UNIVERSITY OF CALIFORNIA
Los Angeles

Silicon Photonics
for chemical sensing and spectroscopy,
diagnosis and therapy

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Electrical Engineering

by

Kam Yan Hon

2012
Silicon Photonics is a research area that utilizes silicon as the optical medium. It has been attracting a lot of research interests in past few years. Due to the mature silicon fabrication industry, Si is a low cost and well understood material. The tight optical confinement in Si also makes it a promising candidate for high speed on-chip interconnects. Additionally, the high optical confinement in Si further makes it an important platform for active nonlinear optical devices. However, almost all literature results are demonstrated in the optical communication window. We believed that optical communication is not the only area where silicon photonics will have an impact. My research is to explore the new capabilities of silicon. In this dissertation, a number of new applications of silicon photonics, ranging from chemical sensing and spectroscopy to diagnosis and therapy, will be introduced.
In the first part, I will present my studies on extending the working window of silicon from near-infrared to longer wavelength regime. Silicon has excellent linear and nonlinear optical properties in the midwave and longwave infrared spectrum and these properties open up the possibility for a new class of infrared photonic devices. Using silicon as a material platform, applications ranging from gas detection and sensing to free space communications can be realized using low cost, chip-level device integration. To transfer the knowledge from the near infrared and apply it to this new regime, detailed understanding of the material properties is essential. For silicon-based devices, it is known that silicon lacks the second-order optical nonlinearity due to the centrosymmetric atomic configuration. Thus, the lowest-order nonlinearity – third-order susceptibility $\chi^{(3)}$, which gives rise to the Kerr and Raman effects, is key to nonlinear optical effects in silicon. Detailed knowledge is needed because these coefficients may have strong wavelength dependences over the midwave region but, unfortunately, there are spectral “gaps” over which knowledge of the coefficients is missing. (i) To assist researchers, we have compiled the third-order nonlinear optical (NLO) data and predicted or fill in the knowledge gaps using semiconductor nonlinear optical theory using the results from the scientific literature. (ii) Additionally, I will introduce a new class of photonic devices based on periodic stress fields in silicon that enables second-order nonlinearity and achieves quasi-phase matching in silicon simultaneously – periodically-poled silicon (PePSi). This adds the periodic poling technology to silicon photonics and allows the excellent crystal quality and advanced manufacturing capability of silicon to be harnessed for devices based on the linear electro-optic effect and other second-order nonlinear effects. As an example of utility of the PePSi technology, efficient mid-wave infrared generation can be realized though difference frequency generation for a wide range of applications such as gas sensing and
spectroscopy. (iii) To extend further into terahertz regime, I have studied and used large area silicon PIN diodes as terahertz emitters. By engineering the back-side reflection of the diode, quasi-continuous wave and enhanced external emission efficiency is demonstrated using impulsive excitations.

In the second part of this thesis, I will present a new type of tunable dispersive device for single-shot, time-wavelength spectroscopy. This new device overcomes the limitations of operational bandwidth, total dispersion and large spatial footprint by leveraging the large modal dispersion of a multimode waveguide in combination with the angular dispersion of diffraction gratings to create chromatic dispersion. Both multimode silica fiber and silicon waveguide are studied for implementing this new concept. Single-shot, time-wavelength spectroscopy is demonstrated in 800 nm (a window that is good for biological applications) using a multimode silica fiber. In addition, I have characterized the devices’ dispersion on silicon waveguides and obtained same amount of dispersion but using a chip scale implementation. This potentially enables on-chip, high speed, time-wavelength mapping devices. This small footprint, high dispersive device can also be used for applications such as high speed time-stretch analog-to-digital conversion and optical imaging.

In the third part, the application of using porous silicon nanoparticles (PSiNPs) for in vivo cancer diagnosis and therapy will be presented. PSiNPs are attractive carriers for targeted drug delivery in nanomedicine. For in vivo applications, the biodegradation property of PSiNPs provides a pathway for their safe clearance from the body. Particles sizes of 80 – 120 nm are of particular interest, however, the biodegradability rate of such particles is often too fast, which limits particles’ in vivo half-life and potentially reduces their delivery efficiency. In this part, I
focus on the degradation of nano-scale particles and study the effect of both thermal oxidation and silica coating on the stability of PSiNPs in phosphate buffered saline solution (a close mimic of a basic biological fluid). The particles produced using both these techniques can be further functionalized using standard surface chemistries suitable for applications in drug delivery or flow cytometry.
The dissertation of Kam Yan Hon is approved.

Jin Hyung Lee
Aydogan Ozcan
Fuyuhiko Tamanoi
Bahram Jalali, Committee Chair

University of California, Los Angeles

2012
TABLE OF CONTENTS

1. Introduction 1
    1.1. Third-order nonlinearity of silicon from midwave to longwave infrared 2
    1.2. Second-order nonlinearity of stressed silicon 3
    1.3. Terahertz generation using large area silicon PIN diodes 3
    1.4. Giant tunable silicon based dispersive element for on chip real time frequency
time mapping 4
    1.5. Study of porous silicon nanoparticles for in vivo cancer diagnosis and therapy 6
    REFERENCES 7

2. The third-order nonlinear optical coefficients of Si, Ge and SiGe in the midwave
and longwave infrared 8
    2.1. Introduction 9
    2.2. Real part of the third-order nonlinear susceptibility $\chi^{(3)}$, 12
    2.3. Two-photon absorption coefficient $\beta_{TPA}$ 19
    2.4. Raman gain coefficient $g_R$ 23
    2.5. Conclusion 27
    REFERENCES 28

3. Periodically-poled silicon (PePSi) – Second-order nonlinearity of stressed silicon 33
    3.1. Mid-wave infrared generation using PePSi waveguide 34
    3.2. Discrepancy between theoretical and experimental $\chi^{(2)}$ values 40
    3.3. Experimental results and further work 43
    REFERENCES 45
4. Silicon PIN diode based quasi-continuous wave terahertz emitter

REFERENCES

5. Giant tunable silicon based dispersive element for on chip real-time frequency-time mapping

5.1. Chromo-modal dispersion

5.2. Demonstration on a multimode fiber

5.3. Discussion

5.4. Methods for multimode fiber

5.5. Preliminary demonstration on a Si slab waveguide

5.6. Conclusion

REFERENCES

6. Tailoring the biodegradability of porous silicon nanoparticles

6.1. Introduction

6.2. Fabrication and characterization of porous silicon nanoparticles (PSiNPs)

6.2.1. Thermally oxidized PSiNPs (TOPSiNPs)

6.2.2. Silica-coated PSiNPs (SCPSiNPs)

6.2.3. Particle characterization

6.2.4. Degradation study in simulated biological conditions

6.3. Results and discussions

6.3.1. Effect of thermal oxidation on the degradation of TOPSiNPs

6.3.1.1. Effect of Rapid Thermal Processing (RTP)

6.3.1.2. Effect of 60°C post-oxidation process

6.3.1.3. Optimized process of TOPSiNPs

REFERENCES
6.3.2. Effect of silica coating on the degradation of SCPSiNPs 86

6.3.3. Further improvement 87

6.3. Summary and conclusion 88

REFERENCES 89
LIST OF FIGURES

2.1. (a) Experimental $\chi^{(3)}_{1111}$ of Si in the literature as a function of wavelength. Data points are determined from Refs. 14-21 (circles), Ref. 22 (upward triangles), Ref. 23 (downward triangles) and Ref. 27 (diamonds). Theoretical dispersion curves discussed here are fitted to Ref. 22 (Bristow et al.) and to Ref. 23 (Lin et al.). Inset lists the $\chi^{(3)}_{1111}$ and $n_2$ at 1.55 $\mu$m (telecommunication wavelength), 2.2 $\mu$m (TPA cutoff), 3.39 $\mu$m (HeNe laser) and 4.26 $\mu$m (a CO$_2$ absorption band). A resulting average dispersion curve is shown which act as a reference for designers dealing with Si (b) Experimental $\chi^{(3)}_{1111}$ of Ge in the literature as a function of wavelength. Ge data points are determined from Refs. 30-33 (circles) and Ref. 34 (diamonds). Two theoretical dispersive curves are used to set the upper and lower bound of the experimental data points. A resulting average dispersion curve is shown which act as a reference for designers dealing with Ge. Inset lists the $\chi^{(3)}_{1111}$ and $n_2$ at 3.17 $\mu$m (TPA cutoff), 3.39 $\mu$m (HeNe laser), and 4.26 $\mu$m (a CO$_2$ absorption band).

2.2. (a) Calculated bandgap energy of $E_{\Gamma_3b}$, $E_{\Gamma L}$ and $E_g$ of Si$_{1-x}$Ge$_x$ alloy with respect to the atomic fraction of Ge. (b) Calculated $\chi^{(3)}_{1111}$ of Si$_{1-x}$Ge$_x$ alloy at the 6 $\mu$m wavelength with respect to the atomic fraction of Ge using the direct and indirect bandgap models.

2.3. Theoretical $\chi^{(3)}_{1111}$ dispersion curves of Si$_{1-x}$Ge$_x$ alloy with $x = 0.2$, 0.4, 0.8 and 0.9.

2.4. (a) Experimental $\beta_{TPA}$ of Si in the literature as a function of wavelength. Data points are determined from Refs. 14-16, 20, 39-41 (circles), Ref. 22 (upward triangles), Ref.
23 (downward triangles), Ref. 27 (diamonds) and Ref. 43 (squares). One theoretical dispersion curve is fitted to Ref. 22 (Bristow et al.) and another to Ref. 23 (Lin et al.).

(b) Experimental $\beta_{TPA}$ of Ge in the literature as a function of wavelength. Data points are determined from Ref. 44 (diamond), Ref. 45 (triangles) and Ref. 47 (circles). A theoretical dispersion curve is fitted to the data points.

2.5. Theoretical $\beta_{TPA}$ dispersion curves of Si$_{1-x}$Ge$_x$ alloy with (a) $x = 0.2$, 0.4 and 0.6 using the indirect bandgap model and (b) $x = 0.8$, 0.9, 0.95 and 1 using the direct bandgap model.

2.6. Calculated TPA cut-off wavelength of Si$_{1-x}$Ge$_x$ alloy with respect to the atomic fraction of Ge.

2.7. Experimental $g_R$ of Si in literature as a function of the Stoke’s wavelength. Data points are determined from Refs. 40, 51-55 (circles) and Ref. 10 (diamonds). A theoretical dispersion curve is fitted to the data points for Si. An estimated dispersion curve for Ge is shown. (b) The estimated $g_R$ of Si$_{1-x}$Ge$_x$ alloy at the 6 $\mu$m Stoke’s wavelength with respect to the atomic fraction of Ge.

3.1. An example of PePSi waveguide formed by covering a silicon channel waveguide with two types of silicon nitride (SiN) stressed films periodically along the waveguide: one induces compressive stress, the other one produces tensile stress.

3.2. Simulated cross-sectional $y$-component normal stress ($\sigma_{yy}$) distribution where the PePSi waveguide is covered by silicon nitride films with 1 GPa (tensile) stress and (b) –500 MPa (compressive) stress. (c) Simulated average $\sigma_{yy}$ and estimated second order susceptibility ($\chi^{(2)}$) induced along the waveguide (z-direction).
3.3. (a) Calculated output spectra in a 2-cm-long PePSi waveguide (Period \( \sim 8 \ \mu m \)): a 12-ps pump pulse at 1.3 \( \mu m \) (input peak intensity = 1.5 GW/cm\(^2\)), a 12-ps signal pulse at 1.75 \( \mu m \) (input peak intensity = 12.5 MW/cm\(^2\)), idler at 5.1 \( \mu m \). (b) Calculated conversion efficiency of MWIR generation.

3.4. Schematic of a 2D centrosymmetric silicon lattice structure under inhomogeneous compressive stress.

3.5. (a) Simulated cross-sectional \( x \)-component normal strain \( (S_{xx}) \) distribution. (The channel waveguide is covered by silicon nitride film with 1 GPa compressive stress) (b) \( S_{xx} \) along \( y \) at \( x = 0 \ \mu m \)

4.1. Experimental setup for autocorrelation measurements of terahertz emission from a large area silicon PIN diode under double optical pulse excitation. Inset 1 illustrates the enhanced external power extraction and quasi-continuous wave emission from the diode by utilizing the internal reflection from the backside substrate. Inset 2 illustrates the selection of particular terahertz frequency through double optical pulse excitation.

4.2. (a) Efficiency of the terahertz generation (b) Terahertz spectra with 40 fs single pulse optical excitations. Measured spectrum shows a narrow-band emission with power concentrated on certain frequency due to backside reflection as compare with the conventional case illustrated by dotted line. The illustration is drawn assuming same terahertz power is generated in both cases (c) Terahertz spectra with 40 fs and 350 fs optical excitations. The figure shows the dependence of terahertz power on the optical pulse widths. This suggests the advantage of higher conversion efficiency using impulsive excitation rather than conventional photomixing.

4.3. The autocorrelation traces of the terahertz signal with double optical pulse excitation.
4.4. Terahertz power spectra with double optical pulse excitation at various pulse repetition periods. The graphs show the enhancement of particular terahertz peaks by double optical pulse excitations.

5.1. Schematic diagram of the chromo-modal dispersion (CMD) device. Light is incident upon two parallel plane gratings, which disperse and collimate the optical spectrum. Angular dispersion is then applied to the spectrum using a lens. The input facet of a multimode fiber is placed at the lens focus such that the various spectral components are coupled into different fiber modes. The figure inset illustrates how different spectral components are coupled into and propagate in the multimode fiber. The dashed line represents the optic axis of the fiber. Although the configuration shown provides anomalous dispersion, the CMD can be tuned throughout both the anomalous and normal dispersive regimes by adjusting the alignment of the lens and fiber relative to the spatially dispersed spectrum.

5.2. (a) Relative delay and (b) dispersion parameter for three different alignments of the CMD. Translating the lens and fiber facet in the horizontal direction relative to the incident spatially dispersed spectrum shifts the CMD’s zero dispersion wavelength.

5.3. Mode profiles at the output of the CMD fiber for alignment position #1, measured 5 mm from the output fiber facet. (a) Mode profiles recorded using various excitation wavelengths. (b) Mode profile using 745-nm excitation after propagating through 20 m of 200-μm core diameter fiber. (c) Mode profile using 745-nm excitation after propagating through 35 m of the same fiber. Blurring in the output mode profile indicates the coupling length has been reached. The scale bar in all cases is 1 mm.
5.4. Time-domain profile and spectrum of an 80-fs optical pulse after propagation through the CMD. (a) Spectrum of the input optical pulse and the CMD dispersion parameter. (b) Measured time-domain output (red) and calculated time-domain profile for the measured input spectrum and dispersion curve.

5.5. Single-shot atomic absorption spectroscopy of rubidium vapor. (a) Optical spectrum after passing though the rubidium cell and the CMD, as measured by a conventional spectrometer. The arrow represents the position of the Doppler-broadened D2-absorption line. (b) Sequence of four consecutive single-shot time-domain traces captured by a photodiode and oscilloscope at a repetition rate of 90.8 MHz. High-frequency amplitude noise in the time-domain signals is a result of ringing in the measurement, and is not fundamental to the time-wavelength spectroscopy technique. The horizontal time axis can be converted to wavelength with knowledge of the CMD dispersion curve. Peaks “1” and “2” are denoted in both (a) and (b) to illustrate correspondence in the time-wavelength mapping.

5.6. Illustration of the different spectral components coupled into and propagate in the Si slab waveguide. This configuration represents normal dispersion. The input facet is polished such that the shortest wavelength component (blue) travels and is confined by total internal reflection at an angle close to the critical angle.

5.7. Schematic of the multimode waveguide made by double side polished wafer. The length is 9 cm and the thickness is 400 μm. The inset shows the polished surface of the input facet.

5.8. Relative delay of the Si-based CMD measured by modulation phase detection (red). Relative delay of the Si-based CMD calculated based on experiment parameters.
5.9. Illustration of an on-chip silicon based CMD device. The wavelength components are spatially dispersed by the fan-shaped reflector grating.

6.1. Fabrication process of the thermal oxidized porous silicon nanoparticles (TOPSiNPs).
Pt: platinum, HF: hydrofluoric acid.

6.2. Representative TEM images of a PSiNP (a) without any thermal oxidization (b) oxidized using RTP for 90s but no post-oxidation (c) Degradation kinetics of TOPSiNPs with various RTP times (no post-oxidation included) showing that the blood retention time can be extended by increasing RTP duration. The inset shows the magnified plot of degradation kinetics from 0 to 200 minutes. The percentage of Si contents released into PBS medium at 37°C is measured using ICP-OES. The half-life indicates the amount of time the percentage of Si content of the solvent increases to 50%. TEM: transmission electron microscopy, RTP: rapid thermal processing, PBS: phosphate buffered saline, ICP-OES: inductively coupled plasma optical emission spectrometry.

6.3 (a) Representative TEM images of TOPSiNPs before and after post-oxidation in a 60°C water bath for 18 hrs. (PSiNPs are oxidized using RTP for 90s) (b) Degradation kinetics of TOPSiNPs with various post-oxidation times (90s RTP is included) showing that the blood retention time can be extended by increasing post-oxidation duration until the optimized post-oxidation treatment of 18 hours.

6.4. Degradation kinetics of TOPSiNPs with various RTP times (post-oxidation at 60°C for 18 hours is included) showing that the optimized thermal oxidation treatment by varying the RTP duration with optimized post-oxidation duration. Using thermal oxidation, the half-life of PSiNPs can be varied from 10 minutes to 3 hours.
6.5 (a) Representative TEM images of PSiNPs with no silica coating, 10 minutes of silica coating and 30 minutes of silica coating. (b) Degradation kinetics of silica coated PSiNPs (SCPSiNPs) with various silica coating times (no thermally oxidation is included) showing that the retention time can be extended by increasing the silica coating time. Using silica coating, we can extend the half-life up to 8 hours.
## LIST OF TABLES

3.1. Comparison between theoretical and experimental results 43
ACKNOWLEDGEMENTS

First, I would like to express my deep appreciation to my advisor, *Prof. Bahram Jalali*, for his countless support and valuable insights thought my PhD studies in UCLA. He trusted and gave me the opportunity and freedom to explore different research directions in silicon photonics – I had to pick up from scratch in every project he has assigned and I am sure this is one of the most influential training in my career. My special thank is given to my MPhil advisor, *Prof. Andrew W. Poon*, for his training. I would not be able to survive in my PhD life without the base he have equipped me.

I would like to thanks my parents and my sisters for their unflagging love and support throughout my life. To my mother, who has always been there for me when I need her, the countless phone calls from her just to wake me up are always touching, never tiring. To my father, who always trust and support my decision. Without them, I would not be able to make to this stage. To my sisters, I am happy to have them who give me a happy childhood. They are truly the backbone of my life.

Last but not the least, I also have to thanks all the postdocs and graduate students who I have been working closely with in the past five years especially *Dr. Kevin Tsia*, for his initial help and guidance when I first stepped into this new country; *Dr. Eric Diebold* for his help and inspiration during my last few projects on silicon photonics; *Dr. Keisuke Goda* for his determination, passion and enthusiasm for research when I work with him on a collaboration project with Olympus Corp.
VITA

2002 – 2003  Wu Ti Hsien Science & Education Foundation and Li Po Chun Charitable Trust Fund Scholarship

2002 – 2005  5 out of 6 semesters Dean’s list Award in Hong Kong University of Science and Technology

2005  B. Eng. Electrical and Electronic Engineering (Information and Communication) (First Class Honors)
The Hong Kong University of Science and Technology

2005 – 2007  Graduate Student Researcher
Graduate Studentship support by the Hong Kong University of Science and Technology

2007  MPhil, Electrical and Computer Engineering
The Hong Kong University of Science and Technology

2007 – 2012  Graduate Student Researcher
Graduate Studentship support by the University of California, Los Angeles
PUBLICATIONS AND PRESENTATIONS

Publications


Lett. 94, 091116 (2009)


Presentations


15. **N. K. Hon** and A.W. Poon, “Surface plasmon resonance enhanced coupling to whispering-gallery modes in optical micropillar resonators,” in Proc. of the Conference on Lasers and


**Book Chapter**


   “Microresonator-based devices on a silicon chip: novel shaped cavities and resonance coherent interference,” a chapter in Practical Applications of Microresonators in Optics and Photonics, Ed. Andrey Matsko, Taylor and Francis, 2008
Chapter 1

Introduction

Undoubtedly, photonics impact every aspect of our lives. One of the biggest impacts we are witnessing is the invention of fiber-optic communication that drives human existence into a new stage – the information age. With the replacement of conventional copper wires with optical fibers, high speed and long distance data transmission is no longer a dream and this opens the door to the convenient, “fast” internet we cannot live without nowadays. Transmission speed is limited in copper wires and the same case happens on silicon computer chips. We, optical engineers, have applied the same tricks on silicon to solve these problems and a new research field named Silicon Photonics is started [1].

Thanks to the advancement in silicon CMOS fabrication and SOI technology developed for electronic devices, strong optical confinement in high index contrast silicon can be demonstrated on compact, low loss silicon waveguides. This fundamental invention draws lots of research attention and the vigorous research force has driven the continuous demonstration of passive devices such as optical filters and active devices such as the Raman laser on this platform [2]. As silicon is transparent at wavelengths near 1550 nm, the main application of silicon photonics is telecommunication. However, optical data communication is not the only area
where silicon photonics will have an impact. My research direction is to explore new capabilities of silicon. In this dissertation, a number of new applications of silicon photonics will be introduced, which range from chemical sensing and spectroscopy to diagnosis and therapy.

1.1. Third-order nonlinearity of silicon from midwave to longwave infrared

Silicon has excellent linear optical properties in the midwave and longwave infrared spectrum. These properties, along with silicon’s excellent thermal conductivity and optical damage threshold, open up the possibility for a new class of midwave and longwave infrared photonic devices [3]. As potential platforms for this new regime, a wide range of applications from gas detection, sensing, spectroscopy to free space communications can be realized on low cost, chip-level integration. The linear optical properties of silicon in this new regime have been well studied. However, this is not the case for the nonlinear optical properties which almost all active devices are based on. To assist nonlinear optical designers, the lowest-order nonlinearity – third-order susceptibility $\chi^{(3)}$, which gives rise to the Kerr and Raman effects, is the key to be studied. Detailed knowledge is needed because these coefficients may have strong wavelength dependences over the mid- to long- wave region (a little-known fact). Unfortunately, there are longwave regions or spectral “gaps” over which knowledge of the coefficients (as revealed in the scientific literature) is missing. Moreover, an examination of the literature shows some inconsistencies in experimental data. In chapter two, we will use a combination of semiconductor theory and experimental results from the scientific literature to compile and plot the key third-order nonlinear optical coefficients of bulk crystalline silicon from 1550 nm to 6.7 $\mu$m. for Si) in order to address these problems.
1.2. Second-order nonlinearity of stressed silicon

Silicon is a centrosymmetric crystal and it lacks second-order optical nonlinearity - a foundation of nonlinear optics. Third order optical nonlinear approaches such as Raman amplification in the mid-wave IR, cascaded Raman amplification or four wave mixing in the near infrared, have been demonstrated recently. However, we know that stress can break symmetry. In chapter 3, I will introduce a new class of photonic devices which enable the otherwise forbidden second order nonlinear in silicon through stress engineering by using periodically stressed fields in silicon. While creating second-order nonlinearity, quasi-phase matching can also be achieved. Periodically-poled silicon (PePSi) adds the periodic poling capability to silicon photonics, and allows the excellent crystal quality and advanced manufacturing capabilities of silicon to be harnessed for devices based on second-order nonlinear effects. As an example of the utility of the PePSi technology, we will present simulations showing that mid-wave IR radiation can be efficiently generated through difference frequency generation from near IR.

1.3. Terahertz generation using large area silicon PIN diodes

To extend the wavelength range over which silicon is a useful material, we study silicon as a material for generating terahertz waves. We know that silicon is also transparent in this new frequency range as high resistivity silicon is used for hemispherical lenses and windows for terahertz applications. Advances in both electronic and optical devices have narrowed the terahertz gap, and there now exist a large variety of sources to choose from based on the desired frequency. On the optical end, quantum cascade laser is one of the main candidates for high-power terahertz sources, but cryogenic cooling of the lasers is required. Purely optical
approaches based on difference mixing of two laser beams in a dielectric with strong second order nonlinearity are also available. Unfortunately, the displacement amplitude of the THz dipoles (very small compared with lattice spacing) created by this optical mixing is small, and thus the efficiency is limited. One approach to enhancing this efficiency is based on the absorption of the THz optical beating, which in turn leads to THz generation of free carriers. The free carriers are then accelerated in a high DC electric field over a distance which is much larger than the lattice contact. This approach is called photomixing. The efficiency of continuous wave photomixing is usually lower than impulsive excitation in which ultrafast (< 0.5ps) light is used to create free carriers. The free carriers are then accelerated in a high DC electric field. The transient transport of the carriers generates THz pulse. To start with, we investigate THz generation on large area silicon PIN diodes under impulsive excitation. Silicon is potentially an attractive platform for THz generation because of its high optical damage threshold and high thermal conductivity, not to mention its CMOS compatibility which can eventually enable integration with electronics. In chapter 4, we will propose and investigate a method for impulsive excitation, with improved coherence, higher energy efficiency, and reduced linewidth on large area silicon PIN THz emitters.

1.4. Giant tunable silicon based dispersive element for on chip real time frequency time mapping

For the purpose of performing spectroscopy, using optical dispersive elements is one of the ways to obtain spectral information. Light can be spatially dispersed using an optical grating and the spectral information can be resolved by either rotating the grating or scanning the detector. This is a typical way in which an optical spectrum analyzer works, but scanning usually
takes time. Another way is to make use of the temporal dispersion to extract the spectrum information as light with different color travels differently. This spectral-to-time-domain mapping enables high speed real time measurements for sensing [4]. So, dispersive optical elements play critical roles in the operation of many important applications, including time-wavelength spectroscopy, time-stretch analog-to-digital conversion, temporal imaging of optical pulses, and high-speed optical imaging. Unfortunately, the devices used to generate large chromatic dispersion exhibit significant tradeoffs between GVD and other performance metrics. Pairs of diffraction gratings commonly used in chirped pulse amplifier laser systems can offer large dispersion but at the expense of footprint because the GVD scales approximately linearly with the grating separation. Dispersion compensating fiber can provide a large amount of GVD in a small footprint, but has limited power-handling capability due to its large effective optical nonlinearity. More importantly, the use of dispersion compensating fiber is limited to the near-infrared telecommunication wavelength (~1550 nm) band. In chapter 5, we will introduce and demonstrate a new type of dispersive device that combines the spatial dispersion of diffraction gratings with the modal dispersion of a multimode waveguide to overcome the footprint, power, wavelength, and bandwidth limitations of conventional dispersive elements. This device, which may aptly be called the chromo-modal dispersion device (CMD), maps different wavelength components of an optical waveform to distinct modes of a multimode waveguide. The CMD then exploits the large modal dispersion inherent in a multimode fiber and silicon waveguide to create chromatic dispersion (GVD).
1.5. Study of porous silicon nanoparticles for in vivo cancer diagnosis and therapy

Many of research interests have been driven to biological applications due to community interests. Some of these researchers have been working on silicon-based optical biosensors for label-free detection. One example of label-free detection is the silicon based interferometer for detecting refractive index changes by the biological sample placed in one arm of the interferometer. Another interesting area of research related to Si is drug delivery system [5]. Conventionally, therapeutic agents are distributed non-specifically inside human body. For cancer therapy, both normal and cancerous cells will be affected. The maximum achievable dosing is therefore limited, which results in suboptimal treatment. Molecular targeted therapy, which is a type of drug delivery system, is a way to overcome the obstacles mentioned. There has been a great deal of attention focused on using nanoparticles drug carriers. The ability of functionalized nanocarriers to target cells for controlled drug release potentially improves delivery efficiency and specificity by orders of magnitude. PSiNPs are self-luminescent, biocompatible, biodegradable and non-toxic. These attractive physical and chemical characteristics have enabled research on PSiNPs to result in their preliminary use as in vivo drug carriers. Motivated by the previous work, we will discuss and investigate the distinct characteristics of PSiNPs and introduce the drug delivery applications in the context of the field of silicon photonics. In this chapter 6, we will present the engineering issues involved in obtaining particles with right luminescence spectrum for tracing or diagnosis, good biodegradability for prolonged lifetime in blood stream and bioconjugation techniques that can be used for surface functionalization.
REFERENCES


Chapter 2

The third-order nonlinear optical coefficients of Si, Ge and SiGe in the midwave and longwave infrared

Using a combination of semiconductor theory and experimental results from the scientific literature, we have compiled and plotted the key third-order nonlinear optical coefficients of bulk crystalline Si and Ge as a function of wavelength (1.5 - 6.7 μm for Si and 1.5 - 14.7 μm for Ge). The real part of third-order nonlinear dielectric susceptibility ($\chi^{(3)}$), the two-photon absorption coefficient ($\beta_{TPA}$) and the Raman gain coefficient ($g_R$), have been investigated. Theoretical predictions were used to curve-fit the experimental data. For a spectral range in which no experimental data exists, we estimate and fill in the missing knowledge. Generally, these coefficient-values appear quite useful for a host of device applications, both Si and Ge offer large $\chi^{(3)}$ and $g_R$ with Ge offering the stronger nonlinearity. In addition, we use the same theory to predict the third-order nonlinear optical coefficients of Si$_{1-x}$Ge$_x$ alloy. By alloying Si and Ge,
device designers can gain flexibility in tuning desired optical coefficients in between the two fundamental components based upon their application requirements.

2.1. Introduction

There are compelling reasons to migrate silicon photonics (known more accurately as group IV photonics) from the telecom wavelength region into the midwave and longwave infrared [1-6]. First, group IV photonics have many potential applications in chemical and biological sensing for realizing the lab-on-a-chip concept. Low cost miniature group IV sensors for trace gas detection, bio-agent sensing, environmental monitoring and industrial process control will attract researchers’ interest. Second, the undesired nonlinear loss, two-photon absorption (TPA), which is a limiting factor for nonlinear optical processes in the near-infrared vanishes at longer wavelengths as the energy of two photons is not enough for a band-to-band transition. Efficient nonlinear processes at longer wavelengths are expected.

In the infrared range, the principal on-chip integrated channel waveguides are likely to be silicon-on-insulator (SOI), silicon-on-sapphire (SOS), silicon-on-nitride (SON) and germanium-on-SOI (Ge/SOI), whose low-loss transmission range extends out to the wavelength of 3.7 μm for SOI, 4.4 μm for SOS, 6.7 μm for SON and 14.7 μm for Ge/SOI. The SOI waveguide has a propagation loss of less than 2 dB/cm over the 1.1 to 2.5 μm and 2.9 to 3.6 μm bands, with fairly high loss over 2.5 to 2.9 μm. Except for SOS, these longwave guides are untested at present. For nonlinear optical processes at longer wavelengths, there has been experimental realization for applications recently such as wavelength conversion [7,8], optical parametric waveguided gain [9], cascaded four-wave mixing for multi-line infrared sources [9], Raman waveguided gain [10] and cascaded Raman lasers [11]. Both Si and Ge possess strong third-order nonlinear optical...
(NLO) coefficients with reduced TPA, and in the aforementioned waveguides, we can project a bright future for third-order NLO effects and devices in the midwave and longwave infrared.

Designers of the new NLO devices mentioned here will surely rely upon knowledge of the relevant third-order NLO coefficients. Detailed knowledge is needed because these coefficients may have strong wavelength dependences over the midwave region (a little-known fact). Unfortunately, there are longwave regions or spectral “gaps” over which knowledge of the coefficients (as revealed in the scientific literature) is missing. Moreover, an examination of the literature shows some inconsistencies in experimental data. We have addressed these problems here.

To assist NLO designers in their new infrared mission, we have compiled the important NLO data and have filled in the knowledge gaps using semiconductor nonlinear optical theory. The resulting combination of experiment and modeling predictions will, we believe, give the community a comprehensive picture of the group IV nonlinear properties over the 1.5 to 14.7 μm wavelength range. In un-doped crystalline Si and Ge, the focus of this paper, the effects can be either resonant (Raman) or non-resonant, and the significant coefficients of interest are: the third-order nonlinear susceptibility, actually the real part of it denoted by $\chi^{(3)}$, the Kerr coefficient ($n_2$), the Raman gain coefficient ($g_R$) and the TPA coefficient ($\beta_{TPA}$). $n_2$ is linked to $\chi^{(3)}$ and we have used $n_2$ results (in some cases) to determine $\chi^{(3)}$. Thus, we investigate $\chi^{(3)}$, $g_R$ and $\beta_{TPA}$ in this paper (all units are in MKS throughout the paper). We are examining NLO effects at photon energies less than the band-gap energy, rather than examining a band-edge effect like the Franz-Keldysh modulation linked to $\chi^{(3)}$.

Besides bulk crystalline Si and Ge, the alloy of these two materials namely Si$_{1-x}$Ge$_x$ has long been investigated due to its importance in transistor fabrication where it provides higher
mobility of charge carriers. Also, the use of Si-rich Si$_{1-x}$Ge$_x$ alloy for the Raman laser has been demonstrated. This gives a tuned Si-Si Raman mode and thus different Raman emissions for same pump wavelength [12]. Although second-order nonlinearity exists in ordered Si$_{1-x}$Ge$_x$ due to the break of inversion symmetry, it is seldom considered as a choice for second-order NLO devices as there are lots of other choices (e.g. LiNbO$_3$) which give designers much stronger second-order effects [13]. So, we focus only upon the third-order nonlinearity of Si$_{1-x}$Ge$_x$ in this paper. We include Si$_{1-x}$Ge$_x$ here as a material that NLO designers can possibly choose based on their need to engineer the third order nonlinearity in between that of bulk Si or Ge. In other words, this alloying gives us an additional tuning parameter to engineer the bandgap and thus the magnitude of $\chi^{(3)}$, $\beta_{TPA}$ as well as the TPA cut-off regime. Moreover, the Raman line can be adjusted which gives flexibility in Raman-based device design.

The organization of this paper is as follows. We first look at the real part of the principal component ($\chi^{(3)}$), then we consider $\beta_{TPA}$, and finally $g_R$. The theories described in the sections below are employed to curve-fit the literature results and to estimate the third-order NLO coefficients in wavelength regions where no data exists for both Si and Ge. Regarding the third order optical nonlinearities of Si$_{1-x}$Ge$_x$, they are rarely explored even at near infrared wavelengths. So, we further use the estimated results and theoretical model to predict the coefficients of Si$_{1-x}$Ge$_x$. We find generally that $\chi^{(3)}$ and $g_R$ are quite strong in the midwave infrared, with Ge being significantly more nonlinear than Si, while the Si$_{1-x}$Ge$_x$ response resides between those of the elements. Generally, the TPA is strong in the near infrared but drops significantly with increasing wavelength and becomes extinguished at 2.2 μm for Si, at 3.2 μm for Ge and 2.2 – 3.2 μm based on the bandgap of Si$_{1-x}$Ge$_x$. Both $\chi^{(3)}$ and $g_R$ have useful device-
worthy response extending out to longer wavelengths. We believe there is tremendous untapped potential here for Si, Ge and Si\textsubscript{1-x}Ge\textsubscript{x}.

2.2. Real part of the third-order nonlinear susceptibility $\chi^{(3)}_{1111}'$

Silicon photonics has attracted considerable attention during the past two decades due to its inherent advantages such as CMOS compatibility and low-cost high-volume production. In addition, the optical nonlinearity of Si has been investigated extensively, primarily at wavelengths longer than those in the near-infrared telecom regime. Similar NLO investigations have been carried out for Germanium. Since Si is a centrosymmetric crystal for which the second-order susceptibility vanishes, optical nonlinearity research has centered on third-order NLO effects that are actually quite strong and thus of practical importance. (The effects are even stronger in Ge, but the optics community has not generally appreciated that fact). The non-resonant optical Kerr effect in Si is significant because it is two orders of magnitude greater than of glass optical fiber. A list of reported Kerr coefficients ($n_2$) within the 1530 - 1560 nm wavelength range [14-23] can be found in a review paper by Tsang et al. [24]. Recently, comprehensive measurements of $n_2$ are reported by Bristow et al. [22] and by Lin et al. [23] from 850 to 2200 nm and from 1200 to 2300 nm, respectively.

In this paper, we are investigating the primary third-order susceptibility which is the $\chi^{(3)}_{1111}'$ coefficient. The Kerr coefficient $n_2$ is related to the real part of $\chi^{(3)}_{1111}'$ by the following equation [25]:

$$n_2 = \frac{3}{4\varepsilon_0 cn^2} \chi^{(3)}_{eff}$$  \hspace{1cm} (1)
where $\chi_{eff}^{(3)}$ is the effective third-order nonlinearity, $n$ is the linear refractive index and $\varepsilon_0$ is the vacuum permittivity. The specifics of this relationship depend upon the crystallographic orientation, namely: $\chi_{eff}^{(3)}=(\chi_{1111}^{(3)}+3\chi_{1122}^{(3)})/2$ along the <110> directions and $(\chi_{1111}^{(3)}+6\chi_{1122}^{(3)})/3$ along the <111> directions. Also, $\chi_{1111}^{(3)} \sim 2.36\chi_{1122}^{(3)}$ [26]. Besides $n_2$, another nonlinear parameter $\gamma$ defined as below:

$$\gamma = \frac{n_2 \omega_0}{c A_{eff}}$$

is widely used in waveguide designing where $A_{eff}$ is the effective cross-section area of the waveguided mode. Since $n_2$ and $\gamma$ can be linked to $\chi_{1111}^{(3)}$, and the focus of this paper is bulk materials, we will mostly consider $\chi_{1111}^{(3)}$ in the rest of this paper.

We have compiled various experimental $\chi_{1111}^{(3)}$ susceptibility results given in the literature for both Si and Ge over the midwave and longwave region, and Fig. 2.1 presents our results. We can see in Fig. 2.1(a) that even for the recent measurements by Bristow et al. [22] and Lin et al. [23], there is a factor-of-six difference in the reported values. The low-loss infrared transmission window of Si extends to around 9 $\mu$m but there are no experimental reports of the susceptibility parameter from 2.3 to 6.7 $\mu$m. It is important to estimate the NLO coefficient in this unexplored regime, therefore we have employed in Fig. 2.1(a) a theoretical dispersion curve based on a two-band model in the transparency region of this indirect-gap semiconductor. We used the curve to fit the experimental data and to fill the spectral “gap” with predicted behavior [28,29].
The spectral dispersion curve of the real susceptibility $\chi^{(3)}$ can be related to the spectral response of the imaginary part of the susceptibility ($\chi^{(3)}_i$) by the nonlinear Kramers-Kronig relation [28]. Since the imaginary part is proportional to $\beta_{TPA}$, the real part is also linked to the TPA coefficient $\beta_{TPA}$ as follows:

$$\chi^{(3)}_{eff}(\omega) = 4\varepsilon_0cn^2 \int_{0}^{\infty} \frac{\beta_{TPA}(\omega, \omega')}{\omega^2 - \omega'^2} d\omega'$$  \hspace{1cm} (3)$$

Then, an understanding of $\beta_{TPA}$ will give us an understanding of the real susceptibility.

Using the approximation $\beta_{TPA}(\omega, \omega') = \beta_{TPA}(\omega + \omega')/2$ for non-degenerate $\beta_{TPA}(\omega, \omega')$, the degenerate $\beta_{TPA}(\omega)$ curve can be calculated using the model derived by Garica et al. [29], which
incorporates the “forbidden - forbidden” (f-f), “allowed - forbidden” (a-f) and “allowed - allowed” (a-a) optical indirect transitions. Bristow et al. have used the same model to curve-fit the experimental data. $\beta_{TPA}(\omega)$ which is the sum of all three transition processes can be expressed as follows [22]:

$$\beta_{WA}(\omega) = \sum_{i=0}^{2} \beta^{(i)}(\omega)$$

with $i = 0, 1, 2$ for a-a, a-f and f-f transitions. And

$$\beta^{(i)}(\omega) = K_i \frac{1}{n^2 E_{ig}^3} F_2^{(i)} \left( \frac{\hbar \omega}{E_{ig}} \right)$$

where $F_2^{(i)}(x) = \frac{(2x-1)^{i+2}}{(2x)^3}$, $E_{ig}$ is the indirect bandgap energy of Si and $K_i$ is a curve-fitting factor.

The f-f transition process is much weaker than the other two processes in the transparent regime [29]. So, the two factors $K_1$ and $K_2$, along with equations (3), (4) and (5), are used for curve-fitting the experimental data. Figure 2.1(a) shows two dispersion curves in which one of them is curve-fitted to results by Bristow et al. and the other one is curve-fitted to results by Lin et al. These two curves also set the upper and lower boundaries of the reported values in literature. Moreover, $\chi^{(3)}_{1111}$ is strongly wavelength dependent up to 3.5 $\mu$m and then declines to an asymptotic value at 4.0 - 6.7 $\mu$m. The reported value of $3.36 \times 10^{-19}$ m$^2$/V$^2$ at 10.6 $\mu$m by Wynne et al. [30] is also consistent with our estimation. An average of the two dispersion curves is plotted in Fig. 2.1 (a). Based on the average curve, NLO designers can have an estimated $\chi^{(3)}_{1111}$ value within a factor-of-three uncertainly from the experimental results. The inset lists the
\( \chi^{(3)}_{1111} \) and \( n_2 \) at 1.55 \( \mu m \) (the telecommunication wavelength), 2.2 \( \mu m \) (the TPA cutoff), 3.39 \( \mu m \) (the HeNe laser wavelength) and 4.26 \( \mu m \) (a CO\(_2\) absorption band).

The NLO effects of diamond-cubic Ge are much less explored than those of Si. Ge has a low-loss transmission window in the 2.0 - 14.7 \( \mu m \) range, a band wider than that of Si. Due to the availability of CO\(_2\) laser sources, the non-resonant third-order non-linearity of Ge was mainly investigated at 10.6 \( \mu m \) [30-33]. All the reported values are plotted in Fig. 2.1(b). The reported values are also summarized in review papers [35,36]. From the graph, we can see that the values show a factor-of-four difference, ranging from \( 0.35 \times 10^{-18} m^2/V^2 \) to \( 1.6 \times 10^{-18} m^2/V^2 \) at 10.6 \( \mu m \). To estimate the most realistic 10.6 \( \mu m \) value, we have chosen the mid-range value that is an average of the data points.

To estimate NLO within the spectral "gaps" of Ge, a theoretical dispersion curve based on the two-band model is used here again. Compared with indirect Si, the TPA of Ge is expected to be dominated by the direct bandgap of Ge which is close to its indirect bandgap [28]. So, we have chosen an expression for \( \beta_{TPA} \) that differs from that implied in equations (3) and (4). The expression of \( \chi^{(3)}_{1111} \) for Ge is determined from [32] to be:

\[
\chi^{(3)}_{1111}(\omega) = \frac{4\varepsilon_0 c n^2}{3} n_2(\omega) = \frac{160\pi\varepsilon_0}{3E_g} K' G_z \left( \frac{\hbar \omega}{E_g} \right)
\]

\( G_z(x) = \frac{-2 + 6x - 3x^2 - x^3 - \frac{3}{4}x^4 - \frac{3}{4}x^5 + 2(1-2x)^{3/2}\Theta(1-2x)}{64x^6} \), \( E_g \) is the direct band gap of Ge,

\( K' \) is a curve fitting parameter and \( \Theta(x) \) is a step function.

An upper and lower bound dispersive curves based on Eqn. (6) are used to enclose the experimental \( \chi^{(3)}_{1111} \) of Ge. A resulting average dispersive curve is then shown in Fig. 2.1(b)
which can give NLO designers an estimation value that is within a factor-of-two uncertainty from experimental results. The NLO coefficient is strongly wavelength dependent up to 6 µm and then settles at an asymptotic value above 8 µm. The inset lists the $\chi^{(3)}_{1111}$ and $n_2$ at 3.17 µm (TPA cutoff), 3.39 µm (HeNe laser), and 4.26 µm (a CO$_2$ absorption band). This graph can act as a reference for NLO designers dealing with Ge. Because $\chi^{(3)}_{1111}$ is quite large as shown, and because the Kerr effect of Ge is also ~10x higher than that of Si, Ge “looks like” a better nonlinear material.

Since experimental results for $\chi^{(3)}_{1111}$ of Si$_{1-x}$Ge$_x$ do not exist, we can use a combination of above two models, the indirect bandgap model that was used for estimating Si and the direct bandgap model that was used for estimating Ge, to estimate the $\chi^{(3)}_{1111}$ of Si$_{1-x}$Ge$_x$ in the midwave and longwave infrared range. Both models actually depend on two important parameters: the linear refractive index $n$ and the direct or indirect bandgap ($E_g$ or $E_{ig}$).

The refractive index of Si$_{1-x}$Ge$_x$ can be formulated as follows [37]:

\[
    \begin{align*}
        n(\omega) &= n_0 + n_i (\hbar \omega)^2 \\
        n_0(x) &= 4.01 - 0.81(1-x) + 0.22(1-x)^2 \\
        n_i(x) &= 0.216 - 0.211(1-x) + 0.089(1-x)^2
    \end{align*}
\]

Also, the compositional dependence of the indirect bandgap energy of Si$_{1-x}$Ge$_x$ has been studied, yielding the following expressions [38]:

\[
    \begin{align*}
        E_{\Gamma\Delta} &= 1.11 - 0.34x + 0.227x^2 \\
        E_{\Gamma L} &= 2.01 - 1.46x + 0.227x^2
    \end{align*}
\]

where $E_{\Gamma\Delta}$ is the bandgap energy near the X valley ($E_{ig} = E_{\Gamma\Delta}(x)$ when x = 0 for Si ), and $E_{\Gamma L}$ is the bandgap energy at the L valley. For the direct bandgap energy, we used the linear
approximation of equation (9) to estimate the SiGe bandgap energy based on the direct bandgaps of Si (3.30eV) and of Ge (0.82eV) [38].

\[ E_g = 3.39 - 2.57x \]  

(9)

Fig. 2.2. (a) Calculated bandgap energy of \( E_{\Gamma\Delta} \), \( E_{\Gamma L} \) and \( E_g \) of Si\(_{1-x}\)Ge\(_{x}\) alloy with respect to the atomic fraction of Ge. (b) Calculated \( \chi^{(3)} \)\(_{1111}\) of Si\(_{1-x}\)Ge\(_{x}\) alloy at the 6 \( \mu \)m wavelength with respect to the atomic fraction of Ge using the direct and indirect bandgap models.

Figure 2.2(a) shows the three bandgap energies as a function of \( x \). We can see that the minimum bandgap energy changes from \( E_{\Gamma\Delta} \) to \( E_{\Gamma L} \) at \( x = 0.8 \). We believe that the most realistic model for Si\(_{1-x}\)Ge\(_{x}\) prediction is a two zone approximation in which we use the indirect bandgap model for estimating \( \chi^{(3)} \)\(_{1111}\) of Si\(_{1-x}\)Ge\(_{x}\) for \( 0 \leq x \leq 0.8 \) while we use the direct bandgap model for \( 0.8 < x \leq 1 \). To confirm the validity of that approach, we plot the \( \chi^{(3)} \)\(_{1111}\) of Si\(_{1-x}\)Ge\(_{x}\) at 6 \( \mu \)m (the long wave asymptote) as a function of \( x \) in Fig. 2.2(b) using both models. The estimated value of Si\(_{1-x}\)Ge\(_{x}\) is calculated based on the two average \( \chi^{(3)} \)\(_{1111}\) dispersion curves (Si and Ge) we obtained in Fig. 2.1. The curves coincide at the 0.8 inflection point, and the figure reveals that the two curve-segments mentioned above give a good physical picture of the \( x \)-dependence. Putting these two Si\(_{1-x}\)Ge\(_{x}\) models together in the prescribed way, we next show the calculated mid-wave/long-
wave spectra of $\chi^{(3)}_{1111}$ for $x = 0.2, 0.4, 0.8$ and $0.9$ in Fig. 2.3. We believe that this graph can give the NLO designer a reference to tune the desired $\chi^{(3)}_{1111}$ by bandgap engineering. We remark that the estimated values here have at most a factor-of-three uncertainly as it is based on the average experimental values of Si and Ge.

![Graph](image)

Fig. 2.3. Theoretical $\chi^{(3)}_{1111}$ dispersion curves of Si$_{1-x}$Ge$_x$ alloy with $x = 0.2, 0.4, 0.8$ and $0.9$.

### 2.3. Two-photon absorption coefficient $\beta_{TPA}$

Although two-photon modulators are feasible, TPA is considered here mainly as an unwanted effect that sometimes obscures a desired nonlinear $\chi^{(3)}_{1111}$ “parametric process”. The effective nonlinear process is often characterized by a nonlinear figure of merit (FOM) defined as $n_2/(\lambda \beta_{TPA})$. That’s why knowledge of TPA is “essential”. In addition to TPA, it is well known that free carrier absorption (FCA) is another counteracting effect induced by TPA. Both effects limit the achievable gain and thus a more comprehensive FOM should include both effects.

$\beta_{TPA}$ of Si has been well investigated and reported in the literature. A list of reported $\beta_{TPA}$ over 1530 - 1560 nm wavelengths [14-16,20,22-23,27,39-41] can be found in a review paper by
Tsang et al. [24]. Also, both Bristow et al. [22] and Lin et al. [23] have reported comprehensive measurements of $\beta_{TPA}$ from 850 to 2200 nm and from 1200 to 2300 nm, respectively.

Figure 2.4(a) shows the experimental $\beta_{TPA}$ for silicon in literature as a function of wavelength. Data points are determined from Refs. 14-16, 20, 39-41 (circles), Ref. 22 (upward triangles), Ref. 23 (downward triangles), Ref. 27 (diamonds) and Ref. 43 (squares). One theoretical dispersion curve is fitted to Ref. 22 (Bristow et al.) and another to Ref. 23 (Lin et al.). (b) Experimental $\beta_{TPA}$ of Ge in the literature as a function of wavelength. Data points are determined from Ref. 44 (diamond), Ref. 45 (triangles) and Ref. 47 (circles). A theoretical dispersion curve is fitted to the data points.

Figure 2.4(a) shows the experimental $\beta_{TPA}$ for silicon in literature as a function of wavelength. A theoretical degenerate $\beta_{TPA}(\omega)$ curve determined by Eqns. (3) and (4) as discussed in Sec. 2 is used for the curve-fitting. In a situation similar to that for the reported $\chi^{(3)}_{1111}$ values, the $\beta_{TPA}$ given in the literature show large inconsistencies. We can see that there is a factor-of-three difference between the values reported by Bristow et al. and Lin et al. as Fig. 2.4(a) reveals. Two dispersion curves, one fitted to the Bristow’s data and the other fitted to Lin’s data, are plotted. These curves set the upper and lower boundaries on the reported results. Here we note that the $\chi^{(3)}_{1111}$ in Fig. 2.1(a) is derived by the Kramers-Krönig relation from the $\beta_{TPA}(\omega)$ in Fig. 2.4(a) with an additional scaling factor. This scaling discrepancy can be accounted for by the neglect of Raman linear- and quadratic- Stark effects [22].
For Ge, most of the reported $\beta_{TPA}$ are from references [43-45]. Very early results show unexpectedly high values that are now interpreted as the combined effect of TPA and TPA-induced FCA [43]. A comprehensive measurement of $\beta_{TPA}$ from 2.6 to 3 $\mu$m is reported by Gibson et al. [45]. Their data spectrum was fitted successfully with the simple band-model prediction of Bryant et al. [46]. Also, a recent measurement of $\beta_{TPA}$ reported by Rauscher et al. [47] shows results that are consistent with those of Gibson et al.

Those reported $\beta_{TPA}$ values are shown in Fig. 2.4(b). Here we use a two-band model to fit the experimental data. As mentioned in Sec. 2, the model is based upon the direct bandgap of Ge [32]. The degenerate $\beta_{TPA}$ can be expressed as follows:

$$\beta(\omega) = K \frac{1}{n^2 E_g} F_2 \left( \frac{\hbar \omega}{E_g} \right)$$

(10)

where $F_2(x) = \left( \frac{(2x-1)^{3/2}}{(2x)^3} \right)$ and $K$ is a curve-fitting factor.

![Fig. 2.5. Theoretical $\beta_{TPA}$ dispersion curves of Si$_{1-x}$Ge$_x$ alloy with (a) $x = 0.2$, 0.4 and 0.6 using the indirect bandgap model and (b) $x = 0.8$, 0.9, 0.95 and 1 using the direct bandgap model.](image)

The $\beta_{TPA}$ of Ge in its transparent regime is much stronger than that of Si. This important finding needs to be taken into account when designing Ge NLO devices. We see that Ge’s strong
\( \chi^{(3)}_{1111} \) comes with higher \( \beta_{TPA} \). By alloying Si with Ge, we can work a tradeoff among strong-enough \( \chi^{(3)}_{1111} \), TPA strength and TPA cutoff. To estimate \( \beta_{TPA} \) for \( \text{Si}_{1-x}\text{Ge}_x \), we use the two zone approximation again. Since the approximation works well for \( \chi^{(3)}_{1111} \), it should do well for \( \beta_{TPA} \) because of the Kramers-Kronig linkage. Figure 2.5(a) shows the calculated \( \beta_{TPA} \) for \( x = 0.2, 0.4 \) and 0.6 using the indirect bandgap model and Fig. 2.5 (b) shows the calculated \( \beta_{TPA} \) for \( x = 0.8, 0.9, 0.95 \) and 1, applying directness. However, the \( \beta_{TPA} \) in the two plots differ by a factor of 100.

A possible explanation for this discrepancy is that different curve-fitting parameters were used for \( \chi^{(3)}_{1111} \) and \( \beta_{TPA} \) to fit experimental points. Not many experimental \( \beta_{TPA} \) values are available for Ge, and those data reported may (inadvertently or unknowingly) include linear absorption effects due to FCA, resulting in higher-than-actual \( \beta_{TPA} \) values. Nevertheless, we still include Fig. 2.5(b) to reveal the general dispersion trends for NLO engineers when using Ge or Ge-rich alloy.

In Fig. 2.5, we can see that the wavelength cut-off of \( \beta_{TPA} \) changes with different \( x \) composition. This cut-off depends on bandgap, and in Figure 2.6 we present this wavelength as a function of composition. Designers can thus choose the alloy composition that eliminates TPA at their working wavelength.
2.4. Raman gain coefficient $g_R$

For Si and Ge, as crystalline solids, Raman scattering is another important third-order NLO process that has attracted huge attention for NLO device engineering. The Raman scattering we consider here is the first-order Raman scattering from the zone-center optical phonons. Stimulated Raman scattering has been used for making Raman amplifiers and Raman lasers on a silicon chip. Most of the reported $g_R$ values for Si are for optical pumping performed between 1427 and 1565 nm [39,49-55]. The gain coefficient has already been summarized by Tsang et al. [24]. In addition, mid-infrared pumping at 2.88 $\mu$m for mid-IR Raman amplification has been demonstrated in bulk silicon crystal [10] and in a Si slab waveguide. The reported values are shown in Fig. 2.7(a) with the unexpectedly high values reported in Ref. 39 being neglected. The Raman process in Si is well explored near telecom wavelengths, but little or no research has been done on Ge Raman within its low-loss window. However, it is known that Ge offers stronger Raman gain than Si [56]. Here we include a theoretical dispersion curve of $g_R$ in silicon to probe the unexplored gap in the midwave and longwave infrared range. Based on the result of silicon, the $g_R$ dispersion curve of Ge is also predicted.
In Fig. 2.7(a), it is seen that the reported $g_R$ ranges widely from 40 cm/GW down to 4 cm/GW. The different pumping schemes used might account for this wide variation [24]. Typically, lower gain is obtained with pulse pumping [57,58].

The Raman gain coefficient can be expressed as follows [59]:

$$ g_R = \frac{8 \pi e^2}{\hbar \omega_s^3 n_s^2 (N + 1) \Gamma} S $$

(11)

where $\omega_s$ is the Stoke’s frequency, $n_s$ is the linear refractive index at the Stoke's wavelength, $N$ is the Bose occupation factor, $\Gamma$ is one-half the FWHM of the spontaneous lineshape and $S$ is the spontaneous Raman scattering efficiency.

The efficiency for one-photon scattering is given by [60] as

$$ S \propto \omega_s^4 |d|^2 $$

(12)

where $d$ is the absolute cross-section of the first-order Raman scattering or Raman polarizability of Si. $d$ is not dispersive in the transparent region [61]. So,

$$ g_R \propto \omega_s $$

(13)

by choosing $g_R \sim 15$ cm/GW at 1.5 $\mu$m, the dispersive curve of $g_R$ for Si shown in Fig. 2.7(a) is obtained, a curve that fits the reported data in the literature.
Unlike Si, there is no experimental Raman gain coefficient reported for Ge explicitly in the wavelength range studied here, but the $g_{R,Ge}$ can be estimated using a measurement of relative Raman polarizability reported in the literature. Assuming $d_{Ge}$ is not strongly dispersive in its transparent range [62], which is the case in Si, we can then estimate the $g_{R,Ge}$ based on $g_{R, Si}$.

$$\frac{g_{R,Ge}}{g_{R, Si}} = \frac{S_{Ge} n_{Ge}^2 \Gamma_{Si}}{S_{Si} n_{Si}^2 \Gamma_{Ge}}$$

(14)

The ratio of $\Gamma_{Si}/\Gamma_{Ge}$ is 0.87 [56] and the ratio of $S_{Si}/S_{Ge}$ is $\sim 6.8$ ($d_{Si}/d_{Ge} \sim 0.33$) [62]. So,

$$g_{R,Ge} \approx 4.5 g_{R, Si}$$

(15)

Ge has a much larger Raman gain than Si and the estimated curve is shown in Fig. 2.7(a).

Notice that the Raman modes in Si or Ge have a fixed Raman shift, but the mode can be tuned by alloying or by inducing stress/strain. It was demonstrated that Si-rich Si$_{1-x}$Ge$_x$ alloy can be used to make a Raman laser with different emission wavelengths under same pump wavelength [12]. Here we are going to estimate the $g_R$ of Si$_{1-x}$Ge$_x$. For Si$_{1-x}$Ge$_x$, there are three main Raman peaks: the Si-Si mode, the Ge-Ge mode and the Si-Ge mode [63]. The frequency and the intensity of Raman peaks depends upon $x$ and the induced stress/strain due to fabrication...
techniques. In addition, experimental results show that there are extra and disorder modes [64]. For simplicity, we do not consider the effect of stress/strain upon the Raman shift. Also, we know that the intensity of the Si-Ge mode is always weaker than that of Si-Si or Ge-Ge modes [65,66]. Therefore we shall focus only on the $g_R$ of the Si-Si and Ge-Ge modes for various $x$. We shall assume that we can neglect the Raman shift due to $x$ as it is not the dominant effect in $g_R$.

In estimating the change in $g_R$, the scattering efficiency is the main factor. Assuming that the Si and Ge atoms are arranged randomly in the lattice, the effective Raman scattering is $(1-x)^2S_{Si}$ for the Si-Si mode and is $x^2S_{Si}$ for the Ge-Ge mode [63,65-66]. As $g_R \propto S$, the Raman gain coefficient of the Si-Si and Ge-Ge modes in Si$_{1-x}$Ge$_x$ can be estimated using following relations:

$$
g_{R,\text{Si-Si}} = (1-x)^2 g_{R,\text{Si}}$$
$$
g_{R,\text{Ge-Ge}} = x^2 g_{R,\text{Ge}}$$

Using Eqn. 16, Fig 2.7 (b) shows the $g_R$ of both Si-Si and Ge-Ge modes as a function of $x$ at the 6 $\mu$m wavelength. It can be seen that the $g_R$ of the Ge-Ge mode is higher than that of the Si-Si mode in Ge-rich alloy and vice-versa.

Similar to the case of the Kerr effect we mentioned in Sec. 3, TPA and the resulting FCA also limits the achievable gain we can have from Raman scattering. But there is a unique exception; the Inverse Raman Scattering where in the presence of FCA enhances the effect that would otherwise be masked by the Coherent Anti-Stokes Raman Scattering (CARS) signal [67].
2.5. Conclusions

In this paper, we use a combination of two-band theory and experimental results to compile the key third-order nonlinear optical coefficients ($\chi^{(3)}_{1111}$, $\beta_{\text{TPA}}$ and $g_R$) of bulk crystalline Si, Ge and Si$_{1-x}$Ge$_x$ alloy as a function of wavelength in order to assist NLO designers in their new infrared mission. These values should be useful for a host of device applications in the midwave to longwave infrared, with Ge offering the stronger nonlinearity. The effects and devices deserve further exploration because of their tremendous untapped potential in this important infrared regime.
REFERENCES


Chapter 3

Periodically-poled silicon (PePSi) –

Second-order nonlinearity of stressed silicon

As a centrosymmetric crystal, bulk silicon lacks second-order optical nonlinearity – a foundational component of nonlinear optics. Hence, its lowest-order nonlinearity originates from the third-order susceptibility $\chi^{(3)}$, which gives rise to the Raman and Kerr effects [1]. However, silicon’s crystal symmetry can be broken to create second-order nonlinearity by applying a dc-electric field [2], operating at interfaces where the crystal symmetry is interrupted [3], and applying mechanical stress (or equivalently strain) in the material [4,5].

An intriguing property of parametric $\chi^{(2)}$ processes is the possibility of achieving quasi-phase matching (QPM) by periodic poling – a technique that enhances the efficiency of nonlinear interactions. Unfortunately, conventional poling processes, such as those used for lithium niobate and nonlinear polymers, do not apply to silicon because it lacks a dipole moment in its native
form. Here, we propose an approach for realizing, what is in effect, periodically-poled silicon (PePSi), a new technology for efficient second-order nonlinear processes. We achieve this by creating alternating stress fields along a silicon waveguide using a periodic arrangement of stressed thin films. We show that this structure creates appreciable $\chi^{(2)}$ and, simultaneously, achieves QPM. Via comprehensive numerical simulations that include two-photon absorption (TPA), as well as free-carrier plasma effects and its wavelength dependence, we show that PePSi can be used for efficient mid-wave infrared (MWIR) generation. Such a capability has many applications including remote sensing of chemical and biological agents and environmental monitoring [6]. Piezoelectric thin films have previously been used to dynamically control phase matching in silicon wavelength converters that operate based on the $\chi^{(3)}$ nonlinear response [7]. PePSi can also be combined with piezoelectric layers, an powerful technology that offers the ability to dynamically control $\chi^{(2)}$ in silicon using intelligent electronic circuitry.

3.1. Mid-wave infrared generation using PePSi waveguide

MWIR generation in silicon can be implemented by first-order stimulated Raman scattering (SRS) pumped at shorter MWIR wavelengths [8], or by cascaded SRS pumped at near infrared (NIR) wavelengths [9,10]. In contrast, our MWIR generation approach using PePSi is a single-step conversion from NIR via difference frequency generation (DFG). This is more favorable in terms of the wide-availability of the NIR pump sources and also more efficient in terms of circumventing the cumulative TPA and associated free-carrier absorption (FCA) that occurs in the cascaded SRS approach.
The PePSi structure considered here is a channel waveguide integrated with two types of silicon nitride (SiN) stressed films: one with tensile stress and another with compressive stress, periodically deposited along the waveguide, as shown in Fig. 3.1. Hence, the sign of stress induced within the silicon core alternates along the waveguide and results in alternating dipole regions – a new type of periodically-poled structure.

SiN is chosen because it achieves high film stress [11] and is compatible with complementary metal-oxide-semiconductor processing. In addition, the force in SiN films can be readily tailored from compressive to tensile stress depending on the deposition conditions [11]. It thus permits flexible engineering of stress, and hence $\chi^{(2)}$, in silicon. In practice, the design shown in Fig. 3.1 can be realized by two different SiN deposition steps, giving rise to alternating stresses along the waveguide. For MWIR applications, silicon-on-insulator (SOI) is not the desirable platform because of the high losses of silicon dioxide at MWIR wavelengths. Instead, silicon-on-sapphire is employed here as sapphire is transparent to MWIR.

Fig. 3.1. An example of PePSi waveguide formed by covering a silicon channel waveguide with two types of silicon nitride (SiN) stressed films periodically along the waveguide: one induces compressive stress, the other one produces tensile stress.
The PePSi waveguide we consider has the silicon core dimensions of 2 μm × 2 μm, which supports optical waveguiding in both MWIR and NIR regimes. We assume the in-plane stresses in the two different SiN films (with 1 μm thickness) to be +1 GPa (tensile) and −500 MPa (compressive) [11]. The SiN stressed film period is designed to be 8 μm in order to phase-match the interacting waves in the DFG process for MWIR generation. To estimate the stress-induced $\chi^{(2)}$ in the PePSi waveguide, we simulated the stress distribution in the waveguide by a three-dimensional finite-element analysis package (ANSYS). In principle, all the stress components (i.e., the normal and shear stresses in all directions) should be considered for evaluating the stress-induced $\chi^{(2)}$ because it is the anisotropy of the stress responsible for breaking the original crystal symmetry. Nevertheless, it is found that, compared to all other stress

Fig. 3.2. Simulated cross-sectional $y$-component normal stress ($\sigma_{yy}$) distribution where the PePSi waveguide is covered by silicon nitride films with 1 GPa (tensile) stress and (b) −500 MPa (compressive) stress. (c) Simulated average $\sigma_{yy}$ and estimated second order susceptibility ($\chi^{(2)}$) induced along the waveguide (z-direction).
components, the \( y \)-component normal stress (\( \sigma_{yy} \)) gives rise to a highly uniform stress distribution in the silicon core with a considerably higher average stress magnitude in the present PePSi waveguide design. Hence, it is conceivable that stress-induced \( \chi^{(2)} \) here is dominated by \( \sigma_{yy} \). Simulated cross-sectional \( \sigma_{yy} \) distributions in the tensile and compressive SiN cladding regions are shown in Fig. 3.2 (a) and (b), respectively. In one half of the period, when the SiN film exhibits tensile stress that deforms the waveguide, a compressive stress field is induced and confined inside the waveguide core (Fig. 3.2(a)) in order to counteract the deformation under elastic equilibrium. Conversely, in another half of the period, the compressive SiN cladding induces tensile stress within the silicon core (Fig. 3.2(b)). In addition, the silicon core, covered by the conformal SiN stressed cladding, displays a stress distribution, and hence the stress-induced \( \chi^{(2)} \) distribution, with good uniformity (Fig. 3.2(a) – (b)). This feature is important to ensure efficient \( \chi^{(2)} \) interaction of the optical modes within the waveguide. The picture of the periodic poling becomes more appealing when we observe the average stress (\( \sigma_{yy} \)) along the waveguide. As illustrated in Fig. 3.2(c), the periodic oscillation with \(-500\) MPa peak average compressive stress and \(+200\) MPa peak average tensile stress is evident.

Based on a classical anharmonic oscillator model [12], the stress-induced \( \chi^{(2)} \) can be estimated as

\[
\chi^{(2)} = \frac{4q^3 S}{(m^2 \alpha \omega^4 a^4)},
\]

where \( q \) is the electron charge, \( m \) is the electron mass, \( \varepsilon \) is the dielectric permittivity, \( a \) is the lattice constant, \( \omega \) is the angular frequency of light, and \( S \) is the strain, which is related to the stress by Young’s modulus of the material and an element of the stress tensor. Thus, based on the simulated stress values, the induced \( \chi^{(2)} \) is estimated to have oscillatory values from \(-15\) pm/V to \(+6\) pm/V in a period of 8 \( \mu m \) (Fig. 3.2(c)). The estimated \( \chi^{(2)} \) agrees qualitatively with the prior work for the same order of magnitude of stress [4].
The PePSi technology introduced here allows efficient silicon photonic devices based on different $\chi^{(2)}$ processes. As an example, we consider generation of MWIR radiation in PePSi waveguide from two NIR beams using DFG. We numerically investigated the QPM-DFG process in the PePSi waveguide using the nonlinear Schrödinger equation (NLSE), which incorporates (1) stress-induced $\chi^{(2)}$ effects, (2) $\chi^{(3)}$ effects (Kerr effect, TPA). We simultaneously calculate the free-carrier concentration resulting from TPA and utilize known empirical relations, proposed by Soref and Bennet, to determine the associated free-carrier absorption (FCA) and refraction (FCR) [13]. Taking the waveguide dispersion into account, the 8-µm-period of the SiN film pattern is designed to satisfy the QPM condition in the DFG process: pump at 1.3 µm, signal at 1.75 µm and idler at 5.1 µm. We consider the fundamental TE-polarized modes of the three interacting waves. Although the waveguide is multi-mode at the pump and signal wavelengths, the higher-order modes are expected to have no significant effect on the QPM-DFG process if the pump and signal propagate predominantly in the fundamental modes because higher-order modes have different mode profiles and dispersion properties, which lead to different phase-matching conditions.

In the model, we input transform-limited pump and signal pulses (both with pulse width of 12 ps), which have peak intensities of 1.5 GW/cm$^2$ and 12.5 MW/cm$^2$, respectively, into a 2-cm-long PePSi waveguide. The peak intensities are chosen to obtain the highest conversion efficiency. Through QPM-DFG, the MWIR idler pulse can be efficiently generated at 5.1 µm at the waveguide output (Fig. 3.3(a)). During the process, second harmonic generation (SHG) and sum frequency generation (SFG), albeit without phase matching, can generate photons with energies above the silicon bandgap. This inevitably causes single-photon absorption (SPA), which introduces additional FCA and, thus, deteriorates the DFG efficiency. We have
incorporated both SHG and SFG in the model, but found that the efficiencies of both processes are negligibly low along the waveguide because both second-harmonic and sum-frequency waves are highly attenuated by SPA in the early stage of propagation. On the other hand, since the second-harmonic and sum-frequency waves remain weak throughout the waveguide, SPA is insignificant. We find that the average free-carrier concentration along the waveguide generated by SPA (due to SHG and SFG) is $\sim 10^{12}$ cm$^{-3}$, whereas that generated by TPA is $\sim 10^{15}$ cm$^{-3}$. Hence, the contribution of SPA to the overall FCA is still overwhelmed by that of TPA. We also remark that even though the idler wave suffers from larger FCA than the pump and signal waves due to the wavelength-dependence of FCA, only the trailing edge of the idler pulse is suppressed by the FCA since more free carriers are generated toward the trailing edge [14]. As a result, high peak conversion efficiency can still be obtained.

Fig. 3.3. (a) Calculated output spectra in a 2-cm-long PePSi waveguide (Period $\sim 8$ $\mu$m): a 12-ps pump pulse at 1.3 $\mu$m (input peak intensity = 1.5 GW/cm$^2$), a 12-ps signal pulse at 1.75 $\mu$m (input peak intensity = 12.5 MW/cm$^2$), idler at 5.1 $\mu$m. (b) Calculated conversion efficiency of MWIR generation.

As depicted in Fig. 3.3 (b), which shows the peak conversion efficiency as a function of the peak pump intensity, the PePSi waveguide is able to achieve a maximum conversion efficiency of the MWIR generation by QPM-DFG as high as $\sim -3$ dB ($\sim 50\%$) under peak pump
intensity of 1.5 GW/cm². This intensity level is readily achievable under experimental conditions. The slight decrease of conversion efficiency at high pump intensities observed in Fig. 3.3(c) is primarily due to TPA and FCA in the pump, signal, and idler waves, as well as FCR, which introduces additional phase-mismatch in the DFG process.

In summary, we have proposed a new type of photonic device based on periodically-poled silicon (PePSi). Our approach employs periodic stress fields in silicon waveguides such that the crystal symmetry of silicon is broken in a periodically alternating fashion. Introducing, what is in effect, the functional equivalent of periodic poling technology into silicon photonics offers a path to realize efficient wave mixing devices based on second-order optical nonlinearity. As an example of the utility of the PePSi technology, we numerically show that using QPM-DFG, MWIR generation at 5.1 µm in PePSi waveguide with conversion efficiency as high as ~ 50% can be achieved. When combined with piezoelectric stressed layers, this technology will offer the capability to dynamically control the χ(2) in silicon.

3.2 Discrepancy between theoretical and experimental χ(2) values

For the above simulation, we adopt the reported χ(2) reported by Jacobson et al. Interestingly, there appears to be a difference between the experimental [15, 16] and theoretical [4] values of χ(2) reported in the literature for strained silicon, with theoretical values being much lower than their experimental counterparts. Here, we discuss the difference between the theoretical and experimental results and speculate on a possible reason for the discrepancy. We also show a simple, intuitive model for the origin of strain-induced χ(2) in silicon, which produces a result in approximate agreement with previously calculated values.
Here we assume a simplified 2D centrosymmetric lattice structure as shown in Fig. 3.4 to estimate the magnitude of the $\chi^{(2)}$ value in an actual strained silicon diamond lattice. Four silicon atoms separated by the Si-Si bond length $d$ are considered. This assumption is valid for an order of magnitude estimation as the induced E-field is dominated by nearest neighbor Si atoms. In Fig. 3.4, the centrosymmetry is broken by an inhomogeneous compressive stress with a gradient along the $y$-direction. The displacement of silicon atoms induces a net E-field. Hence, the net E-field distribution along the $y$-direction experienced by the bond charges at the center of the lattice can be calculated using Coulomb’s law and expressed as follows:

$$E_{\text{net}}(y) = \frac{2e}{\varepsilon_\infty} \frac{(d/2 + y)}{((d/2 - \nabla S d/2)^2 + (d/2 + y)^2) \sqrt{3}} - \frac{(d/2 - y)}{((d/2)^2 + (d/2 - y)^2) \sqrt{3}}$$

where $\nabla S$ is the strain-gradient.

Assuming 1 GPa in-plane stress provided by a 1-μm-thick silicon nitride strained film deposited on top of a silicon channel waveguide with core dimensions of 2 μm x 2 μm, the simulated strain along the $x$-direction ($S_{xx}$) is shown in Fig. 3.5 (a). Fig. 3.5 (b) shows the $S_{xx}$
value along the y-direction at x = 0. We note that the strain-gradient is $\sim 2 \times 10^3 \text{ m}^{-1}$. We estimate the $\chi^{(2)}$ value from its $\chi^{(3)}$ value as follows:

$$
\chi^{(2)} = \chi^{(3)} E_{\text{net}}
$$

(2)

where $E_{\text{net}}$ is the net induced E-field. Based on the simulated strain-gradient, the induced E-field is $\sim 0.2 \times 10^6 \text{ V/m}$. As the $\chi^{(3)}$ value in silicon is $\sim 0.25 \times 10^{-18} \text{ m}^2/\text{V}^2$ [17], the estimated $\chi^{(2)}$ is $\sim 0.05 \text{ pm/V}$. We can see that the estimated value bears an apparent discrepancy with the experimental results.

Fig. 3.5. (a) Simulated cross-sectional x-component normal strain ($S_{xx}$) distribution. (The channel waveguide is covered by silicon nitride film with 1 GPa compressive stress) (b) $S_{xx}$ along y at x = 0 $\mu$m

Besides, silicon surface under inhomogeneous stress has been widely investigated and sophisticated theoretical models have been developed to describe the $\chi^{(2)}$ due to inhomogeneous strain in bulk crystals with inversion symmetry. These models are based on the sp$^3$-orbitals of the Si-Si bond in a unit cell of silicon diamond structure. The effective induced field is estimated as a gradient of the bandgap caused by non-uniform strain, and as such is proportional to the strain-gradient multiplied by the deformation potential of the crystal. Based on the model [15, 16], it has been estimated that the $\chi^{(2)}$ is $\sim 0.1 \sigma \Gamma$ where $\sigma$ is stress in GPa and $\Gamma$ is the stress gradient.
relaxation rate in \( \mu \text{m}^{-1} \). Assuming the material is under a stress of 0.5 GPa with \( \Gamma = 0.6 \ \mu \text{m}^{-1} \) as in Fig. 3.2, \( \chi^{(2)} \) is \( \sim 0.03 \ \text{pm/V} \) which is consistent with our estimation.

Apart from the experimental work of Jacobson et al. another experimental measurement of \( \chi^{(2)} \) induced by silicon dioxide strained film has been performed at high frequency (surface SHG). The results suggested a \( \chi^{(2)} \) value of \( \sim 10 \ \text{pm/V} \) under 1 GPa stress in silicon. We note that all the experimental results are in good agreement, while the theoretical results are consistent with each other as shown in Table 1. However, the theoretical results underestimate the measurements. The apparent disagreement suggests that there may be important missing physics in the existing theoretical calculations.

Here show a table with the comparison between theoretical and experimental results

<table>
<thead>
<tr>
<th>Theoretical works</th>
<th>Experimental works</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D Coulomb's law model</td>
<td>Jacobson et al. [4, 18]</td>
</tr>
<tr>
<td>( \chi^{(2)} ) is ( \sim 0.05 \ \text{pm/V} ) under 1 GPa stress in the SiN strained film)</td>
<td>( \chi^{(2)} ) is ( \sim 15 \ \text{pm/V} ) under 1 GPa stress in the SiN strained film)</td>
</tr>
<tr>
<td>3D bond additivity model [15]</td>
<td>Govorkov et al. [16]</td>
</tr>
<tr>
<td>( \chi^{(2)} ) is ( \sim 0.03 \ \text{pm/V} ) under 1 GPa stress in the SiN strained film)</td>
<td>( \chi^{(2)} ) is ( \sim 10 \ \text{pm/V} ) under 1 GPa stress in material)</td>
</tr>
</tbody>
</table>

Table 3.1. Comparison between theoretical and experimental results

### 3.3. Experimental results and further work

In order to demonstrate the PePSi idea, we first have to verify the \( \chi^{(2)} \) induced in a stressed Si waveguide experimentally. We repeated the measurement done by Jacobson et al. but on a micron sized waveguide to quantify the value of \( \chi^{(2)} \). We fabricated stressed and control waveguides using UCLA Nanolab facilities. By applying voltage on the waveguide with metal electrodes deposited, the phase shift can be measured using a Mach–Zehnder interferometry. However, we did not see any measureable \( \chi^{(2)} \) induced. We conclude that the \( \chi^{(2)} \) induced in the
silicon core is much less than 1 pm/V which agree with theoretical works. Thus, the high \( \chi^{(2)} \) value published is not come from the stressed bulk silicon and we suspect that it is a surface \( \chi^{(2)} \) with an enhancing effect as the value is higher than conventional surface \( \chi^{(2)} \) (\( \sim 3\) pm/V). Further simulation and experiment will be focused on exploring the physics behind.
REFERENCES


Chapter 4

Silicon PIN diode based quasi-continuous wave terahertz emitter

An important approach for generating broadband THz radiation is based on photo-induced carriers in semiconductors accelerated by a high internal electric field. The transient acceleration of photogenerated carriers in the intrinsic region of PIN diode generates THz radiation [1,2].

With respect to coupling THz to free space, there are two approaches: antenna emitters (AEs) and large area emitters (LAEs). LAEs are attractive as they have lower beam divergence, as well as a much larger active area over which to distribute optical power, thus reducing the thermal load. As a result, higher optical power can be used to generate stronger THz power without reaching the thermal limit and damage threshold of the emitter. In addition, LAE allows much less demanding processing procedures for fabrication as compared to most antenna emitters.

In this chapter, we propose and investigate a method for impulsive excitation, with improved coherence, higher energy efficiency, and reduced linewidth on large area silicon PIN THz emitters. PIN diodes for THz generation are receiving attention as they provide unity
photoconductive gain as opposed to THz photoconductive switches that by virtue reduced lifetime (in low temperature grown material) have lower gain. Silicon is potentially an attractive platform for THz generation because of its high optical damage threshold and high thermal conductivity, not to mention its CMOS compatibility which can eventually enable integration with electronics. THz emission from silicon PIN diodes can also be used as a tool to study the dynamics of carriers transport under high electric field [1,3]. For imaging applications [4-6] where high peak power is the primary objective, impulsive THz generation is typically preferred, but for interference-based applications [2], continuous wave (CW) emission by photomixing is used because it offers narrow linewidth and hence better coherence. Here we investigate means to improve the energy efficiency and spectral coherence by utilizing internal reflection and double optical pulse excitation.

Fig. 4.1. Experimental setup for autocorrelation measurements of terahertz emission from a large area silicon PIN diode under double optical pulse excitation. Inset 1 illustrates the enhanced external power extraction and quasi-continuous wave emission from the diode by utilizing the internal reflection from the backside substrate. Inset 2 illustrates the selection of particular terahertz frequency through double optical pulse excitation.
The experimental setup is shown in Fig. 4.1. Collimated optical pulses from a modelocked Ti:sapphire laser are incident on a silicon PIN diode at an incidence angle of 70°. The illuminated area is 0.5 mm². The diode is biased under -50V and the intrinsic layer thickness is 50 μm to produces an electric field is ~ 10 kV/cm. The optical pulse width is compressed to 40 fs before illuminating the diode. It is first evenly split and recombined by a polarizing beam splitter (PBS) such that two pulses with different polarization are generated with a fixed time delay for double optical pulse excitation. The beam is then further divided and combined by the second PBS. By scanning the tunable delay stage, two identical THz pulses generated from the two paths interfere to provide their autocorrelation. Fourier transform of this time-domain information yields the THz power spectrum for double optical pulse excitation. The THz radiation is detected by a cooled silicon bolometer. From previous studies, it is known that the THz emission is primarily coupled into the substrate instead of exciting the surface [2]. However, we can enhance the THz power extracted from the surface by utilizing internal reflection from the backside metal. This is facilitated by the low THz attenuation in the silicon PIN diode. Collection of this extra power increases the external emission efficiency. At the same time, multiple pulses reflected from the substrate can be engineered to provide periodic pulses with frequency in THz range. A burst of pulses with fixed temporal separations, as in this case, act as a quasi-CW source with a narrower spectrum and better coherence. We further enhance certain THz frequency by double optical pulse excitation to achieve even narrower emission.

The power of THz radiation is measured with a calibrated Golay cell. Figure 4.2 (a) shows the measured THz power with various input optical powers. Using an average excitation power of 130 mW, a maximum of 2.8 nW of THz power is generated.
Figure 4.2 (b) shows the THz spectra obtained from the autocorrelation measurements. Power is concentrated on particular frequencies due to having a burst of pulses as opposed to a single pulse (dashed curve). The autocorrelation is further confirmed as shown in Fig. 4.3. We observe a burst of 5 evenly spaced peaks in the autocorrelation data with single optical pulse excitation. So, an optical pulse is creating three THz pulses, each separated by a delay of 9 ps, which corresponds to one round-trip time inside the diode. Collection of two additional THz pulses from a single optical pulse excitation increases the total extracted THz power. Based on the time
separation between the peaks, we estimate the diode thickness to be approximately 450 μm, which is in good agreement with the known device substrate thickness of 440 μm. The measured THz power spectrum is concentrated at 0.11 THz, 0.22 THz and 0.33 THz as shown in Fig. 4.2 (b). And the THz power depends on the optical pulse widths, as shown in Fig. 4.2 (c) inset, where the THz power decreases linearly with increasing pulse width without much change in the spectrum. This result highlights the advantage of quasi-CW generation compared with the CW photomixing approach.

We further show tuning of the THz spectrum by changing the pulse delays in double optical pulse excitation [7]. The enhanced frequency components are the fundamental frequency at 0.11 THz and its higher order harmonics. The fixed time delay in Fig. 4.1 is set to provide pulses with time delay corresponding to the THz frequency of interest. Figure 4.3 shows the autocorrelation of the emitted THz radiation and Fig. 4.4 shows the corresponding THz spectra. The figure shows the enhancement of particular terahertz peaks by double optical pulse excitations.

In conclusion, we have investigated tunable THz generation in silicon PIN diodes where internal reflection leads to enhanced energy efficiency, improved coherence and narrower linewidth. The THz emission efficiency can be further improved by modifying the intrinsic region to obtain higher electric fields with the same applied voltage [1,3]. In addition, higher power can be obtained using beam combining in a diode array.
Fig. 4.3. The autocorrelation traces of the terahertz signal with double optical pulse excitation.

Fig. 4.4. Terahertz power spectra with double optical pulse excitation at various pulse repetition periods. The graphs show the enhancement of particular terahertz peaks by double optical pulse excitations.
REFERENCES


Chapter 5

Giant tunable silicon based dispersive element for on chip real-time frequency-time mapping

Dispersive optical elements play critical roles in the operation of many high-speed optical systems. The ability to stretch and compress optical pulses using chromatic, or group velocity dispersion (GVD) forms the basis for chirped pulse amplification (CPA) – an indispensable technique for amplifying ultra-short optical pulses [1]. Today, the use of high-energy ultra-short laser pulses is pervasive in applications ranging from basic scientific research [2-4] to industrial material processing [5-7]. An even more ubiquitous use of GVD is dispersion compensation in high bit-rate optical communication networks, which form the backbone of the Internet. In this case, dispersion management is critical for controlling data impairment caused by optical dispersion and nonlinearities. Other important applications of GVD include intracavity dispersion compensation in modelocked lasers [8], time-stretch analog-to-digital conversion
and temporal imaging of optical pulses [11-14]. Beside communications, GVD is also essential in biomedical application which is the focus of this thesis. Important applications include time-wavelength spectroscopy [15,16], and high-speed optical imaging [17, 18] which utilizes GVD to transform the signal or information from frequency domain to time domain for real-time high speed data acquisition. Unfortunately, devices used to generate large chromatic dispersion exhibit significant tradeoffs between GVD and other performance metrics. Pairs of diffraction gratings commonly used in CPA laser systems can offer large dispersion but at the expense of footprint because the GVD scales approximately linearly with the grating separation. Dispersion compensating fiber can provide a large amount of GVD in a small footprint, but has limited power-handling capability due to its large effective optical nonlinearity. More importantly, the use of dispersion compensating fiber is limited to the near-infrared telecommunication wavelength (~1550 nm) band. At other wavelengths, commercial optical fibers have large losses and limited dispersion. Chirped fiber Bragg gratings can produce larger GVD with smaller footprint than dispersion compensating fiber, but suffer from similar power handling limitations due to their single mode operation, and typically operate over a limited bandwidth within the telecommunication band. In addition, much higher dispersion parameter is needed when we want to minimize device dimension from fiber down to chip-scale for potential application such as lab-on-chip. On-chip dispersive device have only limited performance up till now and the GVD are either based on engineering the waveguide dispersion [19] or modulated sidewall Bragg grating [20] which are analogous to principle used in dispersion compensation fiber and fiber Bragg grating respectively. To meet the required GVD on-chip for practical use, dispersion provided by waveguide engineering is limited and require a meter-long propagation length which is not possible to be fabricated. The dispersion provided by sidewall modulated
Bragg grating is high enough such that only centimeter-long waveguide is needed. However, the waveguide requires nanometer precision which can only be provided by e-beam lithography. The total length is limited by the e-beam ditching. Thus, a new idea which gives huge dispersion in a small footprint and easy of fabrication insensitive is essential.

Here, we propose and demonstrate a new type of dispersive device that combines the spatial dispersion of diffraction gratings with the modal dispersion of a multimode waveguide to overcome the footprint, power, wavelength, and bandwidth limitations of conventional dispersive elements. This device, which may aptly be called the chromo-modal dispersion device (CMD), maps different wavelength components of an optical waveform to distinct modes of a multimode waveguide. The CMD then exploits the large modal dispersion inherent in a multimode waveguide to create chromatic dispersion (GVD).

5.1. Chromo-modal dispersion

An illustration of CMD on a multimode fiber is shown in Fig. 5.1. A broad-bandwidth optical pulse is incident on a pair of parallel diffraction gratings, which spatially disperses and collimates the pulse spectrum. The dispersed pulse propagates through a lens, which focuses the various spectral components at a range of angles relative to the optic axis. By placing the tip of a multimode waveguide or fiber (as shown in Fig. 5.1.) at the focus of the lens, each spectral component of the pulse excites a different mode of the optical fiber. Modal dispersion in the fiber imparts either normal or anomalous GVD to the pulse, depending on the transverse alignment of the lens relative to the spatially dispersed spectrum.
The CMD’s use of large-core multimode fiber drastically reduces the undesirable effective optical nonlinearity, making it capable of handling much larger peak optical power than singlemode fiber-based dispersive devices. Additionally, the total footprint and size of the gratings used by the CMD are orders of magnitude smaller than grating-based Treacy-geometry dispersive elements with similar dispersion. Finally, since the CMD’s operation is based on the wavelength-invariant principles of ray optics, it can operate at any wavelength for which multimode and transparent waveguides can be constructed.

In this first part of this chapter, we will demonstrate our CMD concept based on a multimode fiber at 800nm. In the second part of the chapter, preliminary results of CMD using on-chip silicon waveguide are demonstrated at 1550nm and this lead us to the possibility lab-on-chip applications such as on-chip time wavelength mapping in the future.

Fig. 5.1. Schematic diagram of the chromo-modal dispersion (CMD) device. Light is incident upon two parallel plane gratings, which disperse and collimate the optical spectrum. Angular dispersion is then applied to the spectrum using a lens. The input facet of a multimode fiber is placed at the lens focus such that the various spectral components are coupled into different fiber modes. The figure inset illustrates how different spectral components are coupled into and propagate in the multimode fiber. The dashed line represents the optic axis of the fiber. Although the configuration shown provides anomalous dispersion, the CMD can be tuned throughout both the anomalous and normal dispersive regimes by adjusting the alignment of the lens and fiber relative to the spatially dispersed spectrum.
5.2. Demonstration on a multimode fiber

As a first demonstration of the CMD’s functionality, we characterize its dispersion in the 800-nm wavelength band. This is an especially important wavelength band for many biological applications because of its large penetration depth in tissues, which decreases at shorter wavelengths due to Rayleigh scattering and at longer wavelengths due to water absorption. Furthermore, this wavelength band is accessible by powerful and tunable modelocked Ti:sapphire lasers, which should enable the CMD to find use in a variety of biomedical, spectroscopic, and material processing applications.

In this work, we utilize two different implementations of the CMD (see Methods in Sec. 5.6), making use of the fact that the device parameters (alignment, grating groove density, grating spacing, lens numerical aperture, fiber length, fiber numerical aperture) can be adjusted to modify its dispersive properties. To measure the dispersion of the CMD, we utilize the modulation phase shift method [21]. Using a radio-frequency lock-in amplifier as a phase-sensitive detector, we measure the relative phase shift between a 30 MHz amplitude-modulated continuous wave Ti:sapphire laser signal after it has traveled through CMD, and the reference 30-MHz signal used to drive the electro-optic modulator. By tuning the laser wavelength, we measure the wavelength-dependent phase delay of the CMD. To calculate the relative time delay (in ps) as a function of wavelength, we use the equation

\[ \Delta \tau_\lambda = -\frac{\phi_\lambda - \phi_{\lambda_0}}{360 f_m} \times 10^{12} \]

where \( \Delta \tau_\lambda \) is the time delay of a given wavelength, \( \phi_\lambda \) is the relative phase of the signal at a given wavelength measured by the lock-in amplifier, and \( f_m \) is the amplitude modulation frequency. \( \lambda_0 \) is an arbitrary wavelength, which is chosen to have zero relative phase delay. A
3\textsuperscript{rd}-order polynomial least-squares fit of the time delay data is used to calculate the dispersion parameter $D_\lambda$ using

$$D_\lambda = \frac{d\Delta \tau_\lambda}{d\lambda}.$$ 

The results of these measurements and calculations are shown in Fig. 5.2.

Additionally, to demonstrate the tunability of the CMD, we record several relative delay curves at different lateral positions of the focusing lens relative to the diffraction gratings. By maintaining alignment of the fiber tip at the lens focus, it is possible to tune the entire delay curve including zero-dispersion wavelength of the CMD. For the three alignment positions measured, the zero-dispersion wavelengths are 788.7 nm, 799.3 nm and 808.5 nm (curves 1, 2 and 3, respectively).

![Fig. 5.2. (a) Relative delay and (b) dispersion parameter for three different alignments of the CMD. Translating the lens and fiber facet in the horizontal direction relative to the incident spatially dispersed spectrum shifts the CMD’s zero dispersion wavelength.](image)

The CMD achieves large, tunable GVD by mapping the optical spectrum into modes of a multimode fiber. To confirm that different wavelengths occupy distinct spatial modes in the fiber, we measure the far-field mode profiles of the CMD output fiber as a function of excitation wavelength. Using the alignment position that yields the dispersion curve #1 in Fig. 2(b), we
record the mode profiles at various wavelengths using a CCD array placed 5 mm from the output tip of the fiber. The mode profiles for various excitation wavelengths are shown in Fig. 3(a).

Fig. 5.3. Mode profiles at the output of the CMD fiber for alignment position #1, measured 5 mm from the output fiber facet. (a) Mode profiles recorded using various excitation wavelengths. (b) Mode profile using 745-nm excitation after propagating through 20 m of 200-μm core diameter fiber. (c) Mode profile using 745-nm excitation after propagating through 35 m of the same fiber. Blurring in the output mode profile indicates the coupling length has been reached. The scale bar in all cases is 1 mm.

Ideally, the CMD operates on the principle that individual wavelengths excite distinct modes of the multimode waveguide, each of which possesses a different propagation constant. To achieve maximum dispersion for a given waveguide length, it is important to avoid mode coupling which can result from random inhomogeneity or bends in the fiber [22, 23]. As a result, the coupling length sets an upper bound on the length of the fiber and hence the maximum dispersion. Fig. 3(b) and 3(c) show the far-field mode profile of the CMD measured at the end of a 20-m and 35-m lengths of fiber, respectively (using 745-nm excitation). As shown in Fig. 5.3(c), in the 35-m fiber the mode profile changes dramatically, indicating that the coupling length has been exceeded.
To measure the dispersive effect of the CMD on an optical pulse, we propagate pulses from a modelocked Ti:sapphire laser through the device, and monitor the fiber output using a fast photodiode (see Methods). The Ti:sapphire laser emits pulses ($t_{\text{FWHM}} = 80$ fs, $\lambda_{\text{center}} = 812$ nm, $\Delta\lambda = 15.8$ nm, repetition rate = 90.8 MHz) which are coupled into the CMD. By varying the alignment, the dispersion is tuned such that the entire pulse bandwidth lies in the anomalous ($D_l > 0$) dispersion regime (Fig. 5.4(a)). As shown in Fig. 5.4(b), the time-domain output’s full-width at half-maximum and full-width are 634 ps and 1.9 ns, respectively.

![Fig. 5.4. Time-domain profile and spectrum of an 80-fs optical pulse after propagation through the CMD. (a) Spectrum of the input optical pulse and the CMD dispersion parameter. (b) Measured time-domain output (red) and calculated time-domain profile for the measured input spectrum and dispersion curve.](image)

For comparison to the experimentally measured time-domain data, we simulated the dispersive propagation of a Gaussian optical pulse through the CMD. The simulated Gaussian pulse and CMD dispersion were chosen to have the same values as those used in the time-domain measurements. We performed the simulation by adding the appropriate spectral phase to the pulse in the frequency domain, and performing an inverse Fourier transform to return the data to the time domain. The calculated time-domain profile is denoted by the dashed curve in Fig. 5.4(b), which is in excellent agreement with the experimental measurement. Differences between
the experimental and calculated plots of the time-domain data shown in Fig. 5.4(b) are attributed to both the use of an ideal Gaussian pulse spectrum and the use of a fit to the relative delay data, instead of the experimental values shown in Fig. 5.4(a) and 5.2(a).

Loss measurements were performed by monitoring the average power before and after the grating pair, and at the output of the CMD fiber. Using a modelocked Ti:sapphire laser with the CMD in alignment position #1, the diffraction gratings’ loss was 6.0 dB. The loss due to the lens and fiber was 1.52 dB. In this case, the large grating loss is due to the use of gratings not optimized for operation at 800 nm, and can be reduced to approximately 0.4 dB by using high-efficiency dielectric gratings [24].

To demonstrate another application of the CMD in a wavelength range outside the telecommunication band, we performed real-time, single-shot [9, 25, 26] time-wavelength atomic absorption spectroscopy of the D2 (5^2S_{1/2} \rightarrow 5^2P_{3/2}) transition in rubidium vapor. In this experiment, near-transform-limited pulses from a modelocked Ti:sapphire laser (\(\lambda_{\text{center}} = 780\) nm, \(\Delta\lambda = 35\) nm, repetition rate = 90.8 MHz) were transmitted through a 225-mm path length rubidium vapor cell before entering a CMD optimized for anomalous dispersion over a 4-nm bandwidth (see Methods). The optical spectrum transmitted through the CMD and the associated time-domain signals are shown in Fig. 5.5(a) and (b), respectively. Spectra are recorded by an oscilloscope in this fashion every 11.01 ns, which is orders of magnitude faster than is possible using conventional array-detector- or scanning-based spectrometer. In addition, the effective spectrometer shutter speed, or integration time, of this measurement is equal to the optical transit time through the rubidium cell, which is approximately 750 ps. In this spectroscopy experiment, the Doppler-broadened D2 transition in rubidium generates an absorption dip in the transmission spectrum, which is observed in the time-domain profile. While it is clear from the experimental
data that the CMD produces a time-wavelength mapping, it is also important to note that in this case, the amount of dispersion is sufficient to satisfy the uncertainty principle requirement for generating a dispersive Fourier transform of the wavelength-domain signal [27, 28].

**5.3. Discussion**

Fundamentally, the mode coupling length and the numerical aperture (NA) of the multimode fiber determine the maximum dispersion of the CMD. For fiber lengths shorter than the coupling length, modal dispersion dictates that the maximum output pulsewidth of the CMD can be approximated by [29]

\[
\Delta \tau_p = L \frac{N_2}{c} \left( \frac{n_1 - n_2}{n_2} \right)
\]
where $\Delta \tau_p$ is the output pulsewidth, $L$ is the fiber length, $n_1$ and $n_2$ are the material refractive indices of the fiber core and cladding, respectively, and $N_2$ is the group index of the cladding. For an input pulse with a center wavelength of 800 nm propagating in a 20-m long pure silica core fiber with a 0.37 NA, the maximum output pulsewidth is approximately 3 ns, assuming all modes of the fiber are excited. A noteworthy feature of CMD is that the output pulsewidth is independent of the input optical bandwidth, provided that it is sufficiently spatially dispersed to intercept the entire NA of the fiber after being focused by the lens. Thus, the CMD is capable of stretching small-bandwidth pulses by a large amount in a small footprint. As an example, we consider a device consisting of two 30-mm wide, 2200 line/mm gratings spaced by 10.8 cm (aligned for an angle of incidence of 50 degrees from the grating normal), a 25.4-mm diameter 0.37-NA lens, and 20-m of 0.37-NA step-index multimode fiber. For an optical pulse of 1-nm bandwidth centered at 800 nm, this CMD can provide approximately $\pm 3000$ ps of group delay, stretching an 800-nm center wavelength, 1-ps pulse to a duration of 3 ns. To achieve the same dispersion, two gratings with the same groove density placed in a double-pass Treacy geometry [30] would have to be spaced by 1400 cm, and one grating would have to be greater than 25 cm in width. The ability to impart large normal or anomalous dispersion over a small optical bandwidth makes the CMD invaluable for applications in picosecond CPA laser systems, which traditionally employ large-footprint, grating-based stretchers and compressors [30-32]. Finally, in contrast to single mode dispersive fibers, the CMD is much better at accommodating high optical power without introducing distortions caused by optical nonlinear interactions. This is a direct result of the large cross section of the multimode fiber, which reduces its effective nonlinearity. This feature can be further improved by replacing the fiber with a hollow multimode waveguide such as those previously reported [33].
As shown in Fig. 5.3(a), the spectrum of the input pulse defines the output mode profile, which can differ significantly from the input mode. To circumvent this mode-conversion in applications such as CPA, in which the pulse stretcher and compressor spatial modes are critical, one may use a length of fiber in the CMD slightly longer than the coupling length. Coupling the modes using an extended length of fiber generates an output that consists of the sum of all fiber mode profiles (Fig. 5.3(c)), as opposed to only modes excited initially (Fig. 5.3(b)). Higher-order modes can subsequently be suppressed, e.g. using fiber coiling [34] or other techniques [35, 36] which reduce the output beam’s $M^2$ value.

For applications in which the output spatial mode is less important than the time-domain profile, such as in the time-wavelength spectroscopy experiment reported here, it is imperative that the CMD fiber length remains shorter than the fiber coupling length. As shown in Fig. 5.5, the CMD generates a wavelength-to-time mapping of the optical spectrum used to probe the D$_2$ transition in rubidium vapor in real-time. Nonlinearity in this mapping can be attributed to the shape of the CMD’s dispersion curve, which is fundamentally the same as that shown in Fig. 5.2(b), albeit over a bandwidth of $\sim$4 nm. This nonlinearity in the CMD’s dispersion slope is static and deterministic, and hence, it can be compensated using digital post-processing [37-39]. Thus, knowledge of the CMD dispersion combined with the appropriate laser source and detector will enable high spectral resolution, real-time spectroscopy at any wavelength for which the fiber is transparent or a hollow waveguide can be constructed.

5.4. Methods for multimode fiber

The CMD used in all experiments except for absorption spectroscopy is constructed from a pair of 900-grooves/mm ruled diffraction gratings spaced 14 cm apart, a $f = 35$ mm, 25.4-mm
diameter, antireflection-coated achromatic doublet lens, and a 20-m long, 200-µm diameter core, 0.37-NA step-index pure silica core, polymer cladding multimode optical fiber. The lens and fiber are mounted on the same linear translation stage, such that as the transverse alignment of the lens is altered, the fiber tip remains at the focus of the lens. Modulation phase shift measurements were performed using a custom-built continuous wave Ti:sapphire laser. The laser linewidth full width at half maximum for each measurement point was less than 0.20 nm. The RF lock-in amplifier time constant was set to 1 second for all measurements. Time-domain measurements were performed by connecting an unamplified InGaAs photodiode with a 25-ps impulse response time to the output of the CMD. Electrical signals from the photodiode are recorded using a 50-gigasample/s, 16-GHz analog bandwidth oscilloscope. Real-time atomic absorption spectroscopy was performed on a heated 75-mm length rubidium vapor cell (natural isotope mixture) in a triple-pass geometry. A narrow bandpass filter (Δλ = 3 nm) centered at 780-nm was placed after the rubidium cell to limit the optical bandwidth entering the CMD. The CMD configuration used in this experiment consisted of two 30-mm length, 2000 grooves/mm ruled diffraction gratings spaced by 28 cm. The lens and fiber were the same as used in all other experiments.

5.5. Preliminary demonstration on a Si slab waveguide

The high index contrast of silicon waveguide allows more extreme ray path to exist and thus a much smaller footprint is possible to achieve the same dispersion demonstrated in the previous section (800 ps time delay within a 15 nm bandwidth).
As a proof-of-principle demonstration, we use a double side polished silicon wafer as a slab waveguide. Figure 5.6 shows the expected ray path of the long and short wavelength components in order to achieve highest dispersion with shortest waveguide length. Basically, short wavelength components (blue) are designed to travel slower than long wavelength components (red). All of them are confined inside the waveguide by total internal reflection with the shortest wavelength component travels at an angle close to the critical angle (the extreme path which have longest traveling path). Since the light is coupled from air into a silicon which have a much higher refractive index, the input facet need to be polished such that the light can be coupled into the extreme path at normal incidence without being refracted. As a result, the waveguide facet is polished at 16.6° as illustrated in Fig. 5.6 and also as shown in Fig. 5.7.

![Si slab waveguide](image)

**Fig. 5.6.** Illustration of the different spectral components coupled into and propagate in the Si slab waveguide. This configuration represents normal dispersion. The input facet is polished such that the shortest wavelength component (blue) travels and is confined by total internal reflection at an angle close to the critical angle.

![Fig. 5.7. Schematic of the multimode waveguide made by double side polished wafer. The length is 9 cm and the thickness is 400 μm. The inset shows the polished surface of the input facet.](image)
Using modulation phase detection measurement, the preliminary result is shown in Fig. 5.8.

![Graph showing relative delay of Si-based CMD measured by modulation phase detection (red) and calculated based on experiment parameters (blue).](image)

Fig. 5.8. Relative delay of the Si-based CMD measured by modulation phase detection (red). Relative delay of the Si-based CMD calculated based on experiment parameters. (blue).

We can see that a maximum of 170ps/cm can be obtained. The theoretical calculation shown in Fig. 5.8 is based on the setup parameters used in our experiment. The data are reasonably matched with the calculation and the slight mismatch is probably due to the flatness of the polished surface (custom done by hand). Same dispersion used in the previous time-wavelength spectroscopy measurements can be achieved by 4.5 cm long waveguide (same order as the one provided by modulation sidewall Bragg grating but much easier to be achieved in term of fabrication). We remark that the bandwidth of the devices depends on our setup (5 nm is demonstrated but larger bandwidth can be used with same dispersion assuming same incidence angle span is used). The next step is to make use of the all the modes in silicon and we can further minimize the length. Figure 5.9 shows an illustration of a proposed on-chip device.
5.6. Conclusion

In summary, we have demonstrated and characterized a compact device that produces large and tunable group velocity dispersion. This has been achieved by transforming the large modal dispersion of a multimode fiber into chromatic dispersion. The amount of dispersion as well as the zero dispersion wavelength is tunable and produces normal or anomalous dispersion, with maximum dispersion limited only by the coupling length and NA of the multimode waveguide used. As a proof of concept demonstration, we stretch a sub-picosecond optical pulse to nearly 2 ns in 20 m of multimode optical fiber. By adjusting the grating groove density and separation, The CMD can be configured to accommodate larger or smaller bandwidth pulses, while maintaining the same output pulse width. To demonstrate one of its utilities, we demonstrated single-shot atomic absorption spectroscopy of rubidium vapor. The large dispersion and small footprint of the device make the CMD potentially useful for on-chip dispersion compensation using optical components such as integrated gratings and planar multimode waveguides. We have also demonstrated CMD on a silicon waveguide which can allow us to achieve much smaller device dimension. The CMD’s physical compactness,
combined with the magnitude and tunability of its dispersion suggest its use as an additional versatile tool for pulse stretching or compression in a variety of applications in which the capabilities of waveguide dispersion or grating-based dispersive elements will not suffice.
REFERENCES


Chapter 6

Tailoring the biodegradability of porous silicon nanoparticles

Recently, there has been a great deal of attention focused on nanoparticles for use as drug delivery vehicles. The ability of functionalized nanocarriers to target cells for controlled drug release potentially improves delivery efficiency and specificity by orders of magnitude. A number of nanocarriers including liposomes [1], polymeric nanoparticles [2], iron oxide nanoparticles [3] and mesoporous silica nanoparticles [4,5], have been studied for this purpose. Porous silicon (PSi), with its attractive physical, chemical and optical properties [6,7], has been investigated for use in biomedical applications such as optical sensors [8], implantable devices [9], orthopedic tissue engineering [10] and recently for an autonomously functioning delivery platform [11]. Fabricated from PSi wafers, PSi nanoparticles (PSiNPs) are comprised of a nanocrystalline network, which provides a large surface-to-volume ratio for functionalization, as well as nano-size pores for drug loading. Additionally, PSiNPs are inherently fluorescent, biocompatible, biodegradable and non-toxic. These attractive physical and chemical
characteristics have enabled research on PSiNPs to result in their preliminary use as in vivo drug carriers [12,13].

6.1. Introduction

Among their various properties, the biodegradability of PSiNPs is of particular interest. PSiNPs degrade readily in biological environments to form silicic acid, which is a natural trace compound in humans [14]. PSiNPs with sizes in the range of 80 – 120 nm are desirable for drug delivery applications, as they are the right size to take advantage of the EPR effect and enhance tumor uptake while still being large enough to avoid renal clearance [12]. However, the in vivo degradation rate of such particles is often too fast for optimum drug delivery efficiency. To achieve efficient in vivo targeting of cells with high specificity, it is ideal for particles to exhibit blood circulation times of at least 12 hours [15]. In previous work on drug delivery using PSiNPs [13], the fabricated PSiNPs had a half-life (i.e. the time it takes for dissolved silicon content to reach 50% of the initial content) of approximately 10 minutes. The half-life is too short for in vivo usages and thus they reported the use of water to oxidize the particles and increase the half-life up to approximately 30 minutes for their in vivo demonstration. Thus, there is a need to study and control particle stability in biological fluids and to significantly increase particle half-life up to 10 hours to facilitate effective in vivo drug delivery.

Two conventional physical approaches to improve PSi stability have been applied to PSiNPs: thermal oxidization and thermal hydrocarbonization [16,17]. Thermal hydrocarbonization completely halts the in vivo PSi degradation process, while thermal oxidation slows the degradation kinetics [16]. Chemical-based stabilizing methods include functionalization using polymers (e.g. polyethylene glycol (PEG)) [18], and drug loading [13].
All these techniques have been employed to stabilize PSi microparticles, but only limited systematic studies of these approaches exist [18]. Furthermore, PSiNPs, which have a much greater surface-to-volume ratio than PSi microparticles, exhibit much faster degradation rates than their larger counterparts. Due to the increased degradation rate of nanoparticles as compared to microparticles, lifetime study dedicated to nanoparticles is essential and prolonging their lifetime is critical to the successful use of them for in vivo applications. As we mentioned in the second paragraph, Park et al. [13] have reported oxidized PSiNPs with half-life up to 30 minutes. Their particles are stabilized by incubating in water for oxidation. Angelis et al. [12] also performed thermal oxidation on PSi wafer to stabilize the photoluminescence and thus possibly prolong the particle half-life. Although half-life values are reported, there is no systematic study on the effect of these stabilizing processes.

Here, we demonstrate two techniques to increase the lifetime of PSiNPs over the range of time from minutes to nearly 10 hours, which is comparable to that of both PSi microparticles [18] and sol-gel silica microparticles [19] in phosphate buffered saline (PBS) solution which is a close mimic of a basic biological fluid. In this study, we investigate two thermal oxidation processes in details: pre-oxidation using rapid thermal processing (RTP) and post-oxidation using hot aqueous bath which are inspired by the work done by Part et al. and Angelis et al. respectively. These methods enable engineering of the degradation lifetime of PSiNPs with dimensions of 80 to 120 nm. We also propose and study silica coating as an alternate method to prolong the lifetime of PSiNPs. This method retains the properties of PSiNPs that are pore-dimension sensitive, such as photoluminescence. The silica surface coverage technique is also attractive for incorporating other fluorescent labels into the PSiNPs, as fluorescent proteins or rare-earth fluophores embedded in silica exhibit enhanced photostability within an aqueous environment.
While each technique has distinct merits, both thermal oxidation and silica coating of PSiNPs result in a nanoparticle with a silica surface, which is convenient for bioconjugation [4,5].

In the first section of this chapter, we investigate two techniques to thermally oxidize PSiNPs. The effects of both processes are discussed and an optimized combination of these two procedures is adopted to provide stable particles. In the second section, we study direct silica coating of PSiNPs as an alternative method to prolong the lifetime of PSiNPs.

6.2. Fabrication and characterization of porous silicon nanoparticles (PSiNPs)

6.2.1. Thermally oxidized PSiNPs (TOPSiNPs)

Fig. 6.1. Fabrication process of the thermal oxidized porous silicon nanoparticles (TOPSiNPs). Pt: platinum, HF: hydrofluoric acid.

PSi thin layers are formed by electrochemical etching of silicon wafers [6]. A 4” diameter p-type silicon wafer (orientation <100>, boron-doped, resistivity 1.2 – 1.9 mΩcm) is etched in a two half-cell tank filled with 49% HF:ethanol (1:3) electrolyte using a constant current density of 200 mA/cm² for 2.5 minutes, as shown in Fig. 6.1. This results in a 20 µm thick of PSi layer
formed on top of the silicon wafer. The wafer is then thermally oxidized at 800°C using RTP in oxygen (O2 flow rate is 40 sccm). Oxidation times are varied from 0 to 6 minutes. A free-standing PSi layer is obtained in a 49% HF:ethanol (1:90) electrolyte using a current pulse density of 1 mA/cm². After the lift-off process, the hydrophilic PSi surface is used to confirm that oxide formed by RTP still remains. This means that the HF electrolyte used in the lift-off process does not remove the surface oxide totally and thus the effect of RTP is retained. The freestanding PSi layer is collected in ethanol and sonicated using an ultrasonic bath cleaner for 24 hours. Sonication is chosen for the sake of simplicity, although ball milling is another option for creating PSi particles. The particle shape obtained using sonication is less regular compared with ball milling which gives regular ball-shaped particles [17]. It is known that particle shape affects the efficiency of drug delivery [15] but it is not the focus of the study presented here. But the techniques we discuss here for tailoring the biodegradation rate is applicable to PSiNPs prepared by both sonication and ball milling. To obtain particles of diameter smaller than 200 nm, the dispersed particles are filtered using a 0.2 μm membrane filter. To remove particles smaller than 50 nm, the sample is first centrifuged at 14,000 rpm for 20 minutes, and the supernatant is removed carefully. After the pellet is resuspended in deionized (DI) water, the centrifugation and suspension process is repeated several times to wash the particles. Finally, for post-oxidation treatment, the dispersed particles are heated in a 60°C water bath. Post-oxidation times range from 0 to 42 hours. For 1 g of free standing PSi thin film, 2 mg of PSiNPs can be obtained after final processing. This yield of 0.2% is largely due to the fact that the filtering removes the majority of particles, which have diameters larger than 200 nm. The yield can be further improved by using longer sonication times or using another particle formation method such as ball grinding.
6.2.2. Silica-coated PSiNPs (SCPSiNPs)

A conventional sol-gel process, traditionally used for synthesizing silica particles, is used for silica coating of the PSiNPs [19]. Since the above-prepared PSiNP seeds are conveniently terminated with hydroxyl groups, no other surface preparation is required. Initially, 2.5 mg of PSiNPs dispersed in 70 µL of DI water are added to 1 mL ethanol (99%) and 20 µL NH4OH (30% w/v aqueous solution). To initiate the reaction, 60 µL of 99% tetraethyl orthosilicate (TEOS) is added to the mixture, which is then sonicated for 10 to 30 minutes. Centrifugation is performed at 14,000 rpm for 20 minutes to separate the silica-coated PSiNPs (SCPSiNPs) from small silica particles that spontaneously form during the reaction. The pellet of SCPSiNPs is resuspended in DI water. Finally, the particles are washed with DI water several times and kept in DI water.

6.2.3. Particle characterization

Dynamic light scattering is used to characterize the size distribution and electrophoretic light scattering is used to measure the Zeta potential of the particles in DI water (pH7). Both TOPSiNPs and SCPSiNPs fabricated have a mean size of 100 nm with a standard deviation of 20nm. For all samples of both TOPSiNPs and SCPSiNPs, the Zeta potential is measured to be -6 mV. The pore size of these particles is estimated using transmission electron microscopy image analysis. For TOPSiNPs, the average pore diameter is approximately 5 nm, while for SCPSiNPs the average pore diameter ranges from 0 to 5 nm, depending on the silica coating time. The calibrated silica coating rate is 0.2nm/min. Longer coating times correspond to smaller pore diameters. Using gravimetric analysis, the porosity of the PSi formed is estimated to be around 70%. Inductively-coupled plasma optical emission spectrometry (ICP-OES) is used to evaluate
the concentration of PSiNPs in solution. A known concentration of an ICP silicon standard is used to draw a standard curve.

6.2.4. Degradation study in simulated biological conditions

To measure the degradation rate of the various types of PSiNPs, we monitor the dissolved silicon concentration in PBS solution at a constant temperature of 37 °C. For each sample, 0.02 mg of PSiNPs is incubated in 2 mL of PBS solution, which corresponds to an initial silicon concentration of 10 ppm. At each time point, 400 µL of the sample solution is removed from the 37°C oven and centrifuged at 14,000 rpm for 20 minutes. By collecting the supernatant, the dissolved silicon is separated from the solid nanoparticles. 20 µL of 1M NaOH is added to the removed supernatant for digestion at 37°C for 12 hrs. The concentration of dissolved silicon is measured using ICP-OES after dilution to a total volume of 4 mL using 5% HNO3. The results are normalized with respect to a fully digested sample prepared separately (An additional 4 µg of PSiNPs is taken from each sample and dispersed in 400 µL of PBS, digested by 20 µL of 1M NaOH for 12 hours at 37°C).

6.3. Results and discussions

6.3.1. Effect of thermal oxidation on the degradation of TOPSiNPs

Here we investigate thermal oxidation as one avenue to increase the half-life of PSiNPs while avoiding alteration of their chemical properties. Both RTP of the PSi wafer and post-oxidation in a hot aqueous bath are expected to improve the stability of PSi particles [12,13]. Without oxide passivation, PSi microparticles are known to be unstable in biological environments, and the degradation rate is believed to depend on pore size [18]. However, for
particles with diameters on the order of 100 nm, which commonly have pore sizes ranging from 5 to 20 nm, stability does not strongly depend on pore size [13]. Thus, adjusting the pore size of PSiNPs by altering fabrication parameters is not the best route to improve stability.

Figure 6.2 (a) shows a representative transmission electron microscope (TEM) image of PSiNP without any oxidation treatment. The particle diameter is 100 nm with pore diameter of approximately 5 nm. The half-life of the native PSiNPs is 10 minutes, as shown in Fig. 6.2 (c), consistent with the results of Park et al. [13]. These unmodified particles are the baseline standard for the following systematic study of thermal oxidation of the PSiNPs.

6.3.1.1. Effect of Rapid Thermal Processing (RTP)

RTP is performed on the PSi wafer before removal from the substrate. The processing time is varied from 0 to 6 minutes, while the process temperature is held constant at 800oC. Using TEM, we imaged particles dried on a Formvar-coated copper grid. Figure 6.2 (b) shows a TEM image of a TOPSiNP after RTP for 90 seconds. As compared to Fig. 6.2 (a), the dot-shaped pore structure changes to line-shaped as a result of RTP. Figure 6.2 (c) shows the effect of RTP time on the degradation of TOPSiNPs. No post-oxidation was performed prior to obtaining the data shown in Fig. 6.2.
As mentioned before, the biodegradability of PSiNPs with pore size in the range of 5 – 20 nm is not highly dependent on the pore dimension. This result suggests that the change in pore morphology and size from dots of 5 nm to line structures of 5 to 20 nm is not the dominant factor responsible for the prolonged stability in PBS. In this case, the improvement in stability is expected to be a result of the inert silica layer formed during RTP, which protects the silicon core from degradation. However, the fact that the possibility exists to engineer the pore morphology using fabrication process control, which is decoupled from the degradation time, provides control over both the loading and releasing properties of these particles. For RTP times longer than 90...
seconds, the half-life of TOPSiNPs increases to a value of 60 minutes, for a RTP time of 180 seconds. The increase in half-life depends on the thickness of the silica surface layer, which increases with oxidation time. However, it is evident that the particle half-life eventually plateaus. Further oxidization beyond 180 seconds does not improve the stability, and the half-life reaches a constant value. We remark that further oxidization beyond 300 seconds results in a half-life value in the range of 50 – 60 minutes and that the slight decrease in half-life for a RTP time of 300 seconds is within the experimental error. There are two potential explanations of this effect. First, the silicon oxidation rate is not constant but depends on the thickness of the silica [21]. As time increases, the oxidation rate is reduced as the thicker oxide inhibits the amount of oxygen diffusing to the buried Si/SiO2 interface. Second, for particles that have half or more of their volume comprised of silica, the core volume fraction of silicon does not influence the half-life measurement. The overall shape of the lifetime curve will be affected by this parameter, but the half-life value is not.

6.3.1.2. Effect of 60°C post-oxidation process

We also study the post-oxidation processing of particles in a hot aqueous bath. This simple processing technique employs a heated ultrasonic bath to grow oxide layers on the particle surface. In these experiments, PSiNPs are treated with 90 seconds of RTP and dispersed in DI water, before being subjected to this additional oxidation process. The particles are kept in a 60°C water bath for times ranging from 0 to 48 hours. Figure 6.3 (a) shows a TEM image of the particles before and after post-oxidation using this process. Figure 6.3 (b) shows the effects of post-heating time on the degradation of TOPSiNPs. Similar to the results obtained using RTP, the stability increases with processing time and reaches a steady-state value after 18 hours of
treatment. Again, the slight variation in the half-life of TOPSiNPs after 18 hours of post-oxidation treatment is a result of experimental error.

![Image of TEM images and degradation kinetics](image)

Fig. 6.3. (a) Representative TEM images of TOPSiNPs before and after post-oxidation in a 60°C water bath for 18 hrs. (PSiNPs are oxidized using RTP for 90s) (b) Degradation kinetics of TOPSiNPs with various post-oxidation times (90s RTP is included) showing that the blood retention time can be extended by increasing post-oxidation duration until the optimized post-oxidation treatment of 18 hours.

**6.3.1.3. Optimized process of TOPSiNPs**

As shown in Fig. 6.4, the maximum half-life of TOPSiNPs fabricated using aqueous post-oxidation without using RTP is 50 minutes. This value is similar to that of TOPSiNPs fabricated using RTP only, as shown in Fig. 6.2 (c). When the porous silicon layer is removed from the substrate and sonicated, the unoxidized silicon surface attaching the particles to the substrate becomes exposed. This exposed surface leads to rapid biodegradation if there is no post-oxidation. In contrast, the post-oxidation process uniformly oxidizes the silicon surfaces, which
should prolong particle stability. Although the use of RTP is expected to result in particles with a shorter half-life than those fabricated using aqueous post-oxidation, these two techniques result in similar improvements in particle stability. This suggests that the oxide formed by RTP may yield better protection against degradation than that formed by post-oxidation. If the unoxidized region can be protected by post-oxidation, the half-life can be further improved. Using 18 hours of post-oxidation described in Sec. 3.1.3 (steady-state value), we vary and obtain the RTP time that yields the best particle stability. Figure 6.4 shows the combination effect of RTP time and 18 hours of aqueous post-oxidation at 60°C on the degradation of TOPSiNPs. The optimal resulting half-life, obtained using both 90 seconds and 180 seconds of RTP, is consistent with the steady-state RTP time obtained in Sec 3.1.2. The optimized TOPSiNPs half-life improves from 10 minutes to 3 hours – a value that is appropriate for PSiNPs intended to be used in in vivo drug delivery studies [13].

![Figure 6.4](image_url)

Fig. 6.4. Degradation kinetics of TOPSiNPs with various RTP times (post-oxidation at 60°C for 18 hours is included) showing that the optimized thermal oxidation treatment by varying the RTP duration with optimized post-oxidation duration. Using thermal oxidation, the half-life of PSiNPs can be varied from 10 minutes to 3 hours.
6.3.2. Effect of silica coating on the degradation of SCPSiNPs

Rather than oxidizing the silicon particles, which consumes the PSi core, the particle half-life can be improved using silica coating. In this process, a layer of silica is chemically grown on the surface to protect the PSi from degradation, but designed not to fill the pores, which are needed for drug loading. In this case, no oxidization is performed on the particles except when the PSi is exposed to air, and a native oxide forms. Hydroxyl surface termination of the PSiNP surface acts to seed silica formation by the sol-gel process discussed in Section 2.2. Figure 6.5 (a) shows TEM images of SCPSiNPs with various coating times. The particle morphology appears similar to uncoated particles after 10 minutes of coating, but after 30 minutes, the particle pores are filled and the shape becomes rounded. Figure 6.5 (b) shows the degradation kinetics of the SCPSiNPs. In this experiment, equal amounts of PSiNPs were used in all trials. Due to the addition of silica from the TEOS, the total silicon content of the SCPSiNP samples increases with longer coating times. Silica coating for 10 minutes increases particle stability from 10 minutes to 8 hours. However, coating the particles for 30 minutes reduces the total solubility to 60% of the coated particles. This is due to the fact that a longer coating time results in a thicker silica shell and thus, greater total silicon content per volume. The total concentration of silicon in this case is measured to be 100 ppm, which is greater than the solubility of silica in water at 25°C (70 ppm) [18]. In this case, the concentrations of silicic acid and amorphous silica reach equilibrium before the silica protective layer is fully consumed, which halts further dissolution of the particles.
Further improvement

PSiNPs exhibit rapid degradation in biological fluids resulting in limited drug delivery efficiency. The prolonged half-lives of TOPSiNPs and SCPSiNPs are attributed to the presence of a silica surface covering the particles. Once the silica surface is dissolved, the PSi core is exposed and degrades readily. As shown in this study, the degradation rate depends on the thickness of silica. The thickness increases with longer oxidation time in the case of TOPSiNPs and longer coating time in the case of SCPSiNPs.

Fig. 6.5 (a) Representative TEM images of PSiNPs with no silica coating, 10 minutes of silica coating and 30 minutes of silica coating. (b) Degradation kinetics of silica coated PSiNPs (SCPSiNPs) with various silica coating times (no thermally oxidation is included) showing that the retention time can be extended by increasing the silica coating time. Using silica coating, we can extend the half-life up to 8 hours.

6.3.3. Further improvement

PSiNPs exhibit rapid degradation in biological fluids resulting in limited drug delivery efficiency. The prolonged half-lives of TOPSiNPs and SCPSiNPs are attributed to the presence of a silica surface covering the particles. Once the silica surface is dissolved, the PSi core is exposed and degrades readily. As shown in this study, the degradation rate depends on the thickness of silica. The thickness increases with longer oxidation time in the case of TOPSiNPs and longer coating time in the case of SCPSiNPs.

Beside silica thickness, it is known that degradability of silica microparticles depends on the silica quality [19]. The half-life of silica microparticles fabricated using different sol-gel silica precursor solutions ranges from 5 – 20 hours. Thus, we can obtain better stability for a given silica thickness by engineering the silica quality through different sol-gel synthesis process.
Aside from simply pushing the stability limit to dictate the usefulness of the particle, there is also a tradeoff between the drug loading capability and stability for both TOPSiNPs and SCPSiNPs. In case of thermal oxidation, 0.46 nm of silicon is consumed in order to produce 1 nm of silica [21]. Assuming a circular pore, an overall 0.54 nm increase in thickness reduces the pore size by 1.08 nm. In case of silica coating, the coating can be engineered by the reaction time. The thicker the silica coating layer, the less room for drug loading exists. Further analysis of the drug loading and release profile is needed to qualify the in vivo performance of the particles.

6.4. Summary and conclusion

In conclusion, we have investigated two ways to engineer the stability of PSiNPs. By using RTP and post-oxidation, the half-life of TOPSiNPs improved from 10 minutes to 3 hours, which is a reasonable time period for an investigational intravenous drug delivery platform. We have also studied sol-gel based silica coating as a method to extend the half-life of PSiNPs in PBS. This technique was shown to enhance the half-life of SCPSiNPs from 10 minutes to 8 hours. Further improvement can be made to increase the thickness or improve the quality of the silica protective layer, which will further stabilize the PSiNPs. Further analysis of the mechanisms and efficiencies of drug release from stabilized PSiNPs will help determine which method will ultimately be most successful for in vivo drug delivery applications.
REFERENCES


