for our resonator-based delay lines, since the power travelling inside the resonator is significantly enhanced from the resonant effect. For

example, in our microdisk-based delay line, the circulating power in the microdisk resonator is about 15 times larger than what is in the access waveguide. Therefore, as the power increases, the nonlinear effects will first take place in the delay line and distort its response.

We are interested in finding out the range of power that the microdisk-based delay line works properly. While an exact nonlinear analysis for an arbitrary input power is quite complicated, here we only need to know the point at which the delay response starts to deviate from its low-power behavior. To do that, we first calculate the power coupled into the resonator from the input light (assuming no nonlinear effects), then the propagation constant of the resonant mode β_d (see Eq. 124) is modified by the nonlinear process as [48, 58]

$$\beta_d = \beta_{d,\text{linear}} + k_0(\Delta n_{\text{FC}} + n_2 I_d) + i(\alpha_{\text{FC}} + \beta_2 I_d), \tag{129}$$

where n_2 (Kerr coefficient) and β_2 (TPA coefficient) correspond to the real and imaginary parts of χ^3, respectively; I_d is the intensity of the travelling light inside the resonator; and Δn_{FC} and α_{FC} are the refractive index change and absorption loss caused by the free carriers, respectively, whose values are empirically found as

$$\Delta n_{\text{FC}} = -(8.8 \times 10^{-4} N_e + 8.5 N_h^{0.8}) \times 10^{-18}, \tag{130}$$

$$\alpha_{\text{FC}} = (8.5 N_e + 6.0 N_h) \times 10^{-18}, \tag{131}$$

where N_e (N_h) is the electron (hole) density with the unit of cm^{-3} and α_{FC} has the unit of cm^{-1}. Finally, N_e ($=N_h$) is obtained through the following rate equation

$$\frac{dN_e}{dt} = -\frac{N_e}{\tau_{\text{FC}}} + \frac{\beta_2 I_d^2}{2\hbar\omega}, \tag{132}$$

with τ_{FC} being the free carrier lifetime.

The above equations are integrated in a numerical code. We study one narrowband filter which has a 3dB bandwidth of 1 GHz under different input powers, and the results are shown in Fig. 33, where the three columns correspond to the amplitude

and phase responses of the APF and the transmission of the P_2 port, respectively. As clearly seen there, when the input power is on the order of 1 mW, the results start to show deviation from the low-power case. In addition, Fig. 33(a) and (b) correspond to two different free carrier lifetimes, and the enhanced distortions observed in Fig. 33(b) suggest that the free carriers are the major source of nonlinearity. This can also be verified by the blue shift of the resonance frequency with the increased power level, which arises from Eq. 130. Also, Eq. 131 explains the increased insertion loss of the APF and the reduction of the extinction ratio of the P_2 port with the power.

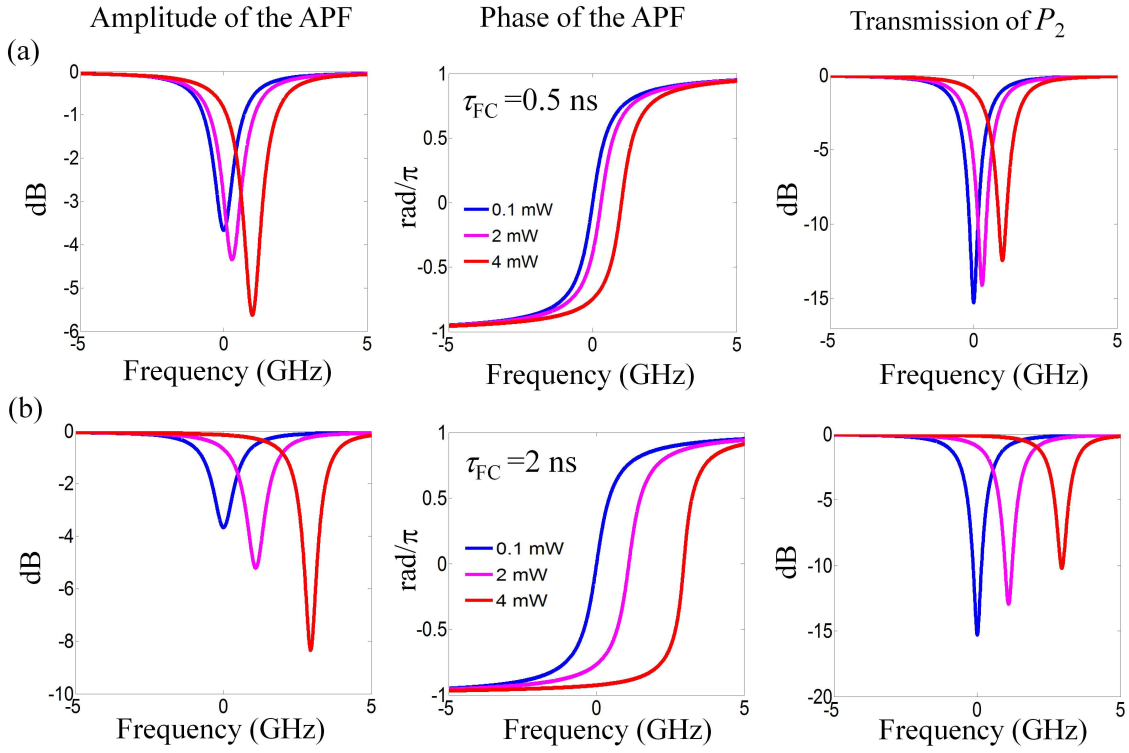

Figure 33: The three columns correspond to the amplitude and phase responses of the APF and the transmission of the P_2 port: (a) $\tau_{FC} = 0.5$ ns, and (b) $\tau_{FC} = 2$ ns.

4.2.5 Conclusions

In conclusion, we have proposed and demonstrated the use of over-coupled microdisk resonators as delay lines for narrowband optical filters. Compared to delay lines

based on single-mode ridge waveguides, they are more than two orders of magnitude smaller in size and at least one order of magnitude lower in insertion loss under the same fabrication conditions. A first-order tunable narrowband filter using the microdisk-based delay line is experimentally demonstrated in an SOI platform, which shows a tunable bandwidth from 4.1 GHz to 0.47 GHz with an overall size of 0.05 mm^2. More complex functionalities can be demonstrated using such a filter as a unit cell in a cascaded structure [60]. However, the current microdisk resonator suffers from two drawbacks: (1) first, the intrinsic Q of the silicon microdisk resonator is still low ($\sim$1.5 million), which limits the minimum bandwidth that can be achieved; (2) second, silicon is a strong nonlinear material, which limits the maximum power that can be sent into the filter (generally less than 1 mW). While the first drawback may be overcome by improved fabrication techniques, the second one is intrinsic to silicon and cannot be easily avoided in the SOI platform. In the next chapter, we will introduce a different material platform to reduce the propagation loss and to increase the power handling capability of such resonator-based delay lines.

CHAPTER V

VERTICAL INTEGRATION OF SILICON NITRIDE TO SOI PLATFORM

5.1 Introduction

As we have discussed in the previous chapter, though silicon (Si) is an excellent photonic material for various applications of integrated optics, it has its own challenges. For example, for certain passive devices such as low-loss delay lines, Si cannot compete with silicon dioxide (SiO_2) or silicon nitride (SiN) in terms of insertion loss and power handling capability for the following two reasons: (1) Si has a relatively large propagation loss (> 0.1 dB/cm) due to its large refractive index contrast and small mode volume [28], while in SiO_2 or SiN, the refractive index contrast is lower and the mode volume is larger, and consequently the propagation loss can be one order of magnitude smaller (< 0.01 dB/cm) [61]; (2) Si has strong nonlinear effects, primarily due to the free carriers generated via the two-photon absorption [48, 58]. This nonlinearity can strongly distort the optical signal at high power levels, while in SiO_2 or SiN, free carriers are absent and the corresponding nonlinear effects are much weaker. On the other hand, active elements such as modulators and phase shifters are difficult to realize in SiO_2 or SiN due to the difficulty in tuning the refractive index of the host material, while Si devices can be easily tuned using thermal or free-carrier-injection processes.

There are several groups which have explored SiN for a range of applications. Dr. Lipson's group (Cornell) has successfully grown thick SiN films up to 700 nm, and compact microresonators with radius down to 20 μm and intrinsic Qs around 3 million have been demonstrated [62], which have served as an enabling element for

applications such as optical frequency comb sources [63] and efficient second-harmonic generations [64]. Dr. Bower's group (UCSB), on the other hand, employs a very thin layer of SiN (less than 100 nm) as the guiding core, and very low-loss waveguides (with propagation loss less than 0.1 dB/m) have been achieved [61, 65]. High-Q microresonators are also reported based on the same platform, though the bending radii have to be on the order of millimeters to avoid significant radiation loss.

Here, we propose a material platform based on the monolithic integration of SiN on SOI to enable reconfiguration, high power handling and also low-loss performance. As illustrated by Fig. 34, SiN is vertically integrated on top of SOI with SiO_2 as the buffer. In this platform, active and reconfigurable devices (e.g., modulators and tunable phase shifters) as well as short-length devices (e.g., couplers) are formed in Si, and passive components which require either low loss or high-power tolerance are formed in the SiN layer (see Fig. 34(b)). To experimentally demonstrate such a hybrid material platform, the following steps have to be carried out: (1) high-quality SiN film growth; (2) demonstration of compact SiN microresonators with very high Qs; and (3) vertical integration of SiN into SOI with proper coupling between devices fabricated in these two different layers.

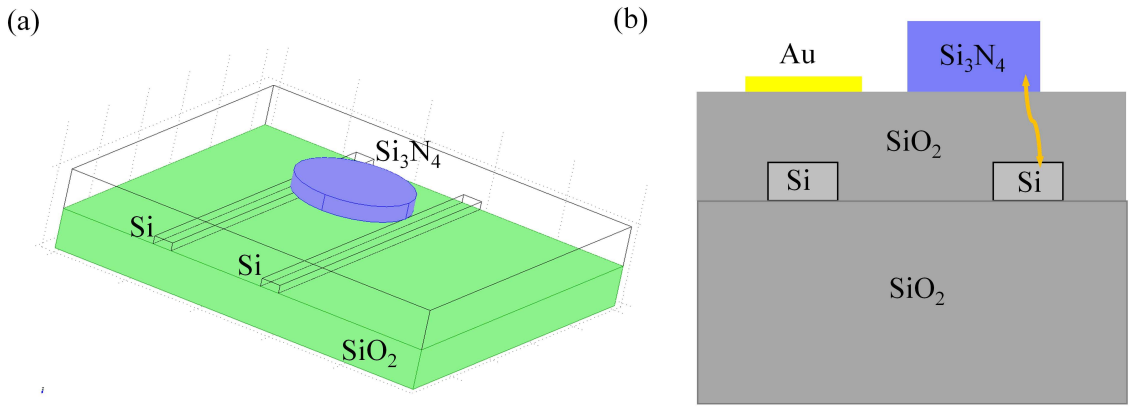

Figure 34: (a) Illustration of the vertical integration of SiN to the SOI platform. (b) Cross section of the SiN-on-SOI platform with metallic heaters implemented on top of the Si devices for reconfigurability.

5.2 SiN growth

For the dense integration purpose, the SiN film thickness has to be large enough to warrant compact microresonators and sharp waveguide bends. Numerical simulations show that a 400-nm-thick silicon nitride permits the realization of microresonators with a radiation Q more than a billion for radii as small as 40 μm. Low-pressure chemical vapor deposition (LPCVD) method is conventionally used for stoichiometric SiN (Si_3N_4) deposition at temperature around 800°C [66]. The standard source gases are dichlorosilane (SiH_2Cl_2, or DCS) and ammonia (NH_3), with the gas ratio being the dominant factor for the film stress and material absorption Low-stress SiN films, which are routinely used as mechanical membrane material and isolation/buffer layer in micromachining, can be grown using a large dichlorosilane to ammonia ratio ($3 \sim 6$) with thicknesses up to a few micrometers [66]. However, low-stress SiN films usually have a high hydrogen content, which is responsible for the strong N-H bond material absorption around the wavelength of 1.55 μm and is not desired for optical applications at the telecommunication wavelengths (N-H bond has a vibrational frequency around the wavelength of 3 μm and its second harmonic is around 1.55 μm [66]). Decreasing the dichlorosilane to ammonia ratio ($0.1 \sim 1$) can reduce the hydrogen content and therefore make the SiN film more transparent to light. The stress of the film, however, then increases significantly and the thickness of the SiN is typically limited to a few hundred nanometers. For our applications, it is clear that a careful balance between the film thickness and material absorption has to be found. Post annealing process has also to be developed to further reduce the hydrogen content. Table 2 lists the details of the recipe we have optimized for the SiN growth in the Tystar Nitride LPCVD tool available in Gatech Petit cleanroom facility.

Table 2: SiN growth recipe

Pressure(mT)	DCS(sccm)	NH_3 (sccm)	T (°C)	Deposition rate (nm/min)
165	50	140	800	4.15

To fabricate high-Q SiN microresonators, devices are first patterned using a JEOL JBX-9300FS electron-beam lithography (EBL) system. ZEP (ZEP520A by Zeon co-operation) is chosen as the e-beam resist, which is capable of defining fine features with a relatively good etch resistance. Here, HSQ is no longer is good mask because the selectivity we can get for SiN to HSQ is around 1, and HSQ is too thin (maximum 200 nm) to etch a 400-nm SiN film. One problem with SiN is that it is an insulating material, and electrons can accumulate at the SiN surface as the exposure progresses. This charge-up effect will somehow disturb the beam writing process, and under certain circumstances, it can become strong enough to cause fracture errors. Figure 35 shows two such examples: in Fig. 35(a) a line is missing and in Fig. 35(b) the circular boundary becomes jagged. To solve this problem, ESPACER, a conducting polymer developed by Shawa Denko K.K., is spin-coated on top of the ZEP before the e-beam exposure, which shall be removed by dipping the sample into DI water before the development of ZEP.

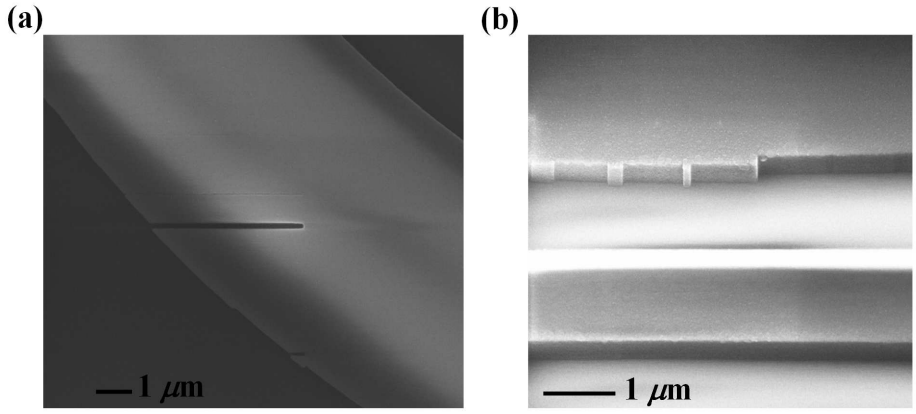

Figure 35: SEM of SiN microring resonators (8 μm width, after etching) with fracture errors: (a) a line is missing, and (b) the microring boundary becomes jagged.

Next, the pattern is transferred to the SiN layer using plasma etching with a

CF_4/CHF_3 gas mixture. For the etching process, there are three important parameters: etch rate, selectivity and anisotropicity. High selectivity is required in our process since the e-beam resist (ZEP) is only a few hundred nanometers thick. Roughly speaking, increasing the etching pressure will improve the selectivity at the cost of anisotropicity, since at a higher pressure, the etching is more likely to take place through chemical reaction rather than physical bombardment. Adding more CHF_3 to CF_4 will also increase the selectivity, but its amount should be carefully controlled to avoid a strong polymerization. We have optimized the etching recipe using the SiN-on-Si wafers, where the conducting Si substrate helps reduce the charge-up effect from SiN so good SEMs can be taken to examine the etching quality. Figure 36(a) shows the SEM of the cross section of a 400-nm-thick SiN waveguide, which shows fairly vertical sidewalls have been achieved. Figure 36(b) is the SEM for the corresponding sidewall, where the sidewall roughness is estimated to be less than 10 nm. Table 3 lists the details of the recipe we use for the SiN etch in an Oxford Endpoint RIE tool in Gatech Marcus Inorganic cleanroom facility.

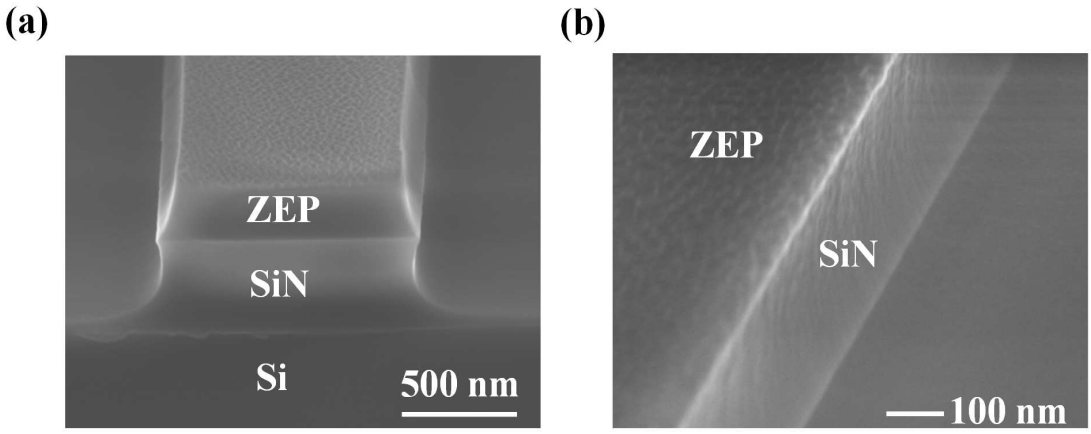

Figure 36: (a) SEM of the cross section of a SiN waveguide structure with a width of 1.2 μm and a height of 400 nm.(b) SEM of the sidewall of the waveguide shown in (a).

Table 3: SiN plasma etch recipe

Pressure(mT)	Gases(sccm)		Power (W)	DC bias(V)	Etch rate(nm/min)	
	CF_4	CHF_3			ZEP	SiN
250	50	5	175	215	22	43

The sidewall roughness can be further reduced using a post wet-etch process. Phosphoric acid (H_3PO_4) is used due to its selective etching of SiN to Si and SiO_2. After stripping the remaining ZEP,the SiN sample is put in 180°C 85% H_3PO_4 solution for 10 minutes, during which time around 26 nm SiN is etched away. Figures 37(a) and (b) show the cross section and sidewall of a wet-etched SiN waveguide, respectively. Comparing Fig. 37(b) to Fig. 36(b), we conclude that the sidewall roughness has been visibly improved by the post wet-etch process. Buffered oxide etcher (BOE) is also a good choice for the wet etching of SiN with a etch rate around 0.75 nm/min at room temperature.

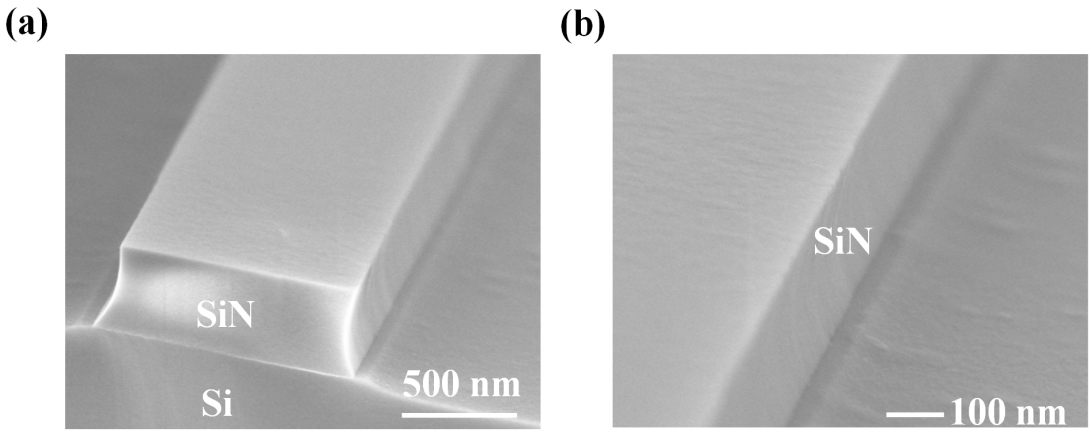

Figure 37: (a) SEM of the cross section of the SiN waveguide as shown in Fig. 36 after the post wet-etch process.(b) SEM of the sidewall of the waveguide shown in (a).

5.3 High-Q SiN microresonators

With the developed fabrication recipes, high-Q SiN microresonators are fabricated. One important parameter we want to know is their intrinsic Q. To do that, a series of microrings with different gaps to the access waveguides are fabricated so the critical coupling condition can be satisfied for one of the resonators. In addition, different annealing conditions have been tested to achieve the best results. Figure 38 shows the transmission measurements of a 60-μm-radius microring with a width of 8 μm and a height of 400 nm, with the oxide cladding on top. The access waveguide has a width of 1.2 μm. The gap between the SiN microresonator and the access waveguide is 700 nm. The blue curve shown in Fig. 38(a) depicts the transmission of the microring before the annealing process and the red curve depicts the transmission measurement of the same device after the annealing treatment. From the blue curve shown in Fig. 38(a), we observe that the resonance dips are much shallower for the wavelength range of $1490 - 1555$ nm compared to those outside of this range, indicating that the material absorption is strong at these wavelengths. This absorption spectrum is characteristic of the overtone absorption of the N-H bond [66]. After proper annealing at high temperatures (8 hours in an O_2 ambient and 4 hours in a N_2 ambient at 1100°C), the resonance dips become more uniform across the whole wavelength range (red curve), indicating that the hydrogen content has been significantly reduced. Figure 38(b) shows the zoom-in figure for one specific resonance around the wavelength of 1530 nm, which is marked by star in Fig. 38(a). As can be seen, before the annealing, the absorption-limited intrinsic Q is around 150,000, corresponding to a propagation loss around 2.35 dB/cm; after the annealing, the intrinsic Q dramatically increases to 6,000,000, corresponding to a propagation loss around 0.06 dB/cm (6 dB/m).

Next, we want to know what are the limiting factors of the intrinsic Q that we obtain. For SiN, there are two major sources of loss: the first one is the material absorption loss from the remaining N-H bonds after the annealing process, and the

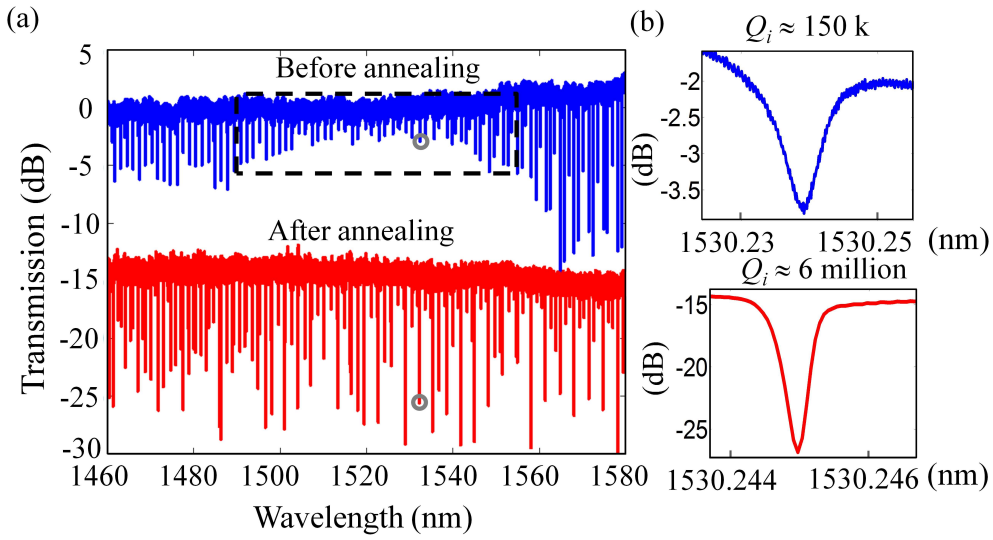

Figure 38: (a) Transmission measurements for a 60-μm-radius microring resonator fabricated with a 400-nm-thick SiN layer: the blue curve is before the annealing and the red curve is after the annealing (intentionally moved down by 15 dB for a better comparison). (b) Zoom-in figures for one resonance around 1530 nm: the upper figure is before the annealing and the lower figure is after the annealing.

second one is the scattering loss which results from the sidewall roughness caused by fabrication imperfections. There is an easy way to distinguish these two loss terms. As discussed in Chapter II, Q is defined as

$$Q \equiv \frac{\omega \cdot U_c}{P_{\text{loss}}}, \tag{133}$$

where U_c is the energy of the resonant mode which is proportional to the mode volume and P_{loss} is its power dissipation rate. We can exploit different scaling behaviors of these loss terms with the size of the resonator (i.e., R) to differentiate them. For example, the material absorption is a volume effect, which means as R increases, its value increases as R^2. On the other hand, the scattering loss is a surface effect, so it varies linearly with R. Since our microring is very thick, the first fundamental mode does not interact with the inner sidewall, and its mode volume scales with the radius R as R^2 [1]. Combing these results, we conclude that if the resonator is limited by the material absorption loss, its intrinsic Q will not change much with the radius, while

[1] If the microring is very thin, then the mode volume just scales with the perimeter.

if it is limited by the scattering loss, its intrinsic Q will linearly increase with the size of the resonator.

In Fig. 39(a), we summarize the statistically averaged intrinsic Qs for radii ranging from 40 μm to 240 μm. Note that although the example provided in Fig. 38(b) shows an intrinsic Q of 6 million, the ensemble averaged Q (i.e., average from many independently fabricated samples under the same condition) for 60-μm-radius microresonators is around 4 million (such a variation in the intrinsic Q has been explained in Ref. [49]). From the linear behavior of the intrinsic Q with the radius R, we conclude that for our resonators, the intrinsic Q is still limited by the sidewall scattering loss. In Fig. 39(b), the resonance for a 240-μm-radius microresonator is plotted, which shows an intrinsic Q around 16 million (equivalent to 2.2 dB/m propagation loss). Such a low-loss microresonator with a moderate size is an ideal candidate for delay lines that can be densely integrated.

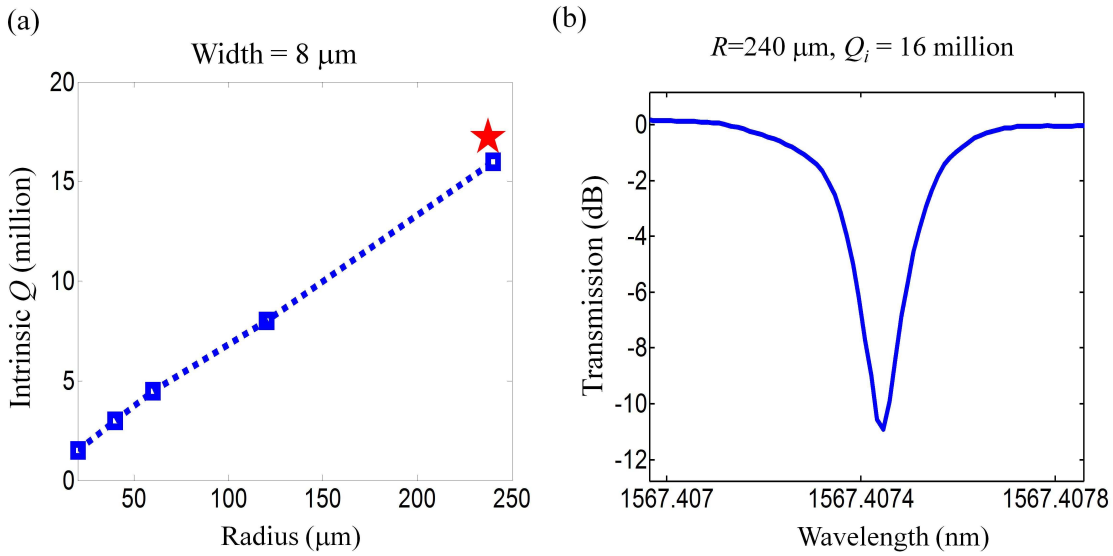

Figure 39: (a) Statistically averaged intrinsic Qs of SiN microresonators with different radii. The width of the microrings is fixed as 8 μm. (b) One resonance of a 240 μm-radius SiN microring showing an intrinsic Q of 16 million.

5.4 Vertical coupling between the SiN and SOI

As illustrated by Fig. 40(a), SiN is grown on top of SOI with the oxide as the buffer. The coupling between the SiN and the Si structures is achieved through evanescent field overlap, which can be controlled by adjusting the oxide thickness (i.e., the vertical gap). Since the fundamental mode of the 400-nm-thick SiN microresonators is TE polarized (electric field is predominantly parallel to the device layer), the Si waveguides have to be designed for the same polarization. Moreover, their propagation constants ought to be similar to achieve an efficient power exchange. For that purpose, we use a thin SOI with a thickness around 110 nm (instead of the conventional 220 nm SOI). Another benefit of using a thinner SOI is that the optical mode of the Si waveguide is less confined, leading to a larger field overlap with the SiN mode. In Fig. 40(b), we plot the power coupling coefficient between a typical Si waveguide (450×110 nm^2) and SiN waveguide (2200×400 nm^2), whose value falls into a wide range and could be fully controlled by the vertical gap.

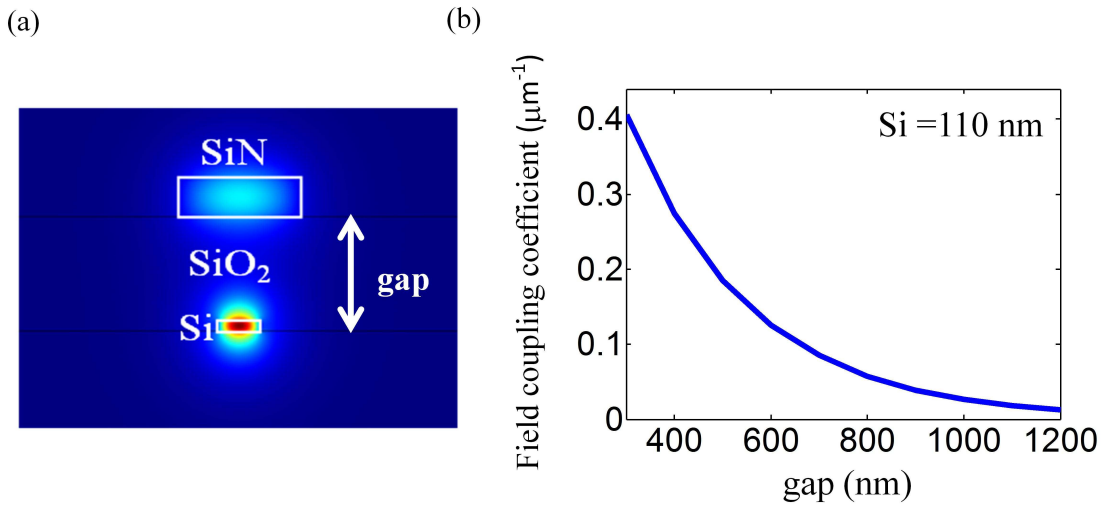

Figure 40: (a) Illustration of the vertical coupling between the SiN and the Si layers with SiO_2 as the buffer. (b) The power coupling coefficient between a typical Si waveguide (450×110 nm^2) and SiN waveguide (2200×400 nm^2) as a function of the vertical gap.

5.5 Integrating the SiN to the SOI

The fabrication procedures of integrating SiN to SOI are illustrated in Fig. 41 and explained below:

- Steps 1-4: starting with an SOI wafer with a 110-nm-thick Si layer and a 3-μm-thick buried oxide layer, the Si devices are fabricated using the HSQ process;
- Step 5: Flowable oxide (FOx) is spin coated, which serves as the buffer layer between the SiN and Si layers. As shown by the adjacent SEMs, the top surface of the FOx layer is almost flat, which makes the planarization unnecessary. In addition, after annealing in the oxygen ambient above 800°C, the FOx is converted to SiO_2 (see SEM in Step 6);
- Step 6: 400 nm SiN is deposited by LPCVD, and annealing is performed;
- Step 7: In order to align the SiN microresonators to the underneath Si waveguides, we have markers made in the Si layer. However, the top SiO_2 and SiN make them invisible under the SEM of the EBL system. To solve this issue, we use photolithography to open a window on top of these markers, and the SiN and SiO_2 are removed subsequently by a dry etching process. We also note from the corresponding optical micrograph there that a few cracking lines appear after the SiN annealing, which are acceptable as long as they have no overlap with the devices.
- Step 8: SiN microresonators are fabricated, which have to be aligned to the underneath Si waveguides using the markers in the Si layer. The adjacent optical micrographs show that a good alignment between the SiN microresonator and the underneath Si waveguide has been achieved.

The fabricated samples are characterized using the same setup discussed in Chapter II. As illustrated in Fig. 42(a), the light is firstly coupled into the Si waveguide

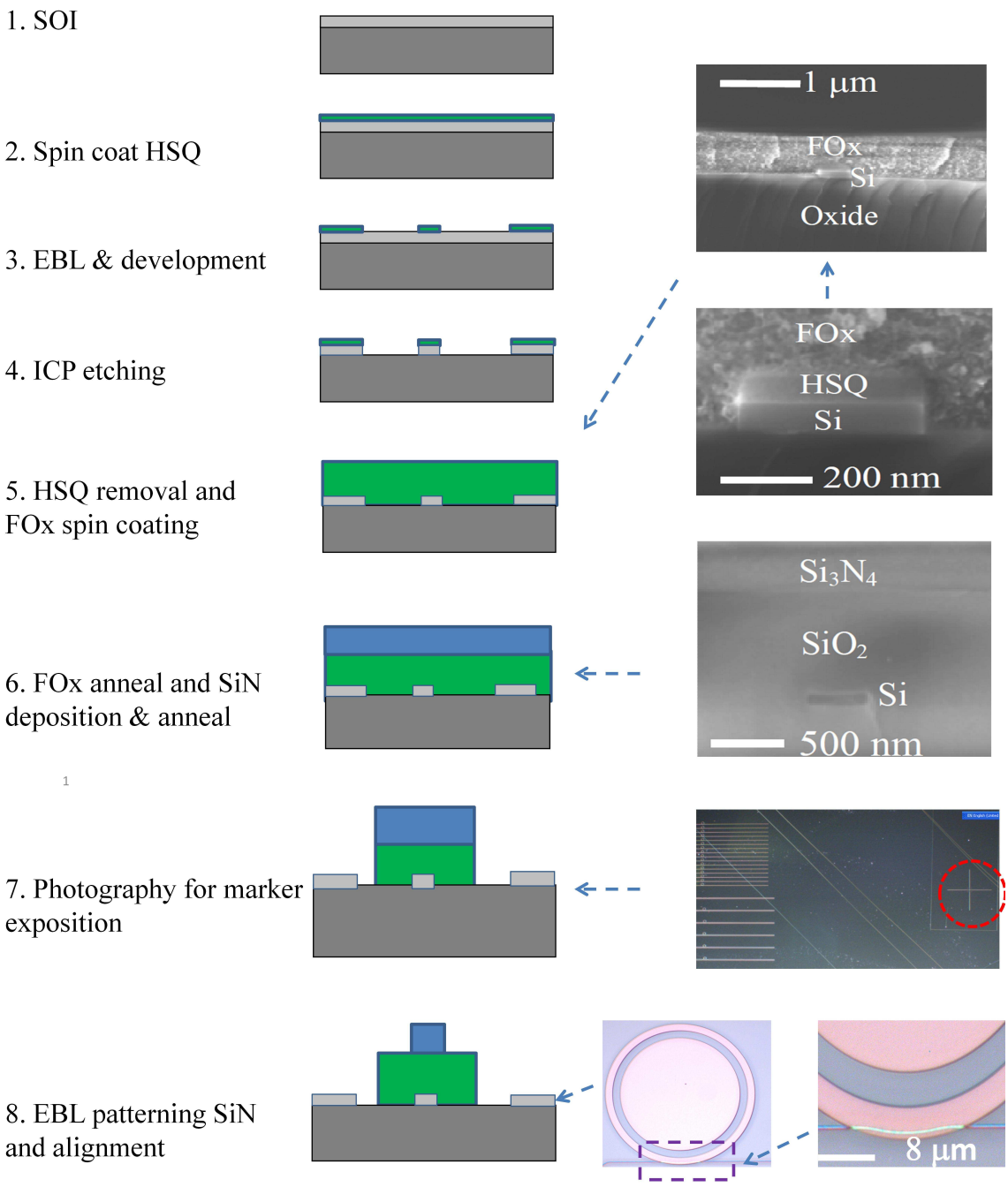

Figure 41: Fabrication procedures of integrating SiN to the SOI platform. Detailed description for each step is provided in the text.

input and then collected at the Si waveguide output. When the SiN microresonator is on resonance, strong scattering light can be observed there from the top infrared camera (see Fig. 42(a)), indicating a good coupling has been achieved between the Si and SiN layers. The preliminary characterization results are shown in Fig. 42(b), where an intrinsic Q around 2 million is measured for a 60-μm-radius SiN microresonator. Note that this Q is about half of what we can get from a SiN-on-SiO_2 structure. We believe the decrease of the intrinsic Q is mainly due to the degradation of the SiN material quality deposited on the nonuniform FOx surface (which is almost flat but not strictly so). In the future, a chemical-mechanical polishing should improve the surface quality of FOx.

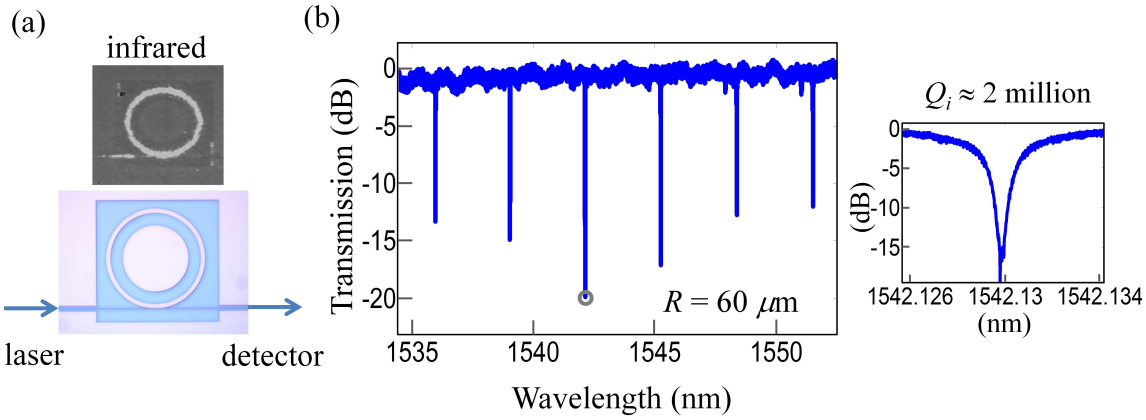

Figure 42: (a) Illustration of the characterization process for the SiN-on-SOI samples: light is coupled from the tunable laser to the Si waveguide input and then collected at the Si waveguide output before sent to the detector. The top image is captured by the infrared camera when the SiN microresonator is on resonance. (b) Characterization result of a 60-μm-radius SiN microresonator fabricated on top of the SOI, showing an intrinsic Q around 2 million.

5.6 Conclusions

We have proposed and demonstrated a new material platform by vertically integrating SiN to the SOI platform. High-Q SiN microresonators have been demonstrated; especially, microresonators with a 16 million intrinsic Q and a moderate size of 240

μm radius are realized. Our experimental results also show that a good coupling between the SiN and the Si layers can be achieved. In the future, by implementing microheaters on top of the Si devices, low-loss optical devices with reconfigurability and high power handling capability are expected to be demonstrated based on this platform.

CHAPTER VI

CONCLUSIONS AND FUTURE WORK

The theme of this thesis is to apply travelling-wave microresonators such as microdisks and microrings for on-chip optical interconnects and signal processing. Challenges arising from these applications including theoretical and experimental ones are addressed. On the theoretical aspect, we start with a study on the coupled mode theory (CMT), which is widely used in photonic circuit design to obtain the power coupling coefficient between coupled devices. A modified version of CMT is offered in this work for the TM-polarized light in high index contrast material systems. It does not only justify the better performance of the old CMT compared to the one incorporating the cross power overlapping factors, but also predicts better phase responses than all the existing CMTs. Moreover, numerical simulations show that for asymmetric couplers, it is essential to use the modified CMT to obtain the most accurate result.

The coupling-induced phase responses are also studied in this thesis, which is of critical importance to coupled-resonator structures since the phase shift will change the resonance frequency of the coupled resonator. A numerical approach based on the finite element method is developed to study such phase responses. Compared to the conventional approaches based on the finite-difference time domain method, our numerical code is very efficient in the simulation time (minutes versus hours or even days). The numerical study reveals that the phase responses from symmetric couplers which have the symmetry about the wave propagating direction can be simply estimated by a linear expression, while the ones from asymmetric couplers are quite complicated. Therefore, symmetric couplers are recommended to designers for coupled-resonator devices.

Mode splitting and scattering loss are two important features commonly observed in the spectrum of high-Q microresonators. Our review of the existing analytical approaches shows that they have only achieved partial success. Especially, different models have been proposed for several distinct regimes and cannot be reconciled. In this thesis, a unified approach is developed for the general case, and we show conditions under which our model can be reduced to the results of the existing approaches in their respective regimes. A numerical study on the two-scatterer example shows a good agreement between our model and the simulation. In contrast, the existing models have significant errors.

On the experimental aspect, we first develop a new fabrication recipe with a focus on the accurate dimensional control and low-loss performance. For coupled-resonator devices, it is essential to have the dimensional control on the order of nanometers so the fabrication-induced resonance frequency mismatch is negligible. For high-Q microresonators, sidewall roughness has to be minimized to reduce the propagation loss. In this work, HSQ is employed as the electron-beam resist, and the lithography and plasma etching steps are both optimized to achieve vertical and smooth sidewalls. This recipe has been widely used within the Photonics Research Group for many different applications.

A third-order temperature-insensitive coupled-resonator filter is designed and demonstrated in the silicon-on-insulator (SOI) platform, which serves as a critical building block element in terabit/s on-chip networks. Two design challenges are overcome. The first one, being broadband engineering, is addressed by using microring resonators with a small free spectral range instead of racetrack resonators; in addition, the coupling has been engineered so the flat-band response is preserved over a large wavelength range. The second challenge, being a temperature-insensitive design, is addressed by overlaying a polymer with a negative thermo-optic coefficient on top of the silicon waveguide. Moreover, through careful balancing between the dispersion

of the bandwidth and the thermal property of the filter, the redundant bandwidth of the filter channel caused by the dispersion can be employed as a thermal guard band. As a result, the filter can accommodate 21 WDM channels with a data rate up to 100 gigabit/s per wavelength channel, while providing a sufficient thermal guard band to tolerate more than $\pm 15^{o}$C temperature fluctuations in the on-chip environment.

In this thesis, high-Q microdisk resonators are proposed to be used as low-loss delay lines for narrowband filters. To selectively couple to one of the radial modes of the microdisk and also to achieve a strong coupling, pulley coupling scheme is used based the on the modified coupled mode theory. The width of the access waveguide is chosen for the phase matching to the third-order radial mode while the coupling length is optimized for the maximum suppression of other radial modes. A first-order tunable narrowband filter based on the microdisk-based delay line is experimentally demonstrated in an SOI platform, which shows a tunable bandwidth from 4.1 GHz to 0.47 GHz with an overall size of 0.05 mm^2. A few problems associated with silicon microdisk resonators are also identified. First, the intrinsic Q of the silicon microdisk resonator is still low ($\sim$ 1.5 million), which limits the minimum bandwidth that can be achieved. Second, silicon is a strong nonlinear material, which limits the maximum power that can be sent into the filter (generally less than 1 mW). While the first problem may be overcome by improved fabrication techniques, the second one is intrinsic to silicon and cannot be easily avoided in the SOI platform.

To address the challenges for the resonator-based delay lines encountered in the SOI platform, we propose to vertically integrate silicon nitride to the SOI platform, which can potentially have significantly lower propagation loss and higher power handling capability. A complete fabrication procedure, including the silicon nitride film growth, annealing, EBL patterning and plasma etching, is developed and optimized. High-Q silicon nitride microresonators are demonstrated; especially, microresonators with a 16 million intrinsic Q and a moderate size of 240 μm radius are realized, which

is one order of magnitude improvement compared to what can be achieved in the SOI platform using the same fabrication technology. We have also successfully grown silicon nitride on top of SOI and a good coupling has been demonstrated between the silicon nitride and the silicon layers.

In the end, we would like to discuss a few possible extensions of this thesis work in the future, which are listed below.

- For the third-order coupled resonator filter, we envision the addition of active switching functionality to the filter through carrier injection (either all-optically or electrically), and a high-performance switch with a switching speed on the order of nanoseconds is expected. In addition, higher-order filters can be used to achieve a better channel isolation and therefore a higher on-off contrast.

- We have demonstrated a first-order tunable filter using microdisk-based delay lines. In the future, more complex functionalities can be demonstrated using such a filter as a unit cell in a cascaded structure. In fact, a forth-order filter has been demonstrated in Ref. [60]. As the number of stages increases, it becomes very challenging to configure all the tuning components (i.e., microheaters) in a systematic way. Therefore, an efficient algorithm has to be developed for an automatic reconfiguration strategy. Moreover, to apply such a technology for real applications, a number of challenges have to be addressed. For example, the optical chip has to be packed in a way so the external laser can be easily plugged into the chip with a negligible insertion loss.

- For the silicon nitride on the SOI platform, all the necessary elements have been demonstrated. In the future, by implementing microheaters on top of the silicon devices, low-loss optical devices with reconfigurability and high power handling capability are expected to be demonstrated based on this platform.

REFERENCES

[1] http://en.wikipedia.org/wiki/Moore's law.

[2] http://newsroom.intel.com/docs/DOC-2032.

[3] M. Kistler, M. Perrone, and F. Petrini, "Cell multiprocessor communication network: built for speed," IEEE Micro **23**, 10 (2006).

[4] A. B. Miller, "Device requirements for optical interconnects to silicon chip," Proc. IEEE **97**, 1166 (2009).

[5] K. Bergman, "Silicon photonic on-chip optical interconnection networks," in "Frontiers in Nanophotonics and Plasmonics," (Guaruja, Brazil, 2007).

[6] A. Shacham, K. Bergman, and L. P. Carloni, "Photonic networks-on-chip for future generations of chip multiprocessors," IEEE Trans. Comput. **57**, 1246 (2008).

[7] W. Chang, *RF Photonic Technology in Optical Fiber Links* (Cambridge University Press, 2002).

[8] K. Poulton, "GHz ADCs: from exotic to mainstream," in "IEEE Custom Integrated Circuits Conference," (2010).

[9] B. Jalali, M. Paniccia, and G. Reed, "Silicon photonics," IEEE Microwave Magazine **17**, 58 (2008).

[10] L. Pavesi and D. J. Lockwood, *Silicon Photonics* (Springer-verlag, 2004).

[11] G. T. Reed and A. P. Knights, *Silicon Photonics: An Introduction* (John Wiley, 2004).

[12] A. W. Fang, H. Park, O. Cohen, R. Jones, M. J. Paniccia, and J. E. Bowers, "Electrically pumped hybrid AlGaInGs-silicon evanescent laser," Opt. Express **14**, 9203 (2006).

[13] A. Liu, R. Jones, L. Liao, D. Samara-Rubio, D. Rubin, O. Cohen, R. Nicolaescu, and M. Paniccia, "A high-speed silicon optical modulator based on a metal-oxide-semiconductor capacitor," Nature **427**, 615 (2004).

[14] Q. Xu, B. Schmidt, S. Pradhan, and M. Lipson, "Micrometer-scale silicon electro-optic modulator," Nature **435**, 325 (2005).

[15] L. Chen and M. Lipson, "Ultra-low capacitance and high speed germanium photodetectors on silicon," Opt. Express **17**, 7901 (2009).

[16] F. Xia, M. Rooks, L. Sekaric, and Y. Vlasov, "Ultra-compact high order ring resonator filters using submicron silicon photonic wires for on-chip optical interconnects," Opt. Express **15**, 11934 (2007).

[17] F. Xia, L. Sekaric, and Y. Vlasov, "Ultracompact optical buffers on a silicon chip," Nature Photon. **1**, 65 (2007).

[18] http://optics.org/news/1/2/19.

[19] K. J. Vahala, "Optical microcavities," Nature **424**, 6950 (2003).

[20] K. J. Vahala, *Optical Microcavities* (World Scientific, 2004).

[21] Q. Xu and M. Lipson, "All-optical logic based on silicon micro-ring resonators," Opt. Express **15**, 924 (2007).

[22] S. M. Spillane, T. J. Kippenberg, and K. Vahala, "Ultra-high-Q toroid microcavity on a chip," Nature **415**, 621 (2002).

[23] F. Vollmer, S. Arnold, and D. Keng, "Single virus detection from the reactive shift of a whispering-gallery mode," Proc. Natl. Acad. Sci. **105**, 20701 (2008).

[24] J. Zhu, S. K. Ozdemir, Y. F. Xiao, L. Li, L. He, D. R. Chen, and L. Yang, "On-chip single nanoparticle detection and sizing by mode splitting in an ultrahigh-Q microresonator," Nat. Photonics **4**, 46 (2010).

[25] T. J. Kippenberg, H. Rokhsari, T. Carmon, A. Scherer, and K. J. Vahala, "Analysis of radiation-pressure induced mechanical oscillation of an optical microcavity," Phys. Rev. Lett. **95**, 033901 (2005).

[26] T. Aoki, B. Dayan, E. Wilcut, W. Bowen, A. S. Parkins, T. J.Kippenberg, K. J. Vahala, and H. J.Kimble, "Observation of strong coupling between one atom and a monolithic microresonator," Nature **443**, 671 (2006).

[27] H. A. Haus, *Electromagnetic Fields and Energy* (Prentice Hall, 1989).

[28] M. Borselli, T. Johnson, and O. Painter, "Beyond the Rayleigh scattering limit in high-Q silicon microdisks: theory and experiment," Opt. Express **13**, 1515 (2005).

[29] A. Mazzei, S. Gotzinger, L. deS. Menezes, G. Zumofen, O. Benson, and V. Sandoghdar, "Controlled coupling of counterpropagating whispering-gallery modes by a single Rayleigh scatterer: a classical problem in a quantum optical light," Phys. Rev. Lett. **99**, 173603 (2007).

[30] J. Zhu, S. K. Ozdemir, L. He, and L. Yang, "Controlled manipulation of mode splitting in an optical microcavity by two Rayleigh scatterers," Opt. Express **18**, 23535 (2010).

[31] D. S. Weiss, V. Sandoghdar, J. Hare, V. Lefevre-Seguin, J. M. Raimond, and S. Haroche, "Splitting of high-Q Mie modes induced by light backscattering in silica microspheres," Opt. Lett. **20**, 1835 (1995).

[32] L. Deych and J. Rubin, "Rayleigh scattering of whispering gallery modes of microspheres due to a single dipole scatterer," Phys. Rev. A **80**, 061805(R) (2009).

[33] J. Wiersig, "Structure of whispering-gallery modes in optical microdisks perturbed by nanoparticles," Phys. Rev. A **84**, 063828 (2011).

[34] X. Yi, Y. F. Xiao, Y. C. Liu, B. B. Li, Y. L. Chen, Y. Li, and Q. Gong, "Multiple-Rayleigh-scatterer-induced mode splitting in a high-Q whispering-gallery-mode microresonator," Phys. Rev. A **83**, 023803 (2011).

[35] J. A. Kong, *Electromagnetic Wave Theory* (EMW Publishing, 2008).

[36] M. Soltani, "Novel integrated silicon nanophotonic structures using ultra-high Q resonators," Ph.D. thesis, Georgia Institute of Technology, Atlanta, GA (2009).

[37] M. Soltani, S. Yegnanarayanan, and A. Adibi, "Ultra-high Q planar silicon microdisk resonators for chip-scale silicon photonics," Opt. Express **15**, 4694 (2007).

[38] S. Chuang, "A coupled mode formulation by reciprocity and a variational principle," J. Lightwave Technol. **5**, 5 (1987).

[39] H. Haus, W. Huang, and A. Snyder, "Coupled-mode formulations," Opt. Lett. **14**, 1222 (1989).

[40] A. Snyder, A. Ankiewicz, and A. Altintas, "Fundamental error of recent coupled mode formulations," Electron. Lett. (1987).

[41] Http://www.comsol.com.

[42] Q. Li, S. Yegnanarayanan, A. Atabaki, and A. Adibi, "Quantitative modeling of coupling-induced resonance frequency shift in microring resonators," Opt. Express **17**, 23474 (2009).

[43] M. Popovic, C. Manolatou, and M. Watts, "Coupling-induced resonance frequency shifts in coupled dielectric multi-cavity filters," Opt. Express **14**, 1222 (2006).

[44] F. L. Teixeira and W. C. Chew, "Systematic derivation of anisotropic PML absorbing media in cylindrical and spherical coordinates," IEEE Microw. Guided W. **7**, 285 (1997).

[45] Y. Tsuji and M. Koshiba, "Finite element method using port truncation by perfectly matched layer boundary conditions for optical waveguide discontinuity problems," J. Lightwave Technol. **20**, 463 (2002).

[46] A. Taflove and S. Hagness, *Computational Electrodynamics: the Finite-Difference Time-Domain Method* (Artech House Publishers, 2005).

[47] J. D. Jackson, *Classical Electrodynamics* (Wiley, 1999).

[48] R. W. Boyd, *Nonlinear Optics* (Academic Press, 2003).

[49] Q. Li, A. Eftekhar, Z. Xia, and A. Adibi, "A unified approach to mode splitting and scattering loss in high-Q whispering-gallery-mode microresonators," arXiv:http://arxiv.org/pdf/1205.4448v1.pdf (2012).

[50] B. Lee, A. Biberman, P. Dong, M. Lipson, and K. Bergman, "All-Optical comb switch for multiwavelength message routing in silicon photonic networks," IEEE Photon. Technol. Lett. **20**, 767 (2008).

[51] Q. Li, M. Soltani, S. Yegnanarayanan, and A. Adibi, "Design and demonstration of compact, wide bandwidth coupled-resonator filters on a silicon-on-insulator platform," Opt. Express **17**, 2247 (2009).

[52] J. Poon, J. Scheuer, S. Mookherjea, G. Paloczi, Y. Huang, and A. Yariv, "Matrix analysis of microring coupled-resonator optical waveguides," Opt. Express **12**, 90 (2004).

[53] Q. Li, S. Yegnanarayanan, M. Soltani, P. Alipour, and A. Adibi, "A temperature-insensitive third-order coupled resonator filter for on-chip terabit/s optical interconnects," IEEE Photon. Technol. Lett. **22**, 1768 (2010).

[54] C. Madsen and J. Zhao, *Optical Filter Design and Analysis* (John Wiley, 1999).

[55] Y. Vlasov and S. McNab, "Losses in single-mode silicon-on-insulator strip waveguides and bends," Opt. Express **12**, 1622 (2004).

[56] P. Dong, W. Qian, S. Liao, H. Liang, C.-C. Kung, N.-N. Feng, R. Shafiiha, J. Fong, D. Feng, A. Krishnamoorthy, and M. Asghari, "Low loss shallow-ridge silicon waveguides," Opt. Express **18**, 14474 (2010).

[57] E. S. Hosseini, S. Yegnanarayanan, A. H. Atabaki, M. Soltani, and A. Adibi, "Systematic design and fabrication of high-Q single-mode pulley-coupled planar silicon nitride microdisk resonators at visible wavelengths," Opt. Express **18**, 2127 (2010).

[58] Q. Lin, O. Painter, and G. Agrawal, "Nonlinear optical phenomena in silicon waveguides: Modeling and applications," Opt. Express **15**, 16604 (2007).

[59] Q. Li, A. Eftekhar, P. Alipour, A. Atabaki, S. Yegnanarayanan, and A. Adibi, "Low-loss microdisk-based delay lines for narrowband optical filters," IEEE Photon. Technol. Lett. **24**, 1276 (2012).

[60] P. Alipour, A. Eftekhar, A. Atabaki, Q. Li, S. Yegnanarayanan, C. Madsen, and A. Adibi, "Fully reconfigurable compact RF photonic filters using high-Q silicon microdisk resonators," Opt. Express **19**, 15899 (2011).

[61] J. Bauters, M. Heck, D. John, D. Dai, M. Tien, J. Barton, A. Leinse, R. Heideman, D. Blumenthal, and J. Bowers, "Ultra-low-loss high-aspect-ratio Si_3N_4 waveguides," Opt. Express **19**, 3163 (2011).

[62] A. Gondarenko, J. Levy, and M. Lipson, "High confinement micron-scale silicon nitride high Q ring resonator," Opt. Express **17**, 11366 (2009).

[63] M. Foster, J. Levy, O. Kuzucu, K. Saha, M. Lipson, and A. Gaeta, "Silicon-based monolithic optical frequency comb source," Opt. Express **19**, 14233 (2011).

[64] J. Levy, M. A. Foster, A. Gaeta, and M. Lipson, "Harmonic generation in silicon nitride ring resonators," Opt. Express **19**, 11415 (2011).

[65] M. Tien, J. Bauters, M. Heck, D. Blumenthal, and J. Bowers, "Ultra-low loss Si_3N_4 waveguides with low nonlinearity and high power handling capability," Opt. Express **18**, 23562 (2010).

[66] F. Habraken, *LPCVD Silicon Nitride and Oxynitride Films: Material and Applications in Integrated Circuit Technology* (Springer, 1991).

[67] L. Rade and B. Westergren, *Mathematics Handbook for Science and Engineering* (Springer, 1999).

[68] G. N. Waston, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, 1995).

[69] A. B. Matsko and V. S. Ilchenko, "Optical resonators with whispering-gallery modes-part I: basics," IEEE J. Sel. Top. Quantum Electron. **12**, 3 (2006).

[70] Q. Li, A. A. Eftekhar, Z. Xia, and A. Adibi, "Azimuthal-order variations of surface-roughness-induced mode splitting and scattering loss in high-Q microdisk resonators," Opt. Lett. **37**, 1586 (2012).

[71] M. Borselli, T. Johnson, and O. Painter, "Accurate measurement of scattering and absorption loss in microphotonic devices," Opt. Lett. **32**, 2954 (2007).

[72] T. Barwicz and H. A. Haus, "Three-dimensional analysis of scattering losses due to sidewall roughness in microphotonic waveguides," J. Light. Technol. **23**, 2719 (2005).

[73] http://en.wikipedia.org/wiki/Bessel function.

[74] http://www.mathworks.com.