University of Southampton

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING

Optoelectronics Research Centre

Optimization of waveguide enhanced Raman spectroscopy systems

by

Zhen LIU

ORC ID 0000-0002-8005-5380

Thesis for the degree of Doctor of Philosophy

October 2023

University of Southampton

Abstract

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING

Optoelectronics Research Centre

Thesis for Doctor of Philosophy

Optimization of waveguide enhanced Raman spectroscopy systems

by

Zhen Liu

Waveguide-enhanced Raman spectroscopy (WERS) is becoming a promising emerging technology, and this work is aiming at three major problems that general WERS system are currently facing.

Traditionally, WERS focused on single-mode operation, but this study explores the impact of planar waveguide thickness on surface scattering losses with a dedicated theoretical model on scattering loss. By considering multimode cross-polarization signal collection, a new generalized Figure of Merit (FOM) that takes count of the scattering loss is built for planar WERS. It is revealed that multimode pump excitation yields much superior conversion efficiencies, despite the reduced surface intensity. As a by-product, a method for reducing scattering loss and hence strengthen WERS with extra interface in waveguide substrate is also illustrated.

Waveguide grating couplers (WGCs), which offer high coupling efficiencies and relaxed alignment tolerances for WERS sensors, has been modelled with approximately treating the grating teeth as a special optical layer that is filled with the polarized dipole. The in-coupling of WGC is then calculated with out-scattering based on the reciprocal theorem, while the out-scattering is calculated with the theorem on dipole radiation at interfaces. The approximation extensively simplified the analytical difficulty for theoretical modelling. The accuracy of the model has been verified with a numerical model and the important physical mechanism that can be exploited for enhancing the Waveguide grating couplers has been revealed.

Finally, a set of algorithms specifically for processing the signal from WERS that suffers from strong background and low resolution is proposed and implemented, showing good performance. The processing includes denoising, baseline removal and deconvolution. The novel multi-frame nonlocal means method has been proven more effective with simpler parameter settings than traditional methods. The asymmetric least square method has been strengthened by incorporating a radial basis function for more precise baseline removal while proving constraint to deconvolution to suppress the ringing artefact.

Table of Contents

Table of Tables

Table of Figures

Table of Figures

Table of Figures

Table of Figures

Table of Figures

Research Thesis: Declaration of Authorship

Print name: Zhen Liu

Title of thesis: Optimization of waveguide enhanced Raman spectroscopy systems

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:

   Article
   [1] **Liu, Z**., Ettabib, M.A., Wilkinson, J.S., Zervas, M.N., 2023b. Accurate dipole radiation model for waveguide grating couplers. Results in Physics 49, 106526. https://doi.org/doi.org/10.1016/j.rinp.2023.106526 (section 4.2)

   Conference
   [2] **Liu, Z**., Ettabib, M.A., Bowden, B., Marti, A., Bartlett, P.N., Wilkinson, J., Zervas, M., 2021. Iterative non-negative constrained deconvolution for waveguide enhanced Raman spectroscopy signal recovery. Presented at the SMS / Sensors / NanoMed / EGF 2021 Joint International Conferences, Milan, Italy.
   [3] **Liu, Z**., Ettabib, M.A., Wilkinson, J., Zervas, M., 2022. Fully optimized waveguide enhanced Raman spectroscopy sensors, Presented at the SMS / Sensors / NanoMed / EGF 2022 Joint International Conferences, Athens, Greece.
   [4] **Liu, Z**., Ettabib, M.A., Wilkinson, J., Zervas, M., 2022. Accurate dipole radiation model for waveguide grating couplers, Presented at the SMS / Sensors / NanoMed / EGF 2022 Joint International Conferences, Athens, Greece.

[5] **Liu, Z**., Ettabib, M.A., Wilkinson, J., Zervas, M., 2021. Effect of scattering loss on optimized waveguide enhanced Raman spectroscopy sensors. Presented at the SMS / Sensors / NanoMed / EGF 2021Joint International Conferences, Milan, Italy.
[6] **Liu, Z**., Ettabib, M.A., Wilkinson, J.S., Zervas, M.N., 2022. Effect of multilayer substrate interference in planar waveguide scattering loss, in: European Physical Journal Web of Conferences. p. 02018. doi.org/10.1051/epjconf/202226702018

Submitted article
[7] **Liu, Z**., Ettabib, M., Wilkinson, J., Zervas, M., 2023a. Effect of scattering loss on optimization of waveguide enhanced Raman spectroscopy. https://doi.org/0.1364/opticaopen.22691893.v1 submitted to IEEE JLT (major revision) (chapter 2)
[8] **Liu, Z**., Ettabib, M., Bowden, B., Bartlett, P.N., Wilkinson, J.S., Zervas, M.N., Multiframe-based non-local means denoising for Raman spectra. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy (revised) (section 5.1)

Co-author
[9] Ettabib, M., Bowden, B., **Liu, Z**., Marti, A., Churchill, G., Gates, J., Zervas, M., Bartlett, P.N., Wilkinson, J.S., in press. Grating-Incoupled Waveguide-Enhanced Raman Sensor. PLOS ONE. (in press)
[10] Ettabib, M.A., **Liu, Z**., Zervas, M.N., Wilkinson, J.S., 2020. Optimized design for grating-coupled waveguide-enhanced Raman spectroscopy. Opt. Express 28, 37226–37235. https://doi.org/10.1364/OE.410602
[11] Ettabib, M.A., Marti, A., **Liu, Z**., Bowden, B.M., Zervas, M.N., Bartlett, P.N., Wilkinson, J.S., 2021. Waveguide Enhanced Raman Spectroscopy for Biosensing: A Review. ACS Sens. 6, 2025–2045. https://doi.org/10.1021/acssensors.1c00366

Signature: ..... Date: .....

Acknowledgements

First and foremost, I would like to express my gratitude to my primary supervisor, Prof. Michalis Zervas, for his guidance in implementing methodologies and instilling “pictures of physics” in my mind, which had previously been overshadowed by daunting mathematics. His support has made the past four years of my journey truly worthwhile. I extend my heartfelt thanks to my second supervisor, Prof. James Wilkinson, whose unwavering support, patience, guidance, and encouragement have played a crucial role in bringing this document to its final form. Without his guidance, this work would not have come to fruition. I am also grateful to my other second supervisor, Dr. Mohamed Ettabib, for his prompt assistance and enlightening discussions, which have greatly contributed to the development of my research. Special thanks are due to Prof. Phil N Bartlett from the Chemistry department for his valuable contributions to the discussions, particularly in the field of chemistry, where my knowledge was limited. His insights have sparked a newfound interest in this subject for me.

I would like to express my deep appreciation to Dr. James Gates, who has witnessed my progress and provided me the convenience to work in his lab. Additionally, I am grateful to Neil Sessions, whose unwavering support in the clean room has been instrumental whenever I had any issue with any equipment. Many thanks to Bethany Bowden from school of chemistry who provide us the analyte and spectra from Raman microscope to confirm that my algorithms work.

There is a common saying that the office is not a place to make friends, but cheerful lads like Nicolo and Sergei have proven this notion wrong. The presence of the wonderful ladies in the office has brought a lively and vibrant atmosphere, dispelling my misconceptions about science girls.

I would like to extend my gratitude to Gary from the Southampton Archery Club, who has always believed in my potential as a skilled archer. Regrettably, due to the demands of my thesis, I didn’t bring too much to the club. Anyway, it is good to know the tiny gift of mine, even if it came somewhat belatedly. I would also like to thank the young lads from both the Soton-origin FC and ECS FC, allowing me to let off steam accumulated from coding that kept failing.

Acknowledgements

I have come to realize recently that I was destined. My father's plasma striking and my mother’s lenses for building a telescope, is just one of many of their many efforts for bring me here. I believe no child could resist such allure, regardless of their inherent abilities. I am happy so far, and deeply grateful to them.

Lastly, I am happy about the decision I made four years ago, and it is good to know my limits. I am thankful to the China Scholarship Council (CSC) for their financial support throughout this four-year journey.

Definitions and Abbreviations

WERS ..... Waveguide enhanced Raman spectroscopy

MIR ..... Mid-infrared

NIR ..... Near-infrared

SERS ..... Surface enhanced Raman spectroscopy WERS

SNR ..... Signal to noise ratio

SBR ..... Signal to background ratio

FOM ..... Figure of merit

WGC ..... Waveguide grating coupler

CE ..... Coupling efficiency

FWHM ..... full width at half maximum

WTF ..... Wavelet transform filtering

PSF ..... Point spread function

ALS ..... Asymmetric least square

ROI ..... Region of interest

RBF ..... Radial basis function

NN ..... Neuron network

NLM ..... Non-local means

MNLM ..... Multi-frame based non-local means

TV ..... Total variation

Chapter 1 Introduction

Biochemical sensing technologies are driven by the increasing demand for affordable point-of-care clinical diagnostics, environmental detection, and chemical safety and security systems. The progressing silicon-based photonics technologies have been fuelling this growth. These technologies make use of chip-based sensing methods to detect biochemical species by taking advantage of the binding of biomolecules in a biological sample to receptors immobilized on the chip's surface, as well as the ease of integration with follow-up devices (waveguides, waveguide coupler, etc.). Multiple optical phenomena have been employed to detect this binding, allowing for the identification and quantification of the biochemical species. Refractometry [1,2] and fluorescent labelling [3] have emerged as two successful approaches in this field. However, refractometry suffers from low specificity (interference bonding from other biomolecules in the sample) while fluorescent-based approaches suffer from the cost and inconvenience of labelling and the risk of altering the nature of the binding process.

Optical spectroscopy techniques offer improved specificity by acquiring chemically specific compositional information with a reduced need for labels or surface receptors. Techniques such as absorption spectroscopy, fluorescence spectroscopy, and Raman spectroscopy have been utilized in various fields, including medical, environmental, and biological applications. Vibrational spectroscopies like mid-infrared (MIR) (∼2−16 μm) absorption spectroscopy and Raman spectroscopy provide chemically specific information (also known as a “fingerprint”), and they are increasingly used in biochemical sensing and measurements.

Vibrational spectroscopies are based on molecular vibrations so they can provide label-free precise and multidirectional information. The advantages of vibrational spectroscopies, due to the mechanism of detection, are as follows, but not limited to:

- Non-destructive: Vibrational spectroscopy does not require special sample preparation or destructive chemical reactions, allowing for non-destructive analysis of the sample.
- High resolution: Vibrational spectroscopy provides high-resolution information about molecular structures and chemical bonds.

- Quantitative analysis: Vibrational spectroscopy can be used for quantitative analysis as the intensity of Raman peaks is proportional to the concentration of the molecules.
- Versatility: Vibrational spectroscopy is applicable to a wide range of samples, including solids, liquids, gases, and biological samples.

As vibrational spectroscopies, the main difference between MIR absorption spectroscopy and Raman spectroscopy lies in the principle of the measurement and the information obtained. MIR absorption spectroscopy measures the absorption of infrared radiation by the sample that absorbs specific infrared wavelengths corresponding to the vibrational energy levels of the molecules. While Raman spectroscopy detects the scattered light undergoing a shift in wavelength, known as the Raman shift, due to the interaction with molecular vibrations in the analyte [4,5]. Usually, the absorption/emission cross-section of Raman scattering is around 6 to 8 orders of magnitude lower than that of fluorescence emission. This necessitates the detection of extremely low-power spectra. Nonetheless, the vibrational frequencies can be upshifted by the frequency of the laser "pump" source. This enables measurements to be taken at a more convenient wavelength (visible or near-infrared) when utilizing Raman spectroscopy (see diagram in Figure 1-1).. This is significantly important and makes Raman spectroscopy more widely applied in biomedical sensing. MIR absorption spectroscopy works in $\sim$2−16 μm wavelength band is unfavourable for biomedical sensing as biomedical analytes are always in an aqueous environment. While the light source to “pump” Raman signal in visible and near infred waveband is more abundant than that for MIR absorption.

To address the weakness of Raman spectroscopy is low cross-section, many advanced Raman variants have been explored, including resonance Raman scattering [6] coherent Raman scattering [7] and surface-enhanced Raman scattering (SERS) which emerged in the 1970s [8–11]**,** and applied successfully in various fields. Among those approaches, SERS has emerged as a successful approach for identifying biochemical species at low concentrations on chip surfaces. It combines chemical enhancement [12–14] and electromagnetic enhancements [15,16], contributing significantly to the overall enhancement factor. To be specific, chemical enhancement contributes on enhancement factor up to 10^2, while the electromagnetic enhancement is 10^4. SERS benefits from strong electromagnetic fields localized at the metal surface, resulting in surface selectivity and improved intensities of

specific vibrational bands [17,18]. Consequently, this makes SERS sensitive to the orientation of the molecule, providing additional information for greater confidence in the species, and offering improved potential for multiplexed detection as compared to fluorescence-based techniques [19,20]. The mechanism of physical enhancement and the consequent features are very similar to the waveguide-enhanced Raman spectroscopy (WERS), which will be detailed later.

With the enhancements overcoming the drawback of the small-cross section in Raman spectroscopy, SERS has advantages over the matured commercialized fluorescence-based techniques in terms of sensitivity (3 orders of magnitude improvement [21,22]), quenching resistance, and photobleaching. However, limited commercial application is attributed to issues with the reproducibility and robustness of nanostructured metal surfaces. Translating spectroscopic techniques like SERS from research laboratories to point-of-use scenarios [23] poses another challenge.

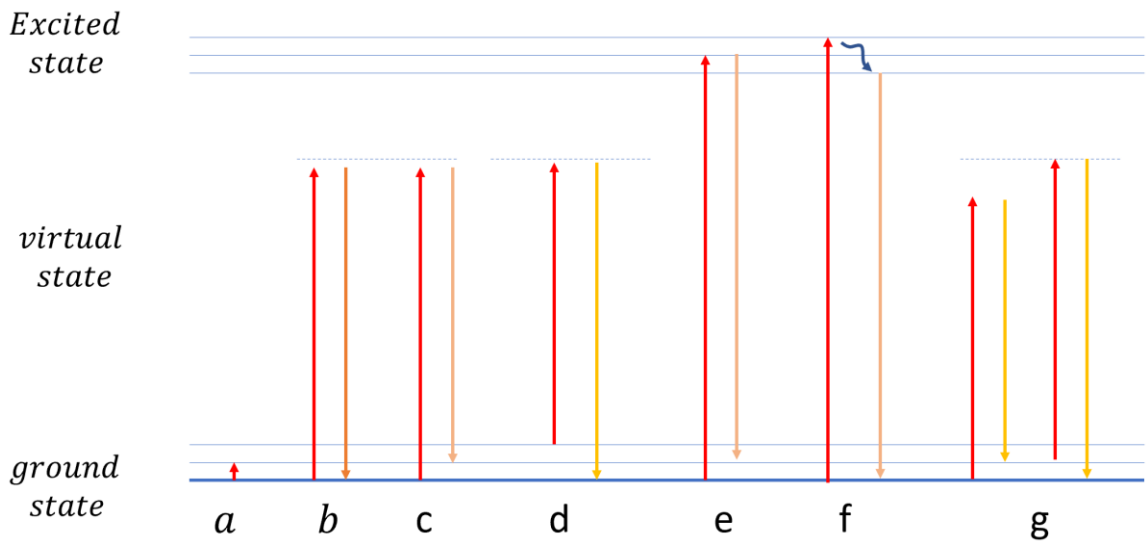

Figure 1-1. Jablonski diagram of (a) infrared absorption, (b) Rayleigh scattering, (c) Stokes Raman scattering, (d) anti-Stokes Raman scattering, (e) resonance Raman scattering, (f) fluorescence, and (g) coherent anti-Stokes Raman spectroscopy.

Waveguide-enhanced Raman spectroscopy (WERS), which also emerged in the 1970s [24–26], provides a robust and powerful optical Raman spectroscopy-based sensing tool. Benefiting from the physical enhancement, which is the major contributor to Raman signal enhancement in SERS, WERS has similar advantages as SERS. WERS can be used for the same applications as those established for SERS, with modified surface chemistry for dielectric rather than metallic surfaces. With the development of technologies in

waveguides and integrated photonics, the use of optical waveguiding to enhance Raman signals at surfaces has re-emerged as a technique with strong potential to address the limitations of SERS [27–29]. In contrast to SERS which faces reproducibility and robustness challenges due to its reliance on complex nanoscale noble metal structures, WERS employs simple, wide-band, nearly lossless robust dielectric waveguides. WERS is based on utilizing the strong evanescent field and long interaction lengths of high-index contrast low-loss optical waveguides to achieve strong excitation of (and collection from) target molecules in the upper cladding of the waveguide. The lower loss of the waveguides in WERS enables it to build up Raman signal with length and shows a "distributed" enhancement, other than the local enhancement in SERS. Moreover, WERS is attracting renewed attention due to its reproducibility, robustness, manufacturability, and ease of integration into a photonic chip.

The latest WERS demonstration has shown a higher signal-to-noise-ratios (SNR), and thus better sensitivity, than classic free space Raman systems both for thin molecular layer sensing and bulk sensing [30], although waveguide core and substrate can both contribute to the background of the collected signal. WERS has also been demonstrated to attain as much as a 300 times stronger Raman signal, with a potential theoretical advantage of 2000-fold, against confocal Raman microscope systems [31].

Regarding the promising future of WERS and current progress, this work is mainly focused on optimizing the WERS system. An illustration of the WERS system and the corresponding optimization is shown in Figure 1-2.

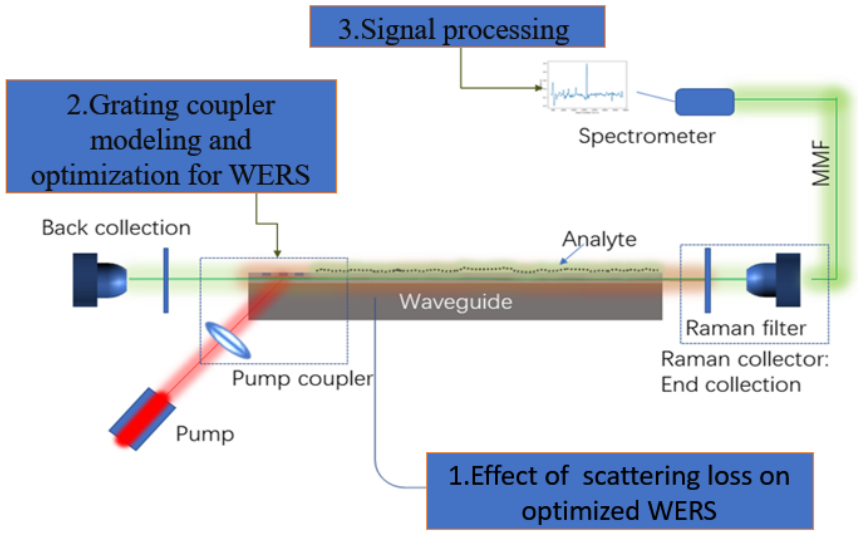

Figure 1-2. An illustration of WERS system and the corresponding optimization carried out in this work (blue boxes).

1.1 Background

As illustrated in Figure 1-2, the waveguide is the core part of the WERS system, where the Raman response from the analyte is excited and collected. Considerable work has been done to optimize waveguides for WERS, including optimizing the waveguide core material and waveguide structure.

1.1.1 Waveguide

1.1.1.1 Waveguide material

The proper material platform choice is critical for WERS performance. Materials with high absorption, surface scattering, or significant Raman or fluorescence response at the desired wavelengths may not be suitable for WERS. Absorption or scattering losses can limit the growth of the Raman signal along the waveguide and dissipate pump light, while high background emissions introduce noise and reduce the signal-to-noise ratio, ultimately affecting the limit of detection.

In addition to low absorption and scattering, an ideal WERS waveguide core material should have a high refractive index for strong optical confinement and increased evanescent field strength [32]. A comparative study evaluated four commonly used photonic platforms (Al_2O_3, Si_3N_4, Ta_2O_5, and TiO_2) in terms of their broadband background emission and Raman conversion efficiency at a pump wavelength of 785 nm [33]. The study found that different materials exhibited varying levels of background emission and Raman signal strength. Experimental comparison of core materials [33] included four candidate waveguide platforms (Al_2O_3, Si_3N_4, Ta_2O_5 and TiO_2), and demonstrated that Ta_2O_5 shows a good compromise in optical properties for WERS. The Ta_2O_5 core typically has a refractive index of 2.1, which is smaller than materials such as Si or TiO_2, resulting in a lower Raman signal conversion efficiency, but the loss is significantly lower (measured in narrow strip waveguides) [33]. Compared to lower index materials, such as Al_2O_3, the pump to signal conversion efficiency (see section 2.2) is much higher, although the loss of Al_2O_3 waveguides were half that of the Ta_2O_5 waveguides [33]. Si_3N_4 has a slightly lower index, so the loss and background noise is slightly lower, however, the Raman signal conversion efficiency is half that of the Ta_2O_5 waveguides [33]. In a recent experiment, fully optimised Ta_2O_5 slot waveguides have been shown to provide x4 higher signal

intensities than their Si_3N_4 counterparts [34]. Due to the good performance of Ta_2O_5 in WERS, it is selected as our waveguide core material for our experimental and theoretical study.

1.1.1.2 Waveguide design

The design process for waveguide enhanced Raman Spectroscopy (WERS) involves selecting the waveguide geometry and optimizing its dimensions. The most commonly used waveguide configurations in WERS are slab, strip, and slot waveguides. Planar (slab) waveguides, as an elementary form, have been investigated for power budget analysis [35] and comparison of collection configurations [36]. Various potential waveguide structures compared using various figures-of-merit (FOMs) for WERS have been investigated. Strip waveguides [37] have been studied due to their stronger confinement of field from two side walls compared to the open slab waveguide, which benefits from a more confined pump and better collection efficiency in the sense of guiding the collected light to the end of the chip and improved potential for device integration. Further, strip waveguides that are etched to form slot waveguides [28,38] and subwavelength grating waveguides [39] have been studied to access more of the evanescent field to excite the analyte.

Slab waveguides have a simple design, where the pump interacts with a large analyte area. Slab waveguides are limited in length due to chip size constraints, typically a few centimetres. The design of slab waveguides focuses on selecting the optimal waveguide thickness to maximize the square of the electric field magnitude at the surface for strong Raman excitation [40] and Raman signal collection[32].

Strip waveguides guide light laterally in the film plane, providing two-dimensional guidance. They have the advantage of guiding all collected light to the end of the waveguide and allowing longer interaction lengths through spiral implementations[37]. However, low propagation loss is crucial to fully utilize this advantage. Slot waveguides feature high refractive index strips adjacent to a lower refractive index slot. They are engineered to enhance the light in the slotted region, reducing background Raman and fluorescence emission from the waveguide material and improving interaction with the analyte. Slot waveguides offer better signal-to-noise ratios, but narrow slots may have mass transport limitations and light coupling challenges.

Compared to slab waveguides, strip and slot waveguides require additional lithographic and etching steps, which can introduce side-wall roughness and increase waveguide losses.

Overall, there is no leading approach for WERS design, as factors such as waveguide loss and material background significantly impact the design considerations. Numerical methods are commonly employed for optimizing the dimensions of the chosen waveguide geometry. A figure of merit (FOM) that considers the Raman excitation efficiency, signal capture efficiency etc. has been introduced for the optimal design [41–44].

Ongoing research and improvements in waveguide technology offer promising avenues for enhancing the efficiency and performance of WERS systems. However, the contradicting requirements between strong excitation and strong scattering loss has not been properly stressed, which are both dependent on the modal field at interfaces, due to the complicated scattering mechanism and the difficulty of including the sub-nanoscale "scatters" in numerical modal. Therefore, to better understand how scattering impacts the efficiency of WERS, in Chapter 2, a theoretical model will be built for the simple slab waveguide and experimentally verified.

1.1.2 Input coupling for pump

The technique used to couple light into WERS waveguide is a crucial aspect of the system design, impacting its performance, complexity, cost, and suitability for different applications. The few mW laser pump power loss, due to the in-coupling to the WERS structure might be easily compensated by using a higher power laser diode. However, thermal dynamics has been proven to have a significant impact on the Raman background [45]. Therefore, the coupling efficiency of the coupler of WERS is also important. Several common light coupling techniques for WERS are discussed:

(a) Prism coupling: This method utilizes evanescent coupling with a high-index prism. While used in early [40] and recent [36] WERS demonstrations, its complex assembly process makes it impractical for real-world applications, especially those requiring disposable chips.

(b) Grating coupling: Grating coupling offers good alignment tolerance, even for the thin waveguide films with small mode sizes used in WERS[46,47], which is critical for the ~100nm scale thin film used in WERS. It integrates the grating during chip production and allows light coupling from below the chip through a transparent substrate, enabling analyte

delivery without interrupting the optical path. Alignment tolerance may be reduced for strip and slot waveguides, but it remains a viable option.

(c) End-fire coupling: This commonly used method in WERS involves focusing the pump beam using an objective or aspheric lens to match the waveguide's mode size closely [48]. The focused beam is directed onto the polished or cleaved end-facet of the waveguide, potentially with a mode-transform for improved coupling efficiency. Translational alignment tolerance is critical, but it can be integrated with conventional Raman microscopes.

(d) Fibre coupling: Similar to end-fire coupling as a conventional fibre-chip integration method, fibre coupling involves connecting the pump light-carrying fibre directly to the waveguide's polished or cleaved end-facet [49]. Alignment tolerances are similar to an end-fire coupling, and permanent attachment of the fibre to the chip allows for continuous monitoring applications.

Optimizing the waveguide design by considering light coupling efficiency enhances the system's figures-of-merit (FOMs) and overall performance. Research has demonstrated the combination of grating coupling efficiency and surface intensity, resulting in an improved FOM and enhanced output Raman signal for incident pump power[42].

Among the above compared in-coupling method, the waveguide grating coupler is selected to couple pumping light for our WERS, due to the stability, high efficiency, and good alignment tolerance, making it an ideal choice for disposable point of care application. Following our previous numerical study of the waveguide grating coupler for WERS, a more detailed theoretical study on the waveguide grating coupler is conducted to compensate for the aspects that the current numerical methods struggle to solve.

1.1.3 Data processing

Data processing for WERS follows similar principles as Raman and SERS spectra processing [50,51]. A good overview of spectral pre-processing for vibrational spectroscopy is given in Ref. [10,52]. However, specific techniques for WERS data processing are limited. Raw WERS spectra include analyte Raman spectra and waveguide background spectra for referencing and subtraction. Background spectra should ideally match analyte spectra conditions, but can hardly be guaranteed. Other mechanisms, such as the wider point spread function

lowering the resolution and wavelength distortion, distinguish the data processing procedures for WERS. Particularly, the wavelength distortion due to the wavelength dependency of scattering loss[53] and signal capture efficiency [54] has not been fully addressed, which contributes to the baseline (residual background) significantly.

Regarding the mechanisms degrading spectra from WERS, background subtraction, denoising, baseline correction and deconvolution are the main steps, and have been implemented and improved in this work.

1.1.3.1 Background subtraction

After minimising the WERS background at its origin using the hardware approaches described above, the background should be subtracted using stored analyte and background spectra (both integrated and averaged appropriately). Sequential analyte and background measurements taken using a cell with flowing samples and blanks aid accurate subtraction [55]. It may be possible to improve the subtraction by first normalising the spectra to a Raman feature emanating from the waveguide core material[56]. Normalization is less critical in WERS than conventional free space Raman spectroscopy due to less variation involved in as the spectra of the background and sample are acquired consistently under similar conditions. However, normalization is still useful to ensure the spectra of the sample and background are comparable so that the subtraction can remove most of the background, revealing the measurable analyte features. Standard normalization methods, such as min-max normalization, 1-norm, vector normalization and standard normal variate [57], apply to WERS spectra while the simplest min-max normalization can produce robust results for background subtraction for most of the cases. Such subtraction is seldom perfect, mostly due to wavelength distortion, waveguide dispersion effects and experimental inconsistencies between measurements, and a procedure to fit and subtract the residual background, namely the baseline removal, is needed.

1.1.3.2 Denoising

Denoising is critical in WERS due to the weak Raman signals and the shot noise associated with the background that is generated within the whole waveguide structure. The first step is to remove spike features due to cosmic rays [58,59]. Single-scan filtering-based methods [58,60,61] have been applied to smooth out the large numbers of spurious features. The

double-exposure-based method [62] is more robust at the cost of increased acquisition time.

Common filtering methods, including low-pass filtering and median filtering in both the space and frequency domains, are typically applied to remove noise [63] [64]. Among these simple filters, Gaussian filtering and median filtering are commonly used in most denoising scenarios, including Raman spectra denoising [64], and often act as baseline techniques to characterize new denoising methods [65]. These filtering techniques, however, have a smoothing effect on the data that can suppress random fluctuations but may result in a loss of contrast between signal and noise, flattening the Raman peaks in the signal, and thereby reducing the effectiveness of the denoising process [63]. Recently, advanced methods, such as Savitzky-Golay (SG) filters [66] and wavelet-based methods [67], have been proposed for processing the Raman spectrum. The SG filter method has shown good performance but can potentially remove genuine Raman peaks with low full width at half maximum (FWHM) if the parameters are not appropriately set [63]. The wavelet transform filtering (WTF) method provides more discriminating frequency filtering than a simple Fourier transform because of multiscale frequency decomposition, so more precise denoising can be achieved than the typical low-pass filtering via Fourier transform in the frequency domain. However, sensitive thresholds for every decomposed frequency component have to be set. In addition, proper wavelets need to be chosen from the large family of wavelets, and complicated optimization algorithms have been developed just for choosing appropriate parameters for WTF [68]. Such sensitive parameters setting can be a challenging task for non-expert users, which limits their applicability in practice and potential for fully automated Raman signal processing. Machine learning has also been used to enhance the signal-to-noise ratio for Raman spectra [69]. As a data-driven method, it can be expected that machine learning will become more effective as standardized Raman data are accumulated enough for training reliable models.

1.1.3.3 Baseline correction

Due to wavelength distortion and waveguide dispersion in WERS, and optical differences between background and analyte matrices, a residual baseline often exists after normalization and background subtraction. Thus, baseline removal is a necessary step. Besides physically driven methods, such as shifted excitation [70,71], mathematical

methods do not need any instrumental modification, they can be equally effective with physical methods and are widely used. Roughly, the methods can be divided into two categories including the fitting base method and the filtering method. The first one is optimization-based fitting methods, which means fitting a smooth curve closely matching the spectra but always underneath the steep Raman peaks. Baseline curve fitting has typically been realized with generic methods [72], asymmetric least square [73], spline smoothing [74], polynomial fitting [75] and others. The baseline varies slowly compared to the Raman signal, therefore the second group of approaches uses filtering in both the space and frequency domains. Fourier transform [76] [77] and wavelet transform [78] [79]are representative approaches. However, current automated methods for baseline correction must have a certain degree of additional prior information about either the signal or the baseline component to succeed [80], In addition, hybrid iterative processing that combines several methods, including manual methods [81], is expected to result in refined baseline removal [80]. Manual methods deploy selecting points for interpolating the baseline manually “by eye”, which is slow and not automated. However, it typically produces high accuracy as it is interpreted by eye, avoiding any artifact peaks. It is possible to use manual method produced data as standards to train models, such as machine-learning, for training accurate fully automated methods.

1.1.3.4 Deconvolution

Typically, the Raman signal from WERS is collected with a multimode fibre to guide light to the spectrometer, and uses a wider spectrometer slit to maintain signal throughput, which results inevitably in a reduced resolution. The resolution of Raman spectroscopy can be improved via mathematical deconvolution, which is beneficial, especially for closely-spaced Raman peaks. Knowledge of the spectral point spread function (PSF) is required to conduct deconvolution. PSF is affected by the pump linewidth and other instrument-broadening effects, which can be difficult to determine. One simple approach is to use a spike-shaped function with estimated width, such as the Lorentz function [82]. Blind deconvolution can be applied to obtain PSF and deconvoluted Raman spectra simultaneously [83–86], in the case that PSF is difficult to be obtained accurately. As an ill-posed process, regularizations are typically needed. State of art regularizations such as Tikhonov regularization[87], maximum entropy regularization [82] and non-local low-rank regularization [85] can be applied.

1.2 Motivation, project building, thesis structure and highlights

Having the background of WERS introduced, this thesis highlights the three main concerns associated with optimizing the WERS system, as illustrated in Figure 1-2. Subsequently, this thesis aims to present and elaborate on three key aspects of WERS optimization. The motivations underlying these aspects are briefly outlined here and will be extensively discussed in their respective chapters. These aspects encompass:

1) Optimization of waveguide for WERS.

As indicated in section 1.1.1.2, the efficiency of WERS is significantly limited by scattering loss. However, existing numerical methods face challenges in accurately incorporating sub-nanometre scale "scatter" into the model for comprehensive optimization. Thus, the initial focus of this thesis is to undertake a theoretical and experimental investigation of scattering loss specifically in slab waveguides. By elucidating the underlying mechanism of scattering loss, this study proposes methods to mitigate its impact. Our new Figure of Merit (FOM) yields up to an up to 7-fold increase in signal power compared to conventional FOM-designed waveguide geometries.

This part of the work is presented in Chapter 2, and highlighted with the following points that have been demonstrated for the first time

- Theoretically predicted and experimentally confirmed that the incorporation of additional interfaces in waveguide substrates leads to a substantial reduction in scattering loss.
- Proposed the implementation of multimode cross-polarization (pump and signal in different polarizations). Initial experiments were conducted, demonstrating a good agreement in terms of the TE/TM polarization ratio.
- Achieved up to a 7-fold increase in signal power by employing waveguides that support multimode cross-polarization collection, whereas conventional waveguides used in WERS are single-mode.

2) Optimization on waveguide grating couplers for WERS

In the context of waveguide grating couplers (WGC), the utilization of numerical methods is almost indispensable, considering the impractical low accuracy of analytical theoretical models that were established decades ago. Nonetheless, the extensive range of grating

parameters poses a challenge for achieving a comprehensive optimization of WGC, primarily due to the overwhelming computational cost, even when employing supercomputers.

As the second part of this work, a new analytical theoretical model for WGCs has been developed. This model has been extensively validated through numerical simulations. In addition to surpassing expectations in terms of accuracy, this new model has revealed important physical mechanisms that are not captured by numerical simulations alone.

Chapter 3 primarily focuses on clarifying the fundamental principles of WGCs and validates several hypotheses by modifying the existing theoretical model. Chapter 4 introduces a novel model that incorporates the verified hypotheses from Chapter 3. This model treats the region of grating teeth as an optical layer filled with dipoles, and it has been successfully validated through numerical simulations, demonstrating excellent agreement. This model has further unveiled additional physical mechanisms that can be leveraged to optimize waveguide grating couplers.

The findings presented in Chapter 3 and Chapter 4, dedicated to WGCs, are highlighted by the following key features

- High accuracy for practical waveguide grating coupler design, enabling full and fast optimization for waveguide grating coupler.
- Important physical mechanisms have been revealed.
- Applicability to general gratings with diverse geometries, allowing for expandability and versatility.

3) Signal processing

The third part of this study encompasses three key aspects: denoising, baseline removal, and deconvolution. In addressing the limitations of conventional denoising methods, as well as the complexity associated with advanced methods requiring sensitive parameter settings, a multi-frame-based nonlocal means approach is proposed and implemented in section 5.1. This approach aims to overcome the drawbacks of current denoising methods and enhance their accuracy.

To address the issue of asymmetric least square (ALS) methods being prone to failure when the gradient of the baseline approaches the Raman peaks, a radial basis function network is incorporated to reinforce the simplicity and robustness of baseline removal. This integration is presented in Chapter 5.

These works are presented in Chapter 5 and highlighted as

- A new precise denoising method with a simple and safe parameter setting has been successfully implemented.
- The asymmetric least square method for baseline removal has been enhanced with a radial basis function network.
- Non-negative baseline constrain has shown an effect on reducing ringing artefact for deconvolution.

Regarding the contributions to development of our WERS system, the theoretical model developed in chapter 2 has been used to guide the waveguide design for WERS. The model formulated in section 4.2 has been used for designing the WGC for WERS. The signal processing algorithms elaborated in chapter 5 have been integrated into a software package with graphic interfaces and applied to our WERS signal for our internal analysis and publications.

Details on the calculation of the electric field in three and multilayer waveguides, which is needed in waveguide and WGCs design, is provided in the Appendix A.

Chapter 2 Effect of scattering loss on optimization of waveguide-enhanced Raman spectroscopy

Waveguide-enhanced Raman spectroscopy (WERS) is a promising sensing technology requiring accurate waveguide optimization to increase the pump/signal surface intensity. Traditionally, WERS has been focused on single-mode operation, which results in stringent waveguide parameter control and increased propagation losses. This chapter studies theoretically and experimentally the impact of planar waveguide thickness on surface scattering losses and waveguide propagation losses. This study considers radiation from a Raman-emitting dipole on a waveguide which can be captured back into the waveguide in polarization and spatial modes different from the pump mode, provided the waveguide can support them. In the case of randomly-distributed dipoles, I consider the Raman gain coefficients corresponding to all possible combinations of pump and signal polarization and spatial modes. This study introduces a new generalized FOM to optimize planar waveguide-based WERS sensors under multimode excitation/collection operation and waveguide-thickness and mode-dependent propagation losses. The FOM is shown to increase with the mode order as a result of the combination of substantial reduction in propagation loss, increased number of collecting modes, and longer optimum sensing lengths, and this occurs despite the concomitant surface intensity reduction. This implies that in the case of randomly distributed dipoles, single-mode pump excitation is not strictly required, and multimode pump excitation will give superior conversion efficiencies.

2.1 Background

As mentioned in introduction, numerical studies are among the main approaches for optimization of the corresponding FOMs of these state-of-the-art waveguides, but they mainly focus on factors including the efficiency of exciting or capturing the Raman signal, or the pump/signal conversion efficiency per unit length [41,88].

However, with a few exemptions (see [41,43]), scattering loss, which is a fundamental factor that limits WERS overall conversion efficiency, has received limited attention. Usually, it has been either neglected or oversimplified as a constant, independent of the waveguide parameters, in current conventional FOMs. As a strong evanescent field

exposed at the surface of the waveguide core is expected in WERS, scattering from the roughness of the surface is unavoidable. Scattering loss is not independent of the waveguide geometry, as surface modal field varies with geometry (i.e. thickness for planar (slab) waveguides), and a full optimization should take the variation of scattering loss as a function of waveguide core thickness into consideration. Kita et al [41,43] have considered the impact of slot waveguide parameters on side-wall induced scattering losses and used it to optimize the proposed FOM. This study, however, is limited to single-mode slot waveguide operation. Even optical fibres, which exhibit minimal losses reaching the limits of Rayleigh scattering, exhibit increased losses when single mode operation and strong evanescent field is needed for sensing or other similar purposes [89]. A ~dB/cm scattering loss could result from a nanometre surface fluctuation in this case [90].

The mechanism of scattering loss from a waveguide has attracted attention since the late 1960s, resulting in four main approaches to estimate it. The simplest one is based on ray optics, and a simple analytical expression was deduced by Tien [91], suitable mainly for multimode waveguides. Marcuse et al. in the 1960s studied the loss caused by imperfections of the surface by using the mode conversion theory for cylindrical waveguides [92] and slab waveguides [93]. Marcuse's model was based on the weakly-guiding approximation, which makes it unsuitable for WERS devices using high index contrast waveguides. The volume-current method [92,93] was introduced in the 1970s as a practical way to calculate scattering losses in waveguides [94,95]. Kita et al. [41,43] adopted this method in their numerical model for calculating FOMs for different waveguides for sensing, and it is the only work so far that takes into account scattering loss into the waveguide optimization and WERS FOM. Payne-Lacey's (PL) theoretical analysis [96,97], based on solving the wave equation via Green's function at a random interface, shows that the scattering loss is a function of the waveguide thickness and spatial frequency (characterized by correlation length) of the rough surface. Moreover, recent theoretical analysis and experimental work of Schmid et al, which is developed from PL's theoretical work, has shown that surface-roughness scattering also interferes with reflections from other interfaces of the waveguide and significantly modifies the resultant scattering loss[53]. In other words, conventionally optimized waveguide parameters for maximum evanescent field strength at the interfaces might not be optimal for WERS due to the concomitant increased scattering loss due to surface roughness.

Another point that has been rarely addressed is the full impact of the multimode signal collection and cross-polarization excitation (i.e. signal and pump in different polarizations, which is discussed by Dhakal et al [98]). Conventionally, the waveguide is designed for the fundamental mode because of its strong surface modal field. However, due to additional requirement for minimized scattering loss, the waveguide parameters may become optimum for WERS when the waveguide is multimode and Raman is collected in multiple modes supported by the waveguide. Thus, the signal could be collected by the waveguide in a higher-order mode and different polarizations as the Raman signal is emitted by randomly oriented dipoles. These additional considerations would further impact the optimum waveguide length (or "saturation" length, in the case of backscattered collection), which is simply the reciprocal of the loss coefficient for single mode excitation and collection.

This work expands previous approaches and extends the WERS waveguide design into the multimode regime by considering the combined effects of reduced surface intensities, reduced scattering losses (increased interaction lengths) and increased number of pump excitation/signal collection spatial/polarization modes. Section 2.2.1 expands the random dipole-to-waveguide coupling into multimode polarization-resolved operation. Section 2.2.2 considers multimode pump-excitation/signal-collection and generalize previously derived FOMs, applicable to single-mode operation, for application in the multimode regime. Section 2.2.3 extends previously derived modal field, coupling efficiency and scattering loss calculations into the multimode regime. The model on scattering loss is extended with the transfer matrix method so that the multilayer interference effect on scattering loss is considered. Section 2.3.2, shows detailed comparison between planar waveguide scattering loss theoretical calculations and experimental measurements. In sections 2.3.3 and 2.3.4, the traditional singlemode-pumping/singlemode-collection FOM calculations are used to compare with the new singlemode-pumping/ multimode-collection FOM. In addition, we have also calculated and compared the signal-background-ratio. Section 2.4 summarizes and present the final conclusions.

2.2 Theoretical model

In Figure 2-1, a typical three-layer WERS waveguide system is shown. The pump is coupled into the waveguide from a coupler and excites the Raman signal from the analyte, which is

modelled as an oscillating dipole, on the top of the core of the waveguide excited by the evanescent field, while the guided power is scattered by the rough core-cladding interface.

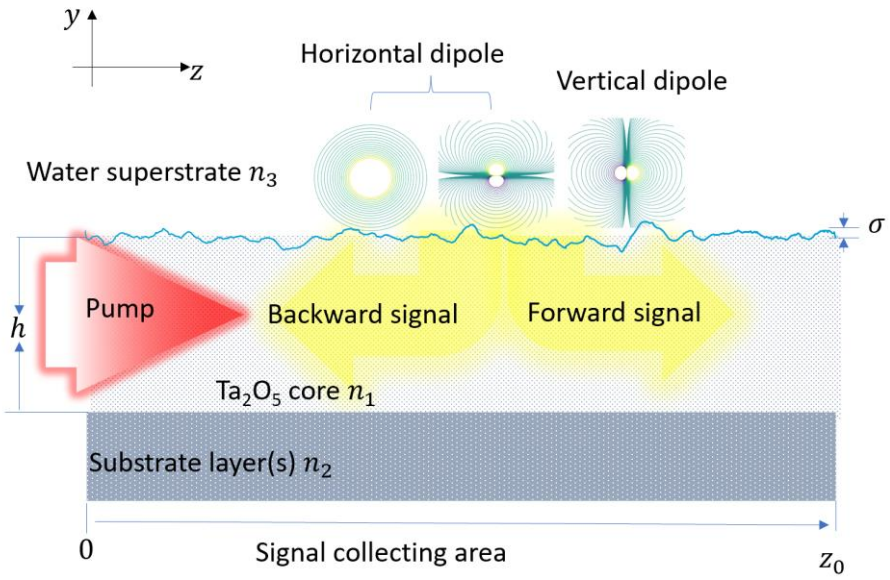

Figure 2-1. Illustration of a WERS system with a three-layer waveguide. The refractive index for cladding, waveguide core, and substrate are n_3, n_1 and n_2 respectively. The geometry of the waveguide for further calculation is depicted. The root-mean-square (RMS) fluctuation is σ, as marked.

The randomly oriented excited molecule radiates the Raman signal incoherently and isotropically. Therefore, the signal can be collected in TE and TM polarization regardless of the pumping polarization from the top of the waveguide (free-space collection) and the ends of the waveguide. However, it has been shown that much more power can be collected from the ends of the waveguide [36], and hence, the top-collection is not considered further here. The forward-/back-collected signal is the integration of the signal excited and captured by the waveguide with the supported spatial mode along the waveguide from $z = 0$ to z_0. Following Ref. [41], the Raman gain coefficient η_j for the j^{th} mode, which indicates the efficiency of converting pump to guided j^{th} mode Raman signal per unit length, is the product of the efficiency of dipole excitation by the pump and the coupling efficiency of the power radiated from the dipole into the guided mode.

2.2.1 Dipole-to-waveguide coupling efficiency; pump/signal polarization dependence

The coupling efficiency into the j^{th} waveguide mode of the emission from a dipole on a layered structure is calculated in terms of the waveguide mode power (P_j) normalized to the dipole free space emission power (P_0) and is, in general, given as [54,98,99]:

$$\frac{dP_j}{P_0} = 1 + Re\int_0^\infty dk_\rho \left[\frac{\mu_\rho^2}{\mu^2}\frac{3}{4}\frac{k_\rho}{k_3^3 k_{3y}}\left(k_3^2 r_{jTE} - k_{3y}^2 r_{jTM}\right) + \frac{\mu_y^2}{\mu^2}\frac{3}{2}\frac{1}{k_3^3 k_{3y}}\left(k_\rho^3 r_{jTM}\right)e^{-ik_{3y}r_0}\right] \quad (2.1)$$

where μ and $k_3 = n_3 k_0$ are the dipole moment and wavenumber in the superstrate, while the subscripts y and ρ denote the vertical and horizontal (x-z plane) components, respectively. r_{jTM} and r_{jTE} are the plane-wave reflectivity for the p (TM) and s (TE) polarizations of the whole structure, including the waveguide and any other interface(s) in the substrate (see Figure 2-1). r_0 is the distance between dipole and waveguide core. For monolayer detection, the analyte is deposited on the waveguide core, so $r_0 = 0$. TM/TE waveguide modes are given by the poles of the r_{jTM}/r_{jTE} reflectivity and the integration is calculated along the infinitesimal semicircle in the complex plane surrounding the corresponding pole [98,99]. The horizontal dipole moment (μ_ρ) can be further analysed into two orthogonal components (μ_x and μ_z). Randomly oriented dipoles can also be divided into three equally-populated x-, y-, and z-oriented subgroups and the horizontal and vertical dipole moment contributions in equation (2.1) are usually taken as μ_ρ/μ = 2/3 and μ_y/μ = 1/3 [98,99]. Strictly speaking, this applies when the dipole excitation is isotropic with respect to the pump field vector (i.e., it depends only on the pump intensity $\propto|E_p|^2$), as is the case when considering gain effects in doped planar waveguides [100]. However, in WERS the x-, y- and z-oriented subgroups will be predominantly excited by the x-, y- and z-components of the pump field, respectively, and, therefore, this study will explicitly use $\mu_x/\mu = \mu_y/\mu = \mu_z/\mu$ = 1/3, along with the appropriate pump field components, in the subsequent analysis. equation (2.1) gives the total collected signal power including TE and TM polarized modes. The collected signal partition into TE and TM waveguide modes is given by the terms containing explicitly the r_{jTE} and r_{jTM} coefficients.

For calculating the reflectivity and transmissivity for the dipole coupling efficiency in equation (2.1)

$$k_{yn} = \sqrt{k_n^2 - k_\rho^2} \quad (2.2)$$

The free-space radiation power normalized to pump power P_p from a dipole excited by the evanescent field of a planar waveguide is [101]

$$\frac{P_0}{P_{pump}} = \frac{4\pi^3}{3k_v\varepsilon_0^2}\frac{\sigma_R(\omega_p,\omega_s)n_g(\omega_p)}{\tilde{A}_{eff}(\omega_p)} \quad (2.3)$$

where σ_R is the scattering cross-section and is given a representative value of $10^{-30}cm^2$ in the following calculations, while $k_v \approx 1.26\times10^{23}\ C^{-2}V^2m^2$ is a constant specified by

Long [52], n_g is the group index and $\tilde{A}_{eff}$ is the effective mode area defined in Ref.[101] and given as

$$\tilde{A}_{eff}(r_0, \omega_p) \equiv \frac{\iint_A \varepsilon_r(r, \omega_p)\left|e(r, \omega_p)\right|^2 dA}{\left|e(r_0, \omega_p)\right|^2} \tag{2.4}$$

$\sigma_R(\omega_p, \omega_s)$ is the cross-section of the scatter. $n_g(\omega_p)$ is the group refractive index for the pump. $k_v = 1.26 \times 10^{23} C^{-2} V^2 m^2$ is a constant given by Long [52]. $\varepsilon(\vec{r}, \omega_p)$ is the relative dielectric function of the pump. $e(\vec{r}, \omega_p)$ is the electric field of in its mode area.

The group refractivity is defined as the velocity of light in vacuum, c, divided by the group velocity of the mode, while Snyder proves the group velocity of a slab waveguide to be the modal power P_j divided by, $\mathcal{W}_j$, the total stored energy per unit length of the waveguide [102]

$$v_{gp} = \frac{P_p}{\mathcal{W}_p} \tag{2.5}$$

For non-dispersive core [102],

$$\mathcal{W}_p = \frac{1}{2}\varepsilon_0 \iint_{A\infty} n(\vec{r})^2 \left|e_p(\vec{r}, \omega_p)\right|^2 dA \tag{2.6}$$

While the power can be calculated by integrating the Poynting vector. So, the item of $\frac{n_g}{\tilde{A}_{eff}}$ now is taking a more straightforward form,

$$\frac{n_g}{\tilde{A}_{eff}} = \frac{\frac{1}{2} c\varepsilon_0 \iint_{A\infty} n(r)^2 \left|e_p(r)\right|^2 dA}{\frac{1}{2}\iint_{A\infty} e_p(r) \times \hbar_p^*(r) \cdot \hat{z}\, dA} \frac{\left|e_p(r_0)\right|^2}{\iint_{A\infty} \varepsilon_r \left|e_p(r)\right|^2 dA}$$

$$= \frac{c\varepsilon_0 \left|e_p(r_0)\right|^2}{\iint_{A\infty} e_p(r) \times \hbar_p^*(r) \cdot \hat{z}\, dA} \tag{2.7}$$

Where the power-normalized modal field can be defined to simplify the formula,

$$\left|E_p(r_0)\right|^2 = \frac{\left|e_p(r_0)\right|^2}{\iint_{A\infty} e_p(r) \times \hbar_p^*(r) \cdot \hat{z}\, dA} \tag{2.8}$$

Therefore, taking the previous derived expression form [101], the free-space radiation due to the evanescent field from a slab waveguide can be written in a simple form,

$$\frac{P_0}{P_{pump}} = \frac{4\pi^3 c}{3k_v\varepsilon_0}\sigma_R(\omega_p,\omega_s)\left|E_p(r_0)\right|^2 \tag{2.9}$$

The Raman gain coefficient, η_j for the j^{th} signal mode is calculated by multiplying the signal coupling efficiency equation (2.1) and excitation efficiency equation (2.9), namely,

$$\eta_j = \frac{dP_j}{P_p}\frac{1}{dz} = \frac{4\pi^3 c}{3k_v\varepsilon_0}\sigma_R(\omega_p,\omega_s)\left|E_p(r_0)\right|^2\frac{dP_j}{P_0} \tag{2.10}$$

From Equation(2.1), (2.9) and (2.10) we can derive the following polarization resolved dipole excitation and signal collection cases. $\left|E_p\right|^2$ is the power normalized squared electric field. The TE polarization $|E_\mathrm{p}|^2 = |E_\mathrm{x}|^2$ while for TM polarization $|E_\mathrm{p}|^2 = |E_\mathrm{y}|^2 + |E_\mathrm{z}|^2$.

Equation (2.10) is also applied to approximately calculate the background generated by the pump in the core and substrate area by Ref. [33,43]. The excitation efficiency for background is written as [33,43],

$$\eta_{bg} = \rho_{bg}\sigma_{bg}n_g^2\lambda^2\int_{cs}\frac{|e(r_0)|^4}{\left(\int_\infty \varepsilon_r|e(r)|^2dr\right)^2 dr_o}. \tag{2.11}$$

In this equation, ρ_{bg} and σ_{bg} are the density and cross-section of background. These terms will cancel each other out during the computation of the relative signal background ratio (rSBR), based on the assumption that the density and cross-section values for both the signal and background are identical. The group index n_g is calculated with equation (2.7). The integration is calculated along the region where background generated. Following Ref. [33,43], equation (2.11) is used for calculating the rSBR that characterizes the background, while the signal generated at the cladding interface, which is key focus of the study, is specified as follows.

2.2.1.1 TE pump

In this case $|E_\mathrm{p}|^2 = |E_\mathrm{x}|^2$, $\frac{\mu_\rho^2}{\mu^2} = \frac{\mu_x^2}{\mu^2} = \frac{1}{3}$ and equation (2.1)and (2.9)reduce to:

$$\eta_j^{TE} = \frac{4\pi^3 c}{3k_v\varepsilon_0}\sigma_R(\omega_p,\omega_s)|E_x(r_0)|^2\,Re\int_l dk_\rho\,\frac{1}{4}\frac{k_\rho}{k_3^3 k_{3y}}\left(k_3^2 r_{jTE} - k_{3y}^2 r_{jTM}\right) \tag{2.12}$$

where *l* denotes the infinitesimal contour of integration in the complex k_ρ plane around the r_{jTE} and r_{jTM} poles. More specifically, $k_\rho=\beta_j+\epsilon\exp(i\theta)$, where β_j is the propagation constant of the captured mode, ϵ is a small number and the integration is done along θ=0 to π. r_{jTE} and r_{jTM} are in general functions of k_ρ and are calculated at $k_\rho=\beta_j$. It is apparent that TE polarized pump excites the x-oriented dipole subgroup and contributes to both TE and TM signal coupling.

2.2.1.2 TM pump

In this case $|E_p|^2 = |E_z|^2 + |E_y|^2$ and $\frac{\mu_\rho^2}{\mu^2}=\frac{\mu_z^2}{\mu^2}=\frac{1}{3}$, $\frac{\mu_y^2}{\mu^2}=\frac{1}{3}$ and from equation (2.1) and (2.9) it can be obtained:

$$\eta_j^{TM}=\frac{4\pi^3 c}{3k_\nu\varepsilon_0}\sigma_R(\omega_p,\omega_s)\,Re\int_l dk_\rho\,\frac{1}{4}|E_z(r_0)|^2\frac{k_\rho}{k_3^3k_{3y}}\left(k_3^2r_{jTE}-k_{3y}^2r_{jTM}\right)$$
$$+\frac{1}{2}\left|E_y(r_0)\right|^2\frac{1}{k_3^3k_{3y}}\left(k_\rho^3r_{jTM}\right) \qquad (2.13)$$

It is evident that TM polarized pump excites both the z- and y-oriented dipole subgroups (with strengths proportional to the $|E_z|^2$ and $|E_y|^2$ pump components, respectively) and contributes to both TE and TM signal coupling. It is also shown that under TM polarized pump excitation both TE and TM signal modes are collected. It should be mentioned that in previous studies [98,99] the full pump field $|E_p|^2$ was considered with and $\frac{\mu_\rho^2}{\mu^2}=\frac{2}{3}$, $\frac{\mu_y^2}{\mu^2}=\frac{1}{3}$ in equation (2.1).

2.2.2 Raman conversion efficiency – effect of scattering propagation loss

Due to the scattering loss, the collected signal at the waveguide ends is a function of both the loss coefficient and waveguide length, and for the $j^{th}-$ mode collected Raman signal, P_{sj} , can be described by the following differential equations, for the forward collection configuration (equation (2.14)(a)) and the back collection configuration (equation (2.14) (b)),

$$\frac{dP_{s,j}}{dz}=\begin{cases}\eta_j P_p(z)e^{-\alpha_{s,j}(z_0-z)} & (a)\\ \eta_j P_p(z)e^{-\alpha_{s,j}(-z)} & (b)\end{cases} \qquad (2.14)$$

where z_0 is the total sensing length, and $\alpha_{s,j}$ is the loss coefficient for a signal collected in the j^{th} mode. The pump decays exponentially along the waveguide with loss coefficient α_p, and is expressed as $P_p(\mathrm{z}) = P_p(0)e^{-\alpha_p z}$. The j^{th} mode collection efficiency is obtained from equation (2.14) as

$$\frac{P_{s,j}}{P_p} = \eta_j f_{f(b),j} \tag{2.15}$$

equation (2.15) determines the ability of the waveguide to convert the pump power into Raman signal in the j^{th} mode, where $f_{f(b),j}$ is the integrated function regarding the loss of signal and pump along the waveguide for forward (indicated by subscript f) and backward collection (indicated by subscript b), and is expressed as

$$f_{f,j}(\alpha, z_0) = \begin{cases} \frac{(e^{a_p z_0} - e^{\alpha_{s,j} z_0}) e^{-z_0(a_p + a_{s,j})}}{\alpha_p - \alpha_{s,j}} & \alpha_p \neq \alpha_{s,j} \\ z_0 e^{-\alpha_p z_0} & \alpha_{s,j} = \alpha_p \end{cases} \tag{2.16}$$

$$f_{b,j}(\alpha, z_0) = \frac{1 - e^{-z_0(\alpha_p + a_{s,j})}}{a_p + a_{s,j}} \tag{2.17}$$

2.2.2.1 Analytical model on scattering loss

As mentioned in the introduction, the geometry of the waveguide (thickness of the planar waveguide core) changes the scattering loss as well, through two mechanisms. Firstly, the core thickness determines the strength of the modal field at the core-cladding interface, and a stronger surface modal field creates stronger scattering. Waveguide scattering loss is usually calculated by the Payne-Lacey's model [97]. This model has been modified by Schmid et al [53] to include the contributions of the interference effects inside the two core interfaces that change the scattering significantly. The modified scattering loss coefficients are given by [53]

$$\alpha_{jTE} = \frac{k_0^3}{4\pi n_{jeff}} (n_3^2 - n_1^2)^2 (\mathrm{E}_0^x)^2 \int_0^\pi \left[\left(1 + r_{jTE}(\theta)\right)^2 + t_{jTE}^2(\theta) \right] \tilde{R}(\beta - n_3 k_0 \cos\theta) \mathrm{d}\theta \tag{2.18}$$

$$\alpha_{jTM} = \frac{n_{\text{jeff}} k_0^3 (n_3^2 - n_1^2)^2}{4\pi b} \int_0^{\pi} \left(\left| \frac{n_1^2}{n_3^2} E_0^y \left(1 + r_{jTM}(\theta)\right) cos\,\theta - E_0^z \left(1 - r_{jTM}(\theta)\right) \sin\theta \right|^2 + \left| \frac{n_1^2}{n_3^2} E_0^y t_{jTM}(\theta) \cos\theta + E_0^z t_{jTM}(\theta) \sin\theta \right|^2 \right) \tilde{R}(\beta - n_3 k_0 \cos\theta) d\theta \quad (2.19)$$

where E_0^x (E_0^y, E_0^z) is the normalized TE (TM) mode field at the n_1/n_3 interface, $n_{j\text{eff}}$ is the effective refractive index of the corresponding waveguide j^{th} mode. $r_{jTE\backslash TM}$ and $t_{jTE\backslash TM}$ are the total reflectivity and transmittance for TE/TM waves reflected and transmitted through the whole structure of the waveguide. $r_{jTE\backslash TM}$ and $t_{jTE\backslash TM}$ are also expressed as functions of θ, the angle of scattered light. $b = \int_{-\infty}^{\infty} n^2(y) \, |E^y(y)|^2 \mathrm{d}y$is the normalization constant. $\tilde{R}$ is the correlation function of the roughness which takes the widely-accepted auto-correlation form [97] $\tilde{R}(k) = \frac{2\sigma^2 L}{1+L^2k^2}$, characterized by the root-mean-square height variation of the fluctuating surface σ, and L is the rough-surface correlation length. The expressions (2.18) and (2.19) are originally for a waveguide core with two rough interfaces, while, for the planar waveguide the substrate interface is normally much smoother and therefore contributes negligible loss. Therefore, the loss coefficients used in the following study are half the ones given by equation (2.18) and (2.19). For given roughness parameters, changes in the waveguide parameters (core thickness and/or refractive indices) will impact the scattering loss through the variation of the n_1/n_3 interface field strength, as well as the changes in $r_{jTE\backslash TM}$ and $t_{jTE\backslash TM}$.

2.2.2.2 Interference from multilayer substrate.

Schmid et al's model only considers the interference effect in a typical three-layer waveguide, i.e. the reflections / transmissions through cladding – core – substrate [53]. However, practical waveguides might be more complicated than the simple three-layer structure, i.e., the substrate could be multilayered. Theoretically, the reflection of the scattering in the substrate could also impact the interference, so original Schmid et al.'s model should be expanded by including this multilayer interference effect into consideration. The transfer matrix method is adopted here to include such concerns, and we employ standard transfer matrix method for calculating the multilayer structure illustrated in Figure 2-2 is given by Ref. [103,104].

Light transfer in the same layer n from upper interface y_{n-1} to lower interface y_n

$$\begin{bmatrix} A_n' \\ B_n' \end{bmatrix} = M_n \begin{bmatrix} A_n \\ B_n \end{bmatrix} \tag{2.20}$$

where the transfer matrix for layer n is

$$M_n = \begin{bmatrix} e^{i\phi_n} & 0 \\ 0 & e^{-i\phi_n} \end{bmatrix} \tag{2.21}$$

in which $\phi_n = k_{yn}h_n$ is the phase shift in layer n, where k_{yn} and h_n are the y component of the wave number in layer n and thickness of layer n, respectively.

And for calculating the loss coefficient (equation (2.18) and (2.19)), assuming the scattering is from layer 0-1 interface, with a scattering angle θ (0~π),

$$k_{yn} = \sqrt{k_n^2 - k_0^2 \cos^2(\theta)} \tag{2.22}$$

The matrix for the interface can be written as

$$m_{nn+1} = \frac{1}{t_{nn+1}} \begin{bmatrix} r_{nn+1} & 1 \\ 1 & r_{nn+1} \end{bmatrix} \tag{2.23}$$

where *r* and *t* are the Fresnel reflectivity and transmissivity from a single layer. The transfer matrix describes the light transfer from layer n to n+1 is

$$\begin{bmatrix} A_n \\ B_n \end{bmatrix} = \boldsymbol{M}_{nn+1} \begin{bmatrix} A_{n+1} \\ B_{n+1} \end{bmatrix} \tag{2.24}$$

where $\mathbf{M}_{\boldsymbol{nn+1}} = M_n M_{nn+1}$.

The transfer matrix for the whole structure is the product of all matrixes for each layer

$$\begin{bmatrix} r \\ 1 \end{bmatrix} = \widetilde{\boldsymbol{M}} \begin{bmatrix} 0 \\ t \end{bmatrix} \tag{2.25}$$

where,

$$\widetilde{\boldsymbol{M}} = \prod_{i=0}^{n} \widetilde{\boldsymbol{M}}_{nn_{+}1} \tag{2.26}$$

i.e.

$$\begin{bmatrix} r \\ 1 \end{bmatrix} = \begin{bmatrix} \widetilde{M}_{00} & \widetilde{M}_{01} \\ \widetilde{M}_{10} & \widetilde{M}_{11} \end{bmatrix} \begin{bmatrix} 0 \\ t \end{bmatrix} \tag{2.27}$$

Hence

$$t = \frac{1}{\widetilde{M}_{11}}, \quad r = \frac{\widetilde{M}_{01}}{\widetilde{M}_{11}} \tag{2.28}$$

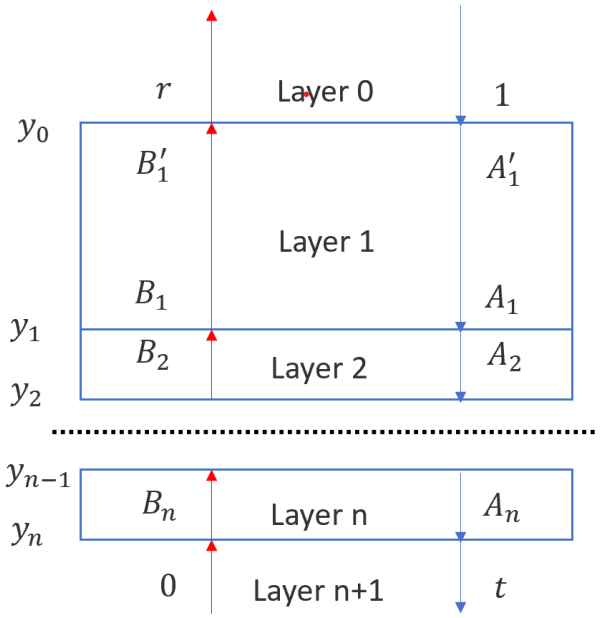

Figure 2-2. Illustration of the light propagation, reflection (r) and transmission (t) in a multilayer structure with an input (marked as 1) from layer 0.

2.2.3 Raman conversion efficiency – Figure of Merit (FOM)

Normally, optimization focuses on strengthening the surface field for a higher Raman conversion coefficient, while the impact on the scattering loss is either ignored or simply treated as a constant, independent of the waveguide parameters. In this case, the fundamental mode is always preferred due to the strongest modal field, and the loss for pump and signal are assumed to be the same ($\alpha_s = \alpha_p$). By taking the conventional optimal length, $z_0 = \frac{1}{\alpha_p}$ [41], into equation (2.16), for forward collection, the function f_f takes its maximum value, an appropriate FOM for forward collection is given by [41]

$$FOM_f = \frac{\eta_j}{e^1 \alpha_p} \tag{2.29}$$

The constant e^1 is omitted in reference [41] for simplicity, but here it is kept because it allows direct quantitative comparison of forward and backward configurations. For backward collection, the signal increases with waveguide length monotonically to its

extreme value $P_{sj}(\mathrm{z}_0 \to \infty) = \frac{\eta_j P_p(0)}{a_p + \alpha_s}$, which is indicated by equation (2.17). The corresponding backward collection *FOM*$_b$ is [33,43]

$$FOM_b = \frac{\eta_j}{2a_p} \tag{2.30}$$

This implies that all else being equal, backward collection is always ~36% better than forward collection.

As can be inferred from equation (2.18) and (2.19), for fixed surface roughness parameters and refractive indices, the loss coefficient is a function of the thickness of the waveguide *h*. It is therefore expected that the thickness that maximizes the surface mode field does not necessarily maximize the overall WERS conversion efficiency. For example, due to reduced scattering loss, the optimum thickness could be greater than the conventional FOM indicates. Furthermore, a thicker waveguide may guide more modes and higher order modes may be used to increase the overall Raman collection. The captured signal and the pump in the core can be guided with different modes, in which case the loss coefficient will no longer be the same. Therefore, here the new FOMs for signal collected in each individual (j^{th}) mode are expressed using a more generalized formula including the ratio between the collected signal power and pump power, following equation (2.15)

$$FOM_{f(b)j} = \eta_j(h) f_{f(b)j}(h, z_0) \tag{2.31}$$

The signal collected at the ends of the waveguide is incoherently excited and captured by all the modes and polarizations that the waveguide supports. In order to calculate the FOM for bulk analyte, this study assumes the dipole is evenly distributed with a density ρ along y direction. In this case, generalizing the single-mode excitation/collection FOM mentioned above, this study introduces a FOM applicable to single-mode excitation/multimode collection WERS arrangements, namely:

$$FOM_f^m = \sum_{k=TE}^{TM} \sum_{j=0}^{N} FOM_{fjk}^m$$

$$= \sum_{k=TE}^{TM} \sum_{j=0}^{N} \eta_{jk}^m(h) f_{fjk}^m(h, z_0) \tag{2.32}$$

In the equation (2.32), FOM_f^m is the total FOM for forward signal collection (subscript f), obtained under m^{th} (TE or TM) mode pumping, by summing up the individual figures of merit FOM_{fjk}^m corresponding to individual j^{th} signal mode with *k* (TE or TM) polarization pumped by the m^{th} mode. η_{jk}^m is the simplified Raman coefficient given by Eqns. (2.12) and (2.13) for the m^{th} pump mode and the j^{th} signal mode with *k* (TE or TM) polarization. f_{fjk}^m is given in equation (2.16) and corresponds to the propagation factor with the propagation loss coefficient of the m^{th} pump mode and the j^{th} signal mode with *k* (TE or TM) polarization given by Eqns. (2.18) and (2.19). A similar formula applies for the case of backward collection (with f replaced with b). The new generalized FOM incorporates geometry dependent loss and multi-interface interference effects and includes collection in multiple modes and both polarizations for polarized waveguide excitation.

This study covers both monolayer and bulk analyte detection, where the analyte fills the entire cladding. For monolayer, the value of density ρ along y direction is apparently 1, while for bulk analyte with a typical molecule size of ~nm, the value 10^9 is taken into our calculation. To calculate the FOM for bulk detection, we integrate the excitation efficiency η over the cladding area. This allows for a comprehensive evaluation of WERS performance for different analyte detection scenarios. FOM calculations are usually for monolayer detection, unless otherwise specified.

2.3 Theoretical calculations

In this section, modal field, coupling efficiency and scattering loss calculations are conducted for the following comparison with experimental waveguide loss measurements. The theoretical calculations also build the foundation for FOM calculations for polarization resolved single-mode pumping and single-mode/multi-mode signal. The calculations presented here correspond to a planar waveguide with Ta_2O_5 core deposited on a SiO_2 substrate. The cladding is assumed to be water. The refractive index for cladding, core, and substrate used throughout this section are 1.333, 2.12, and 1.4535 at a wavelength of 785nm, respectively. The pumping wavelength is set to be 785nm, which is a balanced choice in terms of sensitivity and fluorescence background noise [105]. First, the dispersion equation is solved to obtain the electric field of the waveguide modes, normalized to carry 1W/m power. Details regarding obtaining the electric field of the waveguide modes is elaborated in Appendix A.

2.3.1 Modal field, coupling efficiency and scattering loss calculations

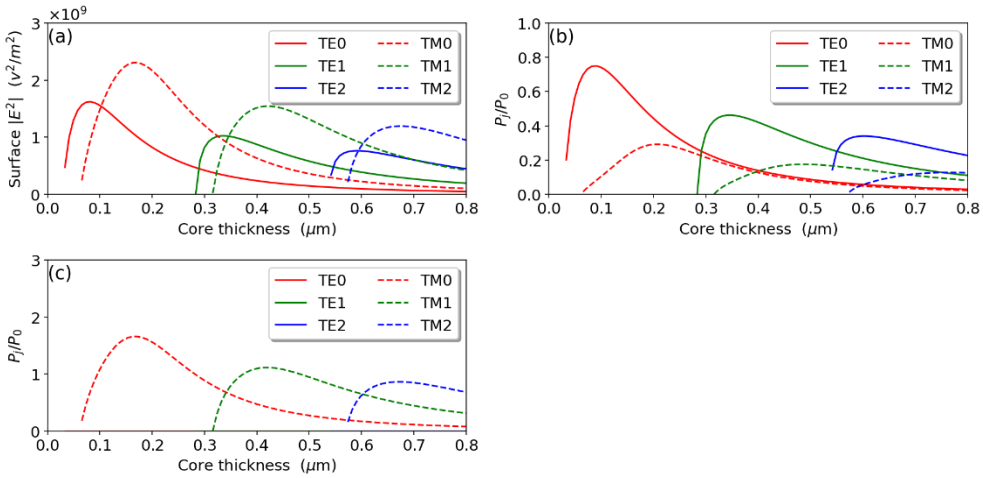

Figure 2-3. Excitation and waveguide coupling for a dipole on the cladding-core interface; (a) Normalized squared electric field at the core-cladding surface. (b) Signal capture efficiency in different modes for horizontal dipole. (c) Signal capture efficiency in different modes for vertical dipole.

The surface electric field for various waveguide modes as a function of core thickness is shown in Figure 2-3(a). TM modes show the highest core/cladding surface field, making them preferable to the TE modes for WERS. In Figure 2-3(b), the efficiency for coupling the radiation from a horizontal dipole to guided modes is shown, assuming the dipole is radiating at the same wavelength as the pump ($\lambda_s = \lambda_p$=785nm). Figure 2-3(c) shows the case for a vertical dipole. It is shown that, in this case, there are no TE signal modes excited or collected. We also observe that for TM_0 pumping the collection efficiency can be >1 due to the strong Purcell effect. Figure 2-3 shows that the TM mode is preferable for pumping. The combined signal collection for vertical and horizontal dipole shows that TM mode is superior than TE mode.

Both the thickness of the waveguide and the roughness parameters impact the scattering loss. The loss coefficient is proportional to the square of the RMS roughness amplitude. In our model, the RMS amplitude is set to 1nm in line with recently developed state-of-art roughness measurements [106–108], which is also a value close to the lattice constant of Ta_2O_5 (a=0.620 nm, b=0.366 nm and c=0.389 nm) [109].

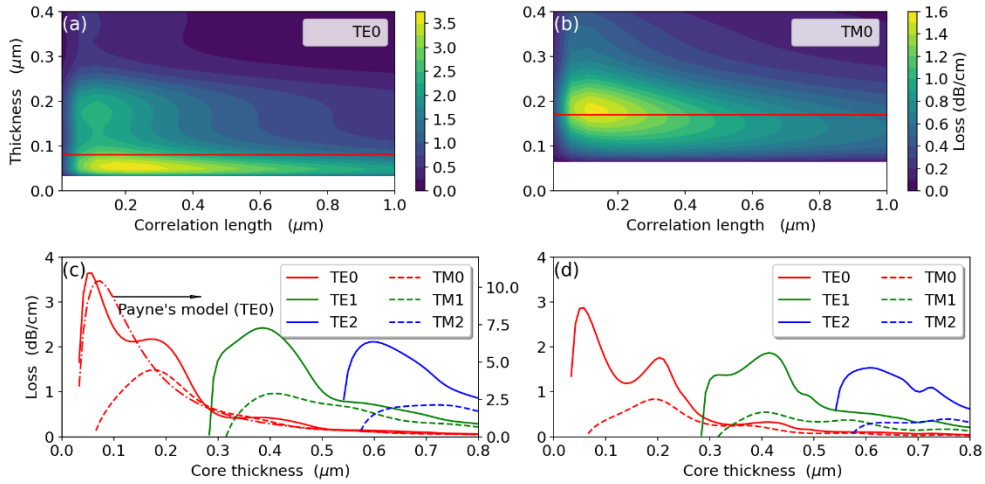

Figure 2-4 The propagation loss coefficient is plotted as a function of core thickness and correlation length for (a) TE_0 mode and (b) TM_0 mode. The surface roughness RMS amplitude is 1nm. The red lines mark the thickness that maximizes the surface modal field. (c) loss coefficient for the first three modes as a function of thickness with correlation length fixed at 180nm. The waveguide comprises a water-Ta_2O_5-SiO_2 three-layer structure. The dot-dash line corresponding to the right y-axis is calculated by the widely used Payne's model. (d) The scattering loss from the film on Si substrate with a $2\ \mu m$ oxidized (SiO_2) layer in between. The waveguide comprises a water-Ta_2O_5-SiO_2-Si four-layer structure. The waveguide structure is used for experimentally validate the model on scattering loss in the following section.

The correlation length is related to the spatial frequency components of the roughness and, combined with the core thickness dependent propagation constant and surface field, impacts the scattering. In contrast to the RMS amplitude, the correlation length has been less systematically studied. Ladouceur et al. considered that the correlation length, L, for a rough waveguide surface is in the range between 0.1 to $1\mu m$ [110]. Kita considered L to be ~100nm[41] and Schmid et al's measured value is 180nm, with which a good experimental and analytical agreement is achieved [53]. For deposited Ta_2O_5 thin films, the measured correlation length range from ~100nm [111], 400nm [112] and 49 to 282nm [113]. Due to the rather wide range of the correlation length reported, the relationship between scattering loss and correlation length is first investigated by calculating the loss of fundamental modes as a function of correlation length and core thickness using equations (2.18) and (2.19) . Figure 2-4(a) and (b), for the TE_0 and TM_0 modes respectively, show that the scattering loss does not change dramatically when the correlation length is in the range from 50nm to 400nm. The correlation length of 180nm measured by Schmid et al is inside range for the measured value for Ta_2O_5 thin film, and both roughness for the sidewall and deposited thin film are of exponential autocorrelation with similar RMS value, so 180nm is

adopted in this work as a representative value for correlation length. Another significant point is that the thickness that maximizes the surface field, marked by the red lines in Figure 2-4(a) and (b), generates strong scattering, although it does not necessarily maximize it. This shows that the impact of the core thickness on scattering loss should be quantified and taken into account into WERS waveguide optimization.

Figure 2-4(c) shows the propagation loss coefficient for both TM and TE polarizations versus core thickness for the first three modes. Obviously, the loss coefficient for different modes is not the same, so that the assumption that the loss coefficient for pump and signal are the same is not valid for signal collected in different modes. To illustrate the effect of the interference effect on scattering, the scattering loss calculated by the widely-used Payne's model, which does not consider interference, is also included in the same figure. The interference effect within the waveguide layers results in pronounced undulations and strong suppression of the scattering from ~10dB/cm down to ~3.5dB/cm.

Figure 2-4(d) shows the scattering loss from a $\mathrm{Ta_2O_5}$ film deposited on the Si substrate coated with a $2\ \mu m$ oxide layer. The reason for simulating this multilayer substrate is that oxidized silicon wafers have excellent surface quality, the silica layer enables low-loss waveguiding on a silicon substrate and they are convenient substrates to experimentally study the scattering loss. Due to the extra layer, the reflectivity and transmissivity for calculating the scattering loss is determined by the transfer matrix method [103,104] specified in section 2.2.2. The calculation shows that the extra layer impacts the interference, and consequently the scattering loss, significantly.

The above study shows that the waveguide core thickness impacts the surface modal field, efficiency of coupling the signal and the scattering loss significantly. Thus, the optimization for WERS should take these parameters into account, especially the scattering loss that strongly depends on the core thickness as it impacts the surface field and interference.

2.4 Waveguide fabrication and loss measurements

The scattering loss is shown to have a more complicated dependency on the core-thickness, but has been always oversighted or oversimplified. Moreover, experimental studies have rarely been carried out especially with the multilayer interference considered. To verify the theoretical prediction for the scattering loss, a systematic measurement of the loss

coefficient vs core thickness was carried out. Ta_2O_5 thin films with thicknesses between 90 nm and 360 nm were deposited on silicon wafers with a 2 µm oxidized layer.

2.4.1 Fabrication of waveguide

The main goal is to deposit thickness controlled, pure, uniform Ta_2O_5 film on silicon wafers. The sputtering, a Physical Vapour Deposition (PVD) method, in the integrated photonics cleanroom is a technique that provides thin film deposition with good adhesion, controlled thickness and good uniformity. To achieve minimal surface roughness, the sputtering parameters, including $\mathrm{A_r}: \mathrm{O_2}$ ratio, chamber pressure, and RF power, were optimized. The deposition parameters are as follows. The RF power is 250 Watts, $\mathrm{A_r}: \mathrm{O_2}$ ratio is 20:6 and chamber pressure is 7 mTorr. The sputtering target is a 150 mm diameter powder-pressed pure Ta_2O_5 bulk. The inclusion of oxygen is to avoid the oxygen deficiency when sputtering the metal oxides, even if the target is the desired material.

To fill any oxygen vacancies lost during the sputtering process and thereby reduce absorption due to the unoxidized metal, a process of annealing is adopted, following deposition. All samples were annealed at 550 ℃ for two hours. The temperature was ramped up to 550 ℃ at a rate of 50 ℃/min, and cooling down with a rate of 20℃/min.

During the fabrication process, the cleanness of the silicon wafer substrate is crucial to the quality of the deposition. Therefore, any contact with skin must be avoided and contamination must be removed before coating. Additionally, the plasma strike on the substrate, which is typically to further clean the substrate, should not be conducted to the wafer due to the side effect of worthening the flatness of the surface.

2.4.2 Characterization of waveguide

The coated Ta_2O_5 surfaces were characterized by a stylus profiler for surface roughness and the thickness of the Ta_2O_5 thin film. The stylus profiler has a similar resolution as atomic force microscopy (AFM) in the vertical direction for the rough surface, while its field of view is much larger, thus making the RMS amplitude measured from a wider range more reliable. The horizontal resolution for the stylus profiler is much lower than an AFM, so the correlation length obtained from the stylus profiler was not applied in this study. Instead, as discussed above, 180nm was adopted in our model as a representative value for the

correlation length. The auto-correlation function was calculated from the raw data from the stylus profiler so that the RMS amplitude is obtained for theoretical calculations, as shown in Table 2-1. As to the thickness of the thin film, a simple method is employed as follow. A line was drawn with a mark pen prior to coating, and then the ink can be removed along with the film after coating. The step height of the mark, which can be easily measured with the stylus profiler, is the thickness of the film.

The loss coefficient measurements were conducted with the prism coupling based Metricon machine. Only when the phase-matching angle is achieved, the incident laser power can be coupled into the waveguide in a certain mode. By recording and analysing the decaying light in the thin film with the built-in photodetector, an exponential curve can be fitted, thus obtaining the loss coefficient. The reliability of the measured loss coefficient could be affected by several factors. Firstly, the measured decaying profile of the coupled light might deviate from a strict exponential form, diminishing the precision of the exponential fitting for the loss coefficient determination. Secondly, the possible residue contamination on the substrate and the possible non-uniform thickness of the film can lead further irregularity and complexity. To mitigate such inaccuracy, multiple measurements were conducted for each sample.

2.4.3 Measured and calculated loss coefficients comparison

The experimental measurements and theoretical calculations are shown in Figure 2-5. The experimental loss coefficients are normalized to the mean square of the fluctuation of the rough surface, σ^2, since the scattering loss coefficient is proportional to it (see equation (2.18) and (2.19)), and the ideal roughness created by an optimized deposition recipe should be around 1nm in practice as mentioned above. The samples are deposited by plasma sputtering with different recipes, which were aiming at exploring the optimal recipe for the surface roughness, and their corresponding mean square of the fluctuation and original loss are given in Table 2-1. The loss measurements were carried out by a Metricon instrument operating at 632.8 nm in air, so the theoretical curves are re-calculated for this wavelength and are different to the ones in Figure 2-4(d). It is worth noting that the 120nm thick film was exposed to a longer plasma strike than the rest of the samples during the deposition process which may have damaged its surface and led to the increased loss values. Moreover, the 359nm core exhibits higher loss coefficients than the theoretical

prediction for fundamental modes most likely because the core is too thick to allow the annealing to oxidize all the tantalum. Therefore, with the fundamental modes mostly confined in the core higher absorption contributes to higher propagation loss. The good agreement between experimental and theoretical results validates the use of the Schmid et al's augmented scattering loss coefficient model and our extension of considering multilayer interference with the transfer matrix method.

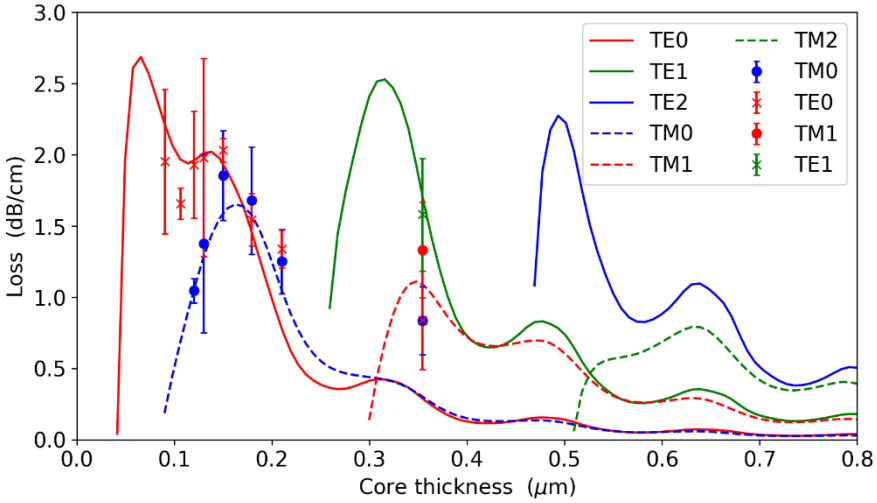

Figure 2-5. Measured loss coefficient and comparison with the theoretical value for different modes. The theoretical value is calculated by assuming the RMS, σ, is 1nm. The measured experimental loss coefficient is normalized to the mean square of the RMS amplitude, σ^2.

		90nm	106nm	120nm	130nm	150nm	179nm	210nm	359nm
$\sigma\ (nm)$		2.06	1.05	2.74	0.91	1.78	1.80	1.35	1.57
TE_0 (dB/cm)	loss	8.31	1.83	14.51	1.66	6.47	5.02	2.44	2.06
TM_0 (dB/cm)	loss	\	\	7.87	1.156	5.90	5.45	2.28	2.05
TE_1 (dB/cm)	loss	\	\	\	\	\	\	\	3.88
TM_1 (dB/cm)	loss	\	\	\	\	\	\	\	3.27

Table 2-1. Mean square of fluctuation and loss for different fabricated core thicknesses.

Both theoretical and experimental values show that a thicker waveguide suffers less scattering loss, and, for a waveguide with a specific thickness, TM modes typically will experience less loss when the Ta_2O_5 core is fabricated with the optimized recipe to reach an ideal roughness.

2.5 FOM Calculations

2.5.1 Single-Mode Pumping / Single-Mode Collection

In a practical device, the Raman signal is collected as the sum of each of the multiple spatial and polarization modes supported by the waveguide. This study first calculated the individual FOMs obtained with a single pump mode and a single collected signal mode, from a conventional three-layer waveguide. Results for TM_0 and TE_0 pumping are shown in Figure 2-6 and Figure 2-7. As shown, the signal collected by the lowest order fundamental mode has the highest efficiency, but the collection efficiency into higher order modes is not negligible, with weaker dependence on waveguide thickness and length. Comparing TM and TE excitation, the former is shown to give larger efficiencies.

The conventional FOM is calculated using Equations (2.29) and (2.30) with the waveguide thickness that maximizes the surface modal field of the pump, as discussed in Section 3.1. The conventional FOM for TE_0 and TM_0 pumping and collection and the corresponding waveguide parameters are listed in Table 2-2. It is shown that for randomly distributed dipoles, compared to TE_0 pumping/collection, TM_0 pumping/collection results in 2.3 times larger Raman conversion efficiency, due to higher optimum surface modal field and lower thickness-dependent scattering loss (see Figure 2-3(a) and Figure 2-4(c)). In addition, the optimum waveguide thickness and length are ~2 times larger. FOM for bulk analyte has also been presented in Table 2-2, providing supplementary information for planar WERS. The value of bulk FOM is much smaller than the monolayer FOM because the analyte density is not included. Considering the analyte molecule with ~nm scale, the bulk FOM will be greater by approximately two orders, which means the signal from a monolayer is only around a percent of that from the bulk analyte. In this scenario, obtaining more signal power is priority and is indicated by our new FOM.

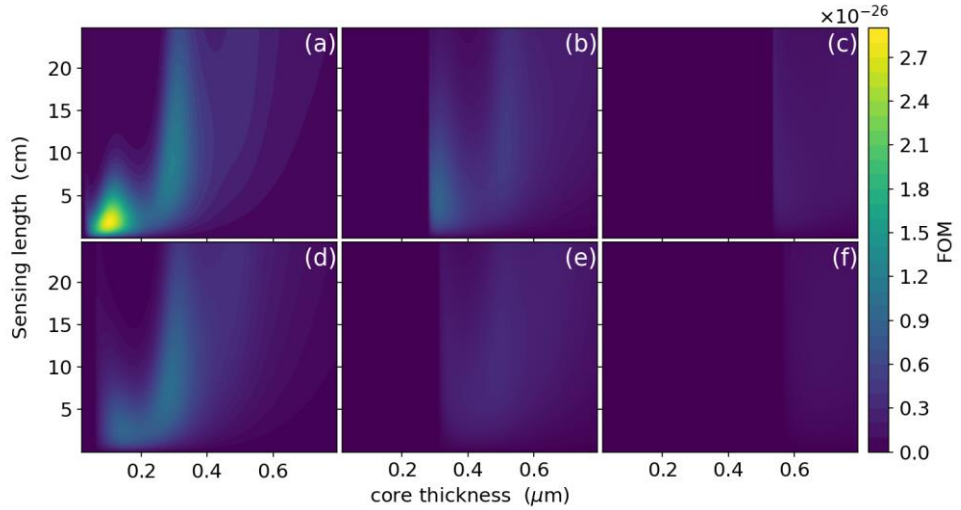

Figure 2-6. FOM calculations for forward collection as a function of core thickness and waveguide length for TE_0 pumping and signal collected as (a) TE_0, (b) TE_1, (c) TE_2 and (d) TM_0, (e) TM_1, (f) TM_2.

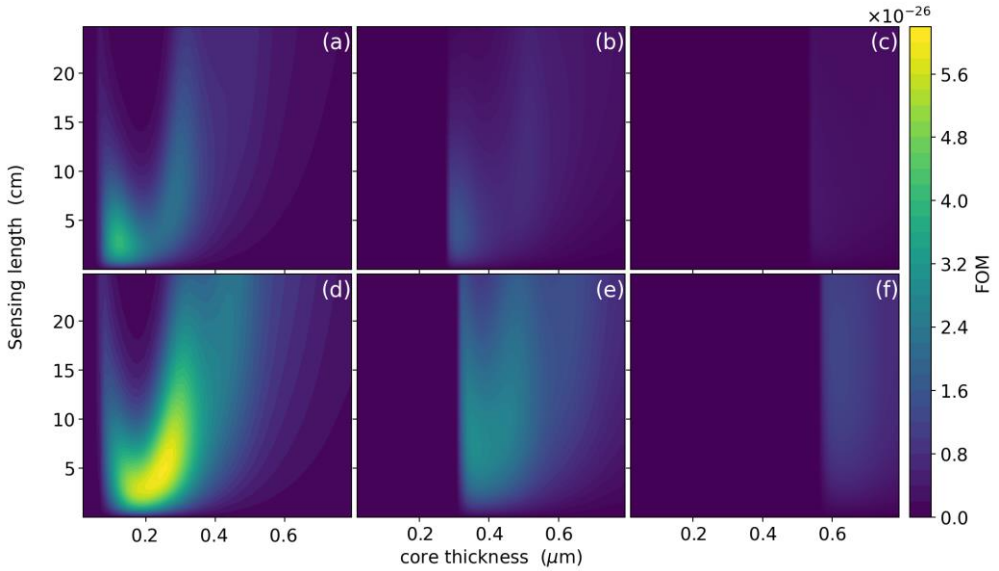

Figure 2-7. FOM calculations for forward collection as a function of core thickness and waveguide length for TM_0 pumping and signal collected as (a) TE_0, (b) TE_1, (c) TE_2 and (d) TM_0, (e) TM_1, (f) TM_2.

The relative FOM (rFOM) (defined as the new FOM (equation (2.31)), which takes into account the thickness dependence of the scattering loss, divided by the conventional FOM of TE_0) and the related optimal parameters are listed in Table 2-3 and Table 2-4 for forward and backward collection, respectively. It is evident that the new optimum waveguide parameters are significantly different from the conventional ones shown in Table 2-2. The new optimized waveguides are of longer length and larger thickness, due to the corresponding lower propagation loss in the presence of roughness. The TE_0

pumping/collection Raman conversion efficiency is 1.18 times larger when compared to the optimum conventional TE_0 case. The TM_0 pumping/collection Raman conversion efficiency, on the other hand, is 2.5 times larger when compared to the optimum conventional TE_0 case. This figure decreases to 1.1 when compared to the optimum conventional TM_0 case, due to the smaller difference in scattering losses in the corresponding optimum thicknesses (c.f. Figure 2-4(c)).

	Forward FOM	Backward FOM	Optimum Thickness (μm)	Length (cm)
TE_0	2.4×10^{-26}	3.26×10^{-26}	0.080	1.46
*TE_0 Bulk	8.51×10^{-25}	1.16×10^{-24}	0.071	1.34
TM_0	5.53×10^{-26}	7.51×10^{-26}	0.168	2.94
*TM_0 Bulk	1.91×10^{-24}	2.59×10^{-24}	0.168	2.94

Table 2-2. The optimal conventional FOM for TE_0 and TM_0 modes pumping and the corresponding optimum core thickness and optimum (forward collection) / effective (backward collection) waveguide length

* "Bulk" indicates the signal is excited and collected over the infinite water-filled cladding

	rFOM(*)		Thickness (μm)		Length (cm)	
Signal	TE_0	TM_0	TE_0	TM_0	TE_0	TM_0
TE_0	**1.181**	1.684	**0.105**	0.122	1.88	2.76
TE_1	0.367	0.664	0.312	0.309	4.27	4.02
TE_2	0.101	0.157	0.549	0.549	7.66	7.79
TM_0	0.421	**2.521**	0.288	**0.248**	7.16	4.65
TM_1	0.143	1.241	0.497	0.357	10.68	7.66
TM_2	0.053	0.519	0.687	0.617	13.94	15.95

Table 2-3. The optimal new relative FOM for TE_0 and TM_0 modes pumping and the corresponding optimum waveguide parameters for forward collected signal in each individual mode

(*) rFOM denotes the relative FOM, the ratio between new FOM and conventional FOM (Table 2-2) of TE_0 pumping

As can be clearly seen, the waveguide parameters are significantly different from the conventional FOM values due to the impacts of scattering loss, and the cross-polarization collection generates nonnegligible signal.

	rFOM		Thickness (μm)		Length (cm)	
Signal	TE_0	TM_0	TE_0	TM_0	TE_0	TM_0
TE_0	**1.183**	1.658	**0.105**	0.124	1.88	2.76
TE_1	0.342	0.630	0.309	0.308	3.77	3.64
TE_2	0.088	0.137	0.547	0.547	6.41	6.41
TM_0	0.421	**2.524**	0.288	**0.248**	7.16	4.65
TM_1	0.134	1.229	0.491	0.354	8.92	7.66
TM_2	0.047	0.460	0.679	0.609	10.68	12.56

Table 2-4. The optimal new relative FOM for TE_0 and TM_0 modes pumping and the corresponding optimum waveguide thickness and effective length for backward collected signal in each individual mode.

2.5.2 FOM Calculations – Single-Mode Pumping / Multi-Mode Collection

The single-mode pumping/collection cases studied in Section 2.5 have clearly demonstrated the impact of the waveguide thickness on the pump/signal surface intensities and the corresponding scattering loss, captured by the new FOM definition. However, in practice, the signal is always collected by all guided modes and the optimum waveguide parameters should maximize the overall Raman conversion efficiency. The FOM is calculated and shown in detailed contour map for the conventional three-layer waveguide at first. The overall rFOM for the four-layer structure, the Ta_2O_5 layer deposited on the silicon wafer with an oxidized layer which creates extra reflections and interference to reduce the scattering loss, is also calculated in this section.

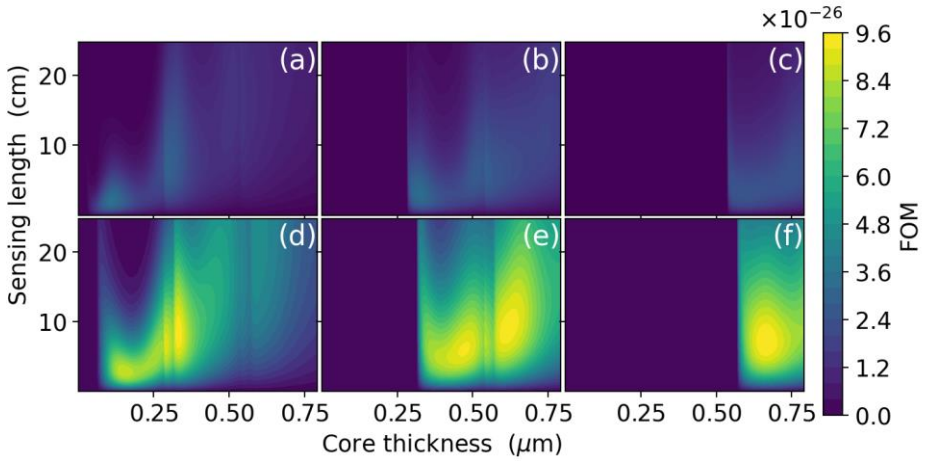

Figure 2-8. Total forward-collection FOM for different pumping polarization and modes. The pumping is in (a) TE_0, (b) TE_1, (c) TE_2 and (d) TM_0, (e) TM_1, (f) TM_2 polarization, for forward collection.

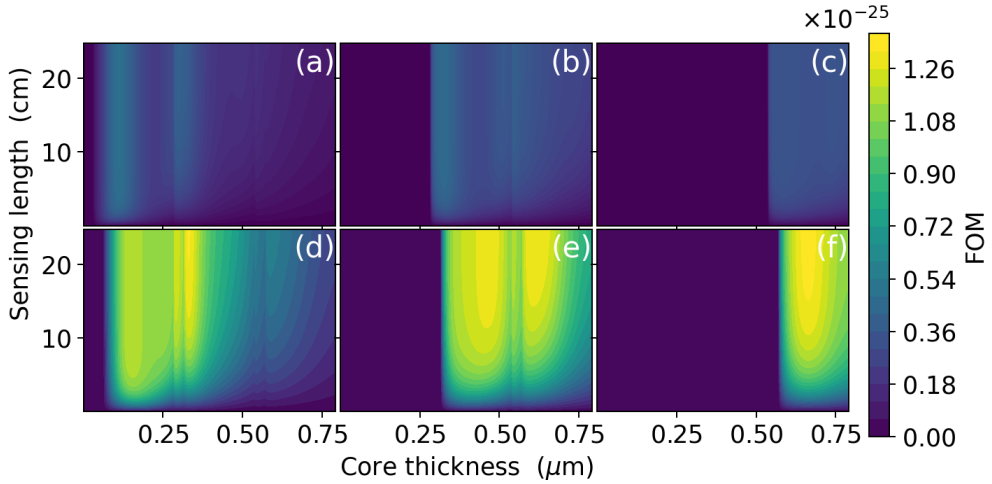

Figure 2-9. Total backward-collection FOM for different pumping polarization and modes. The pumping is in (a) TE_0, (b) TE_1, (c) TE_2 and (d) TM_0, (e) TM_1, (f) TM_2 polarization, for backward collection.

Using equation (2.32) and summing up Raman collection in all the spatial and polarization modes, the FOM that characterizes the total collected signal is shown in Figure 2-8 for forward collection. In general, the TM modes show better performance compared to TE modes excitation in terms of the FOM as well as core thickness and the length fabrication tolerances. From the contour maps in Figure 2-8(b) and (c), it is shown that the total FOM of higher-order mode pumping is comparable to the fundamental mode performance, despite the fact that higher-order modes have a much smaller surface modal field (c.f. Figure 2-2(a)). This is due to larger number of collected modes, the much lower scattering losses (c.f. Figure 2-3(c)) and the corresponding longer optimum sensing lengths.

The backward collection is also calculated and shown in Figure 2-9. As expected from equation (2.17), the FOM increases with the sensing length monotonically and saturates to a maximum value. Compared to forward collection, backward collection results in stronger Raman signal conversion efficiencies and wider sensing length fabrication tolerances. Therefore, the backward collection would be a better choice provided that the presence of the pump in-coupling method does not hinder the Raman signal collection.

The contour maps in Figure 2-8 and Figure 2-9 show multiple peaks due to the multimode collection, and cut-off of the guided modes due to core thickness is clearly indicated by the vertical lines. However, in addition to the FOM value, the waveguide thickness and sensing length fabrication tolerances need to be considered in real WERS systems and applications.

	rFOM		rSBR(*)		Length (cm)		Thickness (μm)	
Pump	Forward	Backward	Forward	Backward	Forward	Backward	Forward	Backward
TE_0	1.46	1.47	1.04	1.04	2.14	2.14	0.113	0.113
TE_0 Bulk	1.34	1.35	0.87	0.87	1.88	2.01	0.101	0.101
TE_1	1.40	1.40	0.63	0.63	3.02	3.02	0.325	0.325
TE_1 Bulk	1.81	1.81	0.69	0.69	3.27	3.27	0.306	0.308
TE_2	1.06	1.05	0.53	0.53	3.14	3.14	0.583	0.583
TE_2 Bulk	1.49	1.48	0.56	0.56	3.89	3.89	0.553	0.553
TM_0	3.96	4.07	0.92	0.92	8.29	8.54	0.330	0.330
TM_0 Bulk	4.08	4.10	2.01	2.00	3.14	3.27	0.128	0.126
TM_1	3.92	4.04	0.96	0.97	8.92	9.30	0.619	0.617
TM_1 Bulk	4.62	4.81	1.56	1.56	6.03	6.41	0.356	0.356
TM_2	3.99	4.12	1.43	1.43	7.04	7.54	0.665	0.664
TM_2 Bulk	4.95	5.14	1.60	1.59	7.29	7.79	0.629	0.628

Table 2-5. The optimal new relative total FOM for each single-mode pumping and the corresponding optimum waveguide thickness and effective length for the forward and backward signal collected by all supported modes. * rSBR indicates the relative signal noise ratio

The optimum rFOM and the corresponding waveguide parameters are listed in Table 2-5. It should be mentioned that in this case the optimum waveguide thickness and sensing length do not correspond to a particular mode but are average values that maximize the overall collection efficiency. With the growing thickness, the propagation loss drops and the corresponding waveguide length becomes longer, and deviates from the conventional reciprocal of the loss coefficient. When compared with the results in Table 2-3 and Table 2-4, it is clear that multimode collection is more efficient than the single-mode collection. It is also shown that the different TE/TM pumping modes provide comparable rFOM. This implies that in the case of randomly distributed dipoles, the single-mode pump excitation is not strictly required, and multimode pump excitation will give comparable conversion efficiencies. This is an important result as it relaxes considerably the fabrication and other experimental tolerances.

In Table 2-5, to present a full picture of the optimization, the rFOM and relevant waveguide parameters for bulk analyte are also provided, along with the relative signal-background-ratio r(SBR), which indicates the signal quality. The signal noise ratio (SBR) is calculated with equation (2.11), following Ref. [33,43]. The SBR normalized to the value calculated from the core thickness from conventional FOM, and is listed in Table 2-5 as rSBR. A concern could be arisen with our optimization that more waveguide material has overlapped with modal field so that more background could be generated. However, the calculated rSBR shows that signal quality is not significantly affected by our optimized waveguide parameters, and configurations for high-order TM modes can even enhance the SBR. The indicators, rSBR and FOM, for bulk analyte are following the trend for monolayer FOM.

The same FOM analysis is also carried out for the waveguide with oxidized silicon wafer substrate, whose scattering loss behaviour is shown in Figure 2-3(d). The results are summarized in Table 2-6 and Table 2-7. Compared to the previous simpler three-media planar waveguide, this composite waveguide provides much superior conversion efficiency performance. This is due to the improved scattering losses, as a result of the additional interference effects from the oxidized SiO_2/Si interface. This is important as it clearly demonstrates the significance of interference effects on the waveguide loss performance and the WERS sensor overall efficiency.

	Forward FOM	Backward FOM	Optimum Thickness (μm)	Length (cm)
TE_0	3.08×10^{-26}	4.18×10^{-26}	0.080	1.88
TE_0 Bulk	1.08×10^{-24}	1.46×10^{-24}	0.071	1.88
TM_0	1.08×10^{-25}	1.47×10^{-25}	0.168	5.77
TM_0 Bulk	3.47×10^{-24}	5.08×10^{-24}	0.168	5.77

Table 2-6. The optimal conventional FOM for TE_0 and TM_0 modes pumping and the corresponding optimum core thickness and optimum (forward) / effective (backward) waveguide length for the multilayer oxidized silicon wafer substrate

Pump	rFom		rSBR		Length (cm)		Thickness (μm)	
	Forward	Backward	Forward	Backward	Forward	Backward	Forward	Backward
TE_0	1.92	1.93	0.95	0.96	3.89	3.89	0.128	0.126
TE_0 Bulk	1.65	1.67	0.73	0.73	3.64	3.77	0.120	0.120
TE_1	1.68	1.68	0.64	0.64	4.65	4.52	0.334	0.334
TE_1 Bulk	2.06	2.06	0.72	0.72	4.65	4.52	0.324	0.324
TE_2	1.21	1.20	0.37	0.37	6.66	6.41	0.696	0.696
TE_2 Bulk	1.64	1.62	0.55	0.55	5.40	5.28	0.553	0.553
TM_0	6.04	5.87	0.92	0.92	17.21	14.57	0.330	0.330
TM_0 Bulk	6.61	6.68	2.04	2.04	6.53	6.78	0.132	0.132
TM_1	6.15	6.17	1.30	1.32	13.07	12.31	0.479	0.476
TM_1 Bulk	6.87	6.99	1.45	1.50	12.94	12.06	0.349	0.352
TM_2	6.30	6.23	1.40	1.40	15.33	13.82	0.685	0.683
TM_2 Bulk	7.37	7.35	1.55	1.56	14.82	13.44	0.656	0.653

Table 2-7. The optimal new relative total FOM for each single-mode pumping and the corresponding optimum waveguide thickness and effective length for the front and backward signal collected by multimode for the multilayer oxidized silicon wafer substrate

2.6 Conclusions

Conventional optimization of waveguides for advanced WERS sensor applications is based on defining the waveguide dimensions that maximize the pump and signal surface intensity under pump/signal single-mode operation. However, this approach overlooks the fact that maximization of the surface intensity results inevitably in increased surface-induced scattering losses, increase the waveguide propagation losses and reduce the overall Raman collection efficiency.

This work has first studied theoretically and experimentally the impact of planar waveguide thickness on surface scattering losses and the waveguide propagation losses. It is shown that surface scattering losses are strongly influenced by interference effects due to reflections at the core/cladding and core/substrate interfaces, as well as additional reflections from other interfaces at adjacent substrate layers. It should be mentioned that this is similar to the reduction of propagation losses in anti-resonant waveguides [114] and hollow-core anti-resonant fibres [115,116]. In the case of Ta_2O_5 waveguides in silica-on-silicon substrates with typical surface roughness parameters, it is shown that the peak propagation losses reduce from ~10dB/cm to below ~3dB/cm for TE_0 modes and below ~2dB/cm for TM_0 modes, when the interference with additional reflections from the substrate interfaces is taken into account. The propagation losses are even lower for higher-order modes. These results are in good agreement with experimental data.

In our study, we consider radiation from a Raman-emitting dipole on a waveguide can be captured back into the waveguide in polarization and spatial modes different from the mode carrying the pump, provided the waveguide can support them. In the case of randomly distributed dipoles, we consider the Raman gain coefficients corresponding to all possible combinations of pump and signal polarization and spatial modes. It should be noted that, in this case, single-polarization pump will produce both co-polarized and cross-polarized Raman signal. Our initial WERS experimental results have confirmed both co-polarized (TE) and cross-polarized (TM) signal collection from a TE pumped waveguide, and the ratio of the observed power in different polarizations agrees well with our theory.

This chapter introduced a new generalized FOM to optimize planar waveguide-based WERS sensors under multimode excitation/collection operation and waveguide-thickness- and mode-dependent propagation losses. Our new FOM suggests, for example, that once the

core thickness increases from ~80nm, which is the optimum value given by traditional FOM for a Ta_2O_5 on SiO_2 TE_0 waveguide at 785nm, the signal power will increase significantly due to the reduced scattering loss and higher number of modes supported by the waveguide, as shown in Table 2-5. It is shown that in comparison with the traditional optimization, our optimized larger-core waveguides result in a WERS signal collection efficiency increase by a factor of ~1.5 when using TE_0 pumping and a factor of ~5 when using TM_0 pumping. It is also shown that the improvement increases to ~4 when TM_1 and TM_2 pump is used. When the waveguide is deposited on a multilayer oxidized silicon wafer substrate, improvements by a factor of ~6 to ~7 are observed when the pumping mode order increases from TM_0 to TM_2 (see Table 2-7). The FOM increase with the mode order is a result of the combination of substantial reduction in propagation loss, increased number of collected modes, longer optimum sensing lengths and occurs despite the concomitant surface intensity decrease. This implies that in the case of randomly distributed dipoles, the single-mode pump excitation is not strictly required, and multimode pump excitation will give superior conversion efficiencies. This is an important result as it relaxes the fabrication and other experimental tolerances considerably, and it is expected to result in more robust, reproducible, and cheaper WERS sensors.

Compared to the singlemode waveguide counterpart, use of multimode waveguides and multimode signal collection is likely to affect the spectroscopic measurements. This will be the case, if the multimode waveguide signal output is collimated and directly fed into the spectrometer. However, in most WERS system implementations, a multimode fibre is used at the output to collect the signal and feed it into the spectrometer [35,36,117,118]. In this case, the input to the spectrometer is highly multimoded even in the case of a singlemode waveguide and it will not further affect the spectroscopic measurements.

Finally, this work has shown the significance of considering the propagation losses, due to surface roughness, when designing for optimised waveguides for WERS applications. Although the study is focused on planar slab waveguides, it is expected that the proposed design strategy will be even more important in the case of other waveguide types, such as rib or slot waveguides, where the surface roughness and associated propagation losses will be mode pronounced due to the additional etching processes.

Chapter 3 Modelling Waveguide grating coupler for WERS I—basic theorems and model

As discussed in introduction chapter, the waveguide grating coupler is selected to couple pump light for our WERS system[42,119], due to the stability, high-efficiency, and good alignment tolerances. Simple grating structures have been studied for over a century [120]. However, it is quite involved to calculate the coupling efficiency when used as an input coupler for waveguides or fibres. Unlike the surface scattering problem, which is stochastic and random by nature, gratings are deterministic and can be numerically solved with good accuracy. Many iterative searching methods, including one of our group [42], using numerical solutions of Maxwell equations, have been proposed for better design of the grating coupler [121]. However, these methods are still facing several difficulties, which will be specified below. Therefore, this chapter aims to build a theoretical model to understand the physical mechanisms and for improved WGC optimization. The theoretical model has been developed and validated with the comparison to our previous publication on numerical WGC optimization for our WERS system [42].

3.1 Background

Waveguide grating couplers (WGCs) are proving to be essential components for coupling light from fibre/free space into integrated waveguide circuits and have found increasingly widespread use in areas such as silicon photonics [122] and, more recently, in bio-medical sensing[42]. Over the years, WGCs have attracted extensive research related to their underlying physical mechanisms and efficiency optimization. The WGC input coupling efficiency (CE) is usually treated analytically by considering its reciprocal problem of out-coupling or output scattering [123–125]. However, the output scattering analytical models are usually based on a rather restrictive assumption that the grating depth is small[123], which makes the modelling inaccurate in a number of practical situations. For more accurate calculations, numerical simulation is the mainstream approach for WGC optimization [42,121].

However, full numerical models suffer from several shortcomings. Firstly, the WGC CE is affected by a number of factors, and a full optimization taking into account all contributing

parameters can result in prohibitively long computing times. The computational cost of the multi-factor-based optimization increases almost exponentially with the number of the impacting factors, and the FDTD or FEM methods for solving Maxwell equations are known to require intense computation. Secondly, numerical searching for the most efficient solutions could converge on a local optimum, especially when the parameter space is large. Finally, important underlying physical mechanisms can be obscured, and significant physical insight can be lost.

In this chapter, the basic theories, including reciprocal theorem, theory on grating scattering order, are introduced. Sychugov et al.'s model [123], as a rare model that can be used to calculated the CE, is adopted for reproducing our previously published numerical results. The inaccuracy of Sychugov et al.'s model has been analysed and the model has been modified for significant improvement. Important underlying physical mechanisms have been revealed with Sychugov et al.'s model and the modifications.

3.1.1 Grating-assisted coupling efficiency (CE) – reciprocity in WGC

Waveguide grating-assisted waveguide out-coupling and in-coupling is a pair of reciprocal problems. Analytically calculating the in-coupling efficiency (CE) can be a quite complex problem. However, designing the in-coupler is easier to be accomplished by analysing the out-coupling problem first. The reciprocal and direct problem is shown in Figure 3-1(a) and (b), respectively.

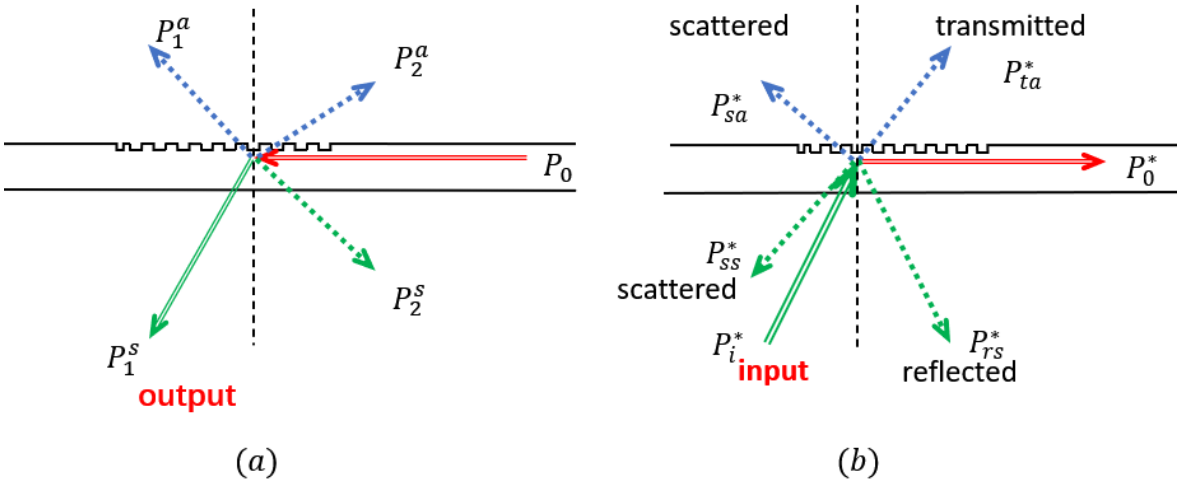

Figure 3-1. Illustration of reciprocity of (a) reciprocal out-coupling and (b) direct in-coupling of a waveguide grating coupler. n_f, n_s and n_a denote the refractive index of core, substrate, and cladding respectively.

Supposing the waveguide is lossless except for the scattering in the grating area, due to the conservation of energy, the power carried by each scattered radiation and surface wave fits the following relation,

$$P_0 = \Sigma_i (P_i^s + P_i^a) \tag{3.1a}$$

$$P_i^* = P_0^* + \Sigma_i (P_{is}^* + P_{ia}^*) + P_{ta}^* + P_{rs}^* \tag{3.1b}$$

The parameters are shown in Figure 3-1. P_0 is the power of guided mode. i denotes the diffraction order, s and a stand for the substrate and superstrate cladding. The asterisk denotes the in-coupling configuration. P_{ta}^* and P_{rs}^* are the transmitted and reflected coupling power. As a reciprocal procedure, the electromagnetic field fits the Lorentz reciprocity theorem,

$$\int_{A\infty} \{\mathrm{E} \times H^* - E^* \times H\} \cdot z dA = 0 \tag{3.2}$$

in which, E and H are the corresponding electric and magnetic fields. According to the Lorentz reciprocity theorem, $\int_A \mathrm{E_j} \times H_k^* \cdot z dA = \int_A \mathrm{E_k^*} \times \mathrm{H_j} \cdot \mathrm{z} dA = 0$ if j and k correspond to waves travelling in different directions, and the only non-zero pairs are the ones that travel in opposite directions. As shown in Figure 3-1, only the [P_i^*,P_{1s}] and [P_0^*,P_0] are the reciprocal pairs. The relation between the reciprocal powers is given by Dalgoutte [125], namely

$$\frac{P_0^*}{P_i^*} = \left[\int_{-\infty}^{\infty} g(z)h(z)dz\right]^2 \frac{P_i^s}{P_0} \tag{3.3}$$

where $g(z)$ and $h(z)$ are the profiles of the normalized field distribution of the in-coupling P_i^* and out-coupled P_i^s fields as shown in Figure 3-2.

For semi-infinite gratings, the profile of scattered power into substrate, P_i^s , is approximately exponentially decaying due to the radiation loss of the grating, namely, $g(z) = \exp(\alpha z)$ for $z < 0$, and $g(z) = 0$ for $z > 0$. Therefore, g(z) is non-zero only over the grating length L. For the in-coupling Gaussian beam, $h(z)$ is the projection on z axis of the normalized Gaussian profile with an incident angle of θ_{in}^*. The two profile functions are normalized on the (y'z') plane such as $\int_{-\infty}^{0} [g(z')]^2 dz' = \int_{-\infty}^{\infty} [h(z')]^2 dz' = 1$. In this case, the two field profile functions take the form [125,126],

$$g(z') = \sqrt{2\alpha_{tot}} exp(\alpha_{tot} z') \tag{3.4}$$

$$h(z') = \sqrt{\frac{(2/\pi)^{1/2}}{\omega_{in}}} exp\left[-\left(\frac{z'}{\omega_{in}}\right)^2\right] \tag{3.5}$$

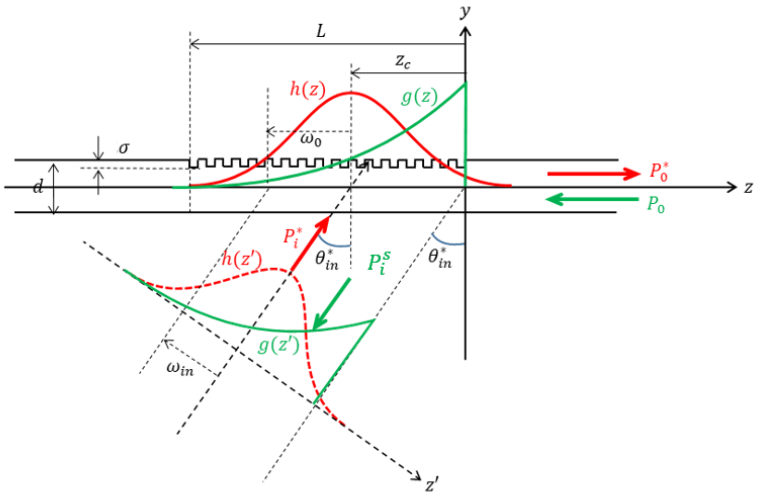

Figure 3-2. Reciprocal modes of in-coupling (red) and out-coupling (green) from the grating. The incident (first-order diffraction) coupled in (out) from the grating with the angle of θ_{in}^*. The full-depth of the grating is σ, and the waveguide thickness is d. The beamwidth of in-coupling plane ($y'z'$) is ω and ω_0 is the projected beamwidth on the grating plane (yz).

Projecting the above functions on the (yz) grating plane ($z = z'/cos(\theta_{in}^*)$), Equation (3.3) takes the form:

$$\eta_i = \frac{P_0^*}{P_i^*} = \eta_{max} \frac{P_i^S}{\Sigma_i(P_i^S + P_i^a)} = \eta_{max} \frac{\alpha_i^S}{\Sigma_i(\alpha_i^S + \alpha_i^a)} = \eta_{max} \frac{\alpha_i^S}{\alpha_{tot}} \tag{3.6}$$

where η_i is the waveguide coupling efficiency (CE), $\alpha_{is(a)}$ are the i^{th} order radiation loss coefficients into substrate (s) and superstrate (a), and [126]

$$\eta_{max} = \frac{2\alpha_{tot}\sqrt{2/\pi}}{\omega_0}\left\{\int_{-L}^{0} exp(\alpha_{tot} z) exp\left[-\left(\frac{z+z_c}{\omega_0}\right)^2\right] dz\right\}^2 \tag{3.7}$$

η_{max} is the maximum achievable grating-assisted waveguide coupling efficiency, given as a function of the grating length L, the total radiation loss $\alpha_{tot} = \Sigma_i(\alpha_{is} + \alpha_{ia})$, the Gaussian beam offset z_c from the grating edge (z=0) and the projected Gaussian field radius $\omega_0 = \omega_{in}/cos\theta_{in}^*$.

In the case of a semi-infinite grating ($L \to \infty$), the integral of Equation (3.7) attains the maximum value $\eta_{max} = 0.81$ when the following conditions are met [124],

$$\begin{aligned} 2\alpha_{tot} w_0 &= 1.36 \\ \frac{z_c}{w_0} &= 0.733 \end{aligned} \tag{3.8}$$

In this case, the optimal coupling efficiency takes the simplified form [123],

$$\eta_i = 0.81 \frac{\alpha_{is}}{\Sigma_i(\alpha_{is}+\alpha_{ia})} \tag{3.9}$$

In Figure 3-3 we plot the maximum coupling efficiency (equation (3.7) for the case of semi-infinite gratings ($L \to \infty$)) as a function of total radiation loss (α_{tot}), Gaussian beam-width (ω_0), and beam offset (z_c). It is shown that $\eta_{max} \approx 0.81$ when $\alpha_{tot} w_0 \approx 0.68$ and $\frac{z_c}{w_0} \approx 0.733$, in close agreement with conditions (3.8).

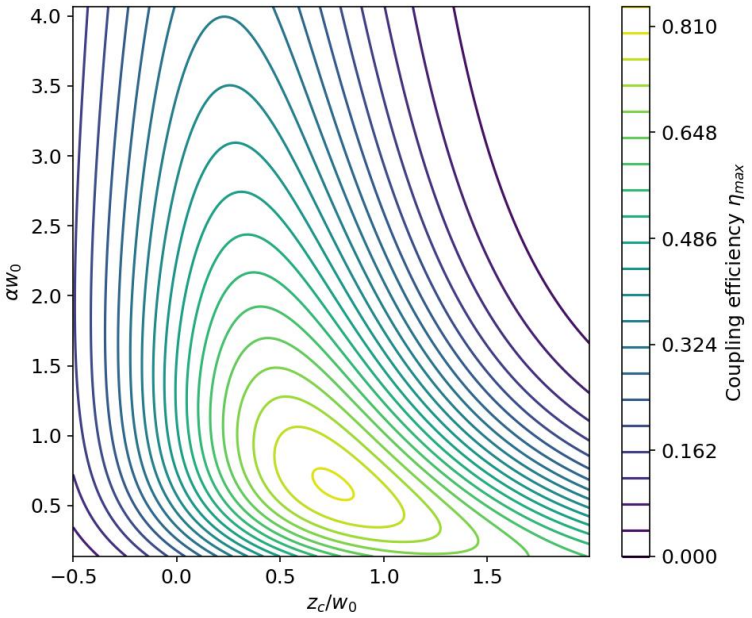

Figure 3-3. Contour map of the coupling efficiency (η_{max}) of semi-infinite gratings ($L \to \infty$) as a function of total radiation loss (α_{tot}), Gaussian beam-width (ω_0), and beam offset (z_c).

From Equation (3.6), it becomes obvious that in order to achieve the maximum coupling efficiency with uniform gratings, in addition to optimizing η_{max} as shown above, we should design the grating/waveguide system in the reciprocal arrangement to radiate only into one substrate order. In this case $\eta_i = \eta_{max} \approx 0.81$. Coupling efficiencies $\eta_i > 0.81$ can be achieved by properly apodising the grating so that the out-coupling shape matches the

in-coming beam profile ($g(z') \cong h(z')$). This, however, can complicate the grating design and fabrication and therefore can increase the cost and reduce manufacturability and it is not considered here.

3.1.2 Grating diffraction orders

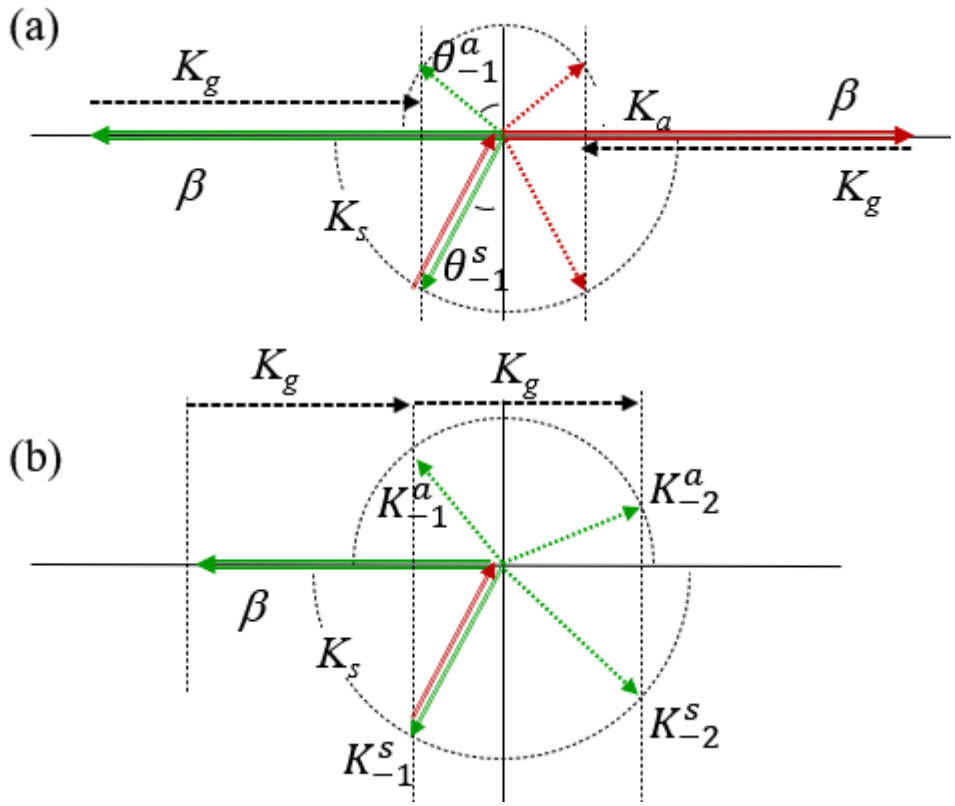

Figure 3-4. Diagram of in/out-scattering orders (a) Reciprocal in/out put scattering orders (b) Multiple scattering orders

As mentioned above, the in-coupling efficiency can be analysed via the reciprocal out-coupling problem. The diffracting grating was initially studied by Rayleigh[120]. The Rayleigh approach is valid for weak gratings and is not suitable for strong deep-grooved gratings [127]. According to Rayleigh expansion, for TE mode, the diffracted electric field of the j^{th}order is,

$$E_n(y,z) = A_n e^{-ik_{i,jz}z} e^{-ik_{i,jy}y}$$

where,

$$k_{i,ny} = k_i \cos\theta_{i,j}, \quad k_{i,nz} = k_i \sin\theta_{i,j}$$

i.e.

$$k_{i,jy}^2 + k_{i,jz}^2 = k_i^2 \tag{3.10}$$

in which $k_i = \frac{2\pi n_i}{\lambda}$ is the wavevector determined by the grating function, and n_i is the refractive index of medium i. For the three-layer waveguide, which is shown in Figure 3-4, media i are cladding (denoted by a), core (denoted by f) and substrates (denoted by s) respectively. The variable $k_{i,jy}$ and $k_{i,jz}$ are the wave vector of j^{th} diffraction order along y and z axis. $k_{i,jz}$ is determined by,

$$k_{i,jz} = k_{f,0z} + jK, \tag{3.11}$$

where K is the spatial frequency of the grating, $K = \frac{2\pi}{P}$, and P is the period (pitch size) of the grating. And $k_{f,0z}$ is the propagating constant of the waveguide, and $k_{f,0z} = \beta_w = \frac{2\pi n_{eff}}{\lambda} = k_{f,0}\cos\theta_{f,0}$. By substituting (3.11) into (3.10)

$$k_{i,ny} = \pm\left[k_i^2 - (\beta_w + jK)^2\right)^{\frac{1}{2}} \tag{3.12}$$

in free space, i.e. the radiative field in cladding or substrate, $|k_{i,ny}|$ and $|k_{i,nx}|$ can be any value between 0 and k_i, otherwise $k_{i,ny}$ or $k_{i,nz}$ would be imaginary. But for the $k_{f,nz}$ in the waveguide, it can only take the discrete values $\pm\beta_m$, the Propagating constant of the m^{th} waveguide mode supported by the waveguide. Here, only the single-mode waveguide is taken into consideration. A typical out-coupling schematic from waveguide grating as the reciprocal of forward coupling setup is shown in Figure 3-4. The guided mode is propagating along the negative z direction, and the -1st-diffraction order goes into the substrate, (which corresponds to the in-coupling beam of the reciprocal arrangement) with a small angle $\theta_{s,-1}$(reciprocal of the in-coupling angle θ_i^*). So, for the -1st-diffraction order $k_{i,-1z} = \beta_w - K$. According to Snell's law, $k_{i,-1z} = k_f sin\theta_{f,-1} = k_a sin\theta_{a,-1} = k_s sin\theta_{s,-1}$. Therefore, the grating period, in this case, is given by

$$P = \frac{2\pi}{\beta_w - k_s sin\theta_i^*}. \tag{3.13}$$

The number of diffracting orders is discussed next. For the reciprocal problem of forward in-coupling configuration, as shown in Figure 3-4, the diffracted $k_{i,-1z}$ must be of same direction as the main wave β_w, which is negative z in Figure 3-4. So, the diffraction order

j can only take negative values, otherwise $|k_{j,-1z}| > \beta_w$ which means it cannot be propagating in any one of the three layers.

For the shown three-layer waveguide grating, typically, the wavevectors in each layer satisfy $k_f > \beta_w = k_f cos\theta_{f,0} > k_s > k_a$. So, only a few orders can be supported. For the negative j, from equation (3.12) it is clear that when $|\beta_w + jK| < |k_i|$, the j^{th} order could be radiated. So, the maximum number of orders is

$$|j|_{max} = \left\lfloor \frac{\beta_w + k_i}{K} \right\rfloor = \left\lfloor \frac{\beta_w + k_i}{\beta_w - k_s sin\theta_i^*} \right\rfloor = \left\lfloor \frac{n_{eff} + n_i}{n_{eff} - n_s sin\theta_i^*} \right\rfloor \quad (3.14)$$

In order to minimize the number of orders, for the forward coupling arrangement, the incident angle θ_i^* should be close to but larger than zero.

In the core region, assuming $n_s sin\theta_i^*$ is positive but small, $|j|_{max} = \left\lfloor \frac{2n_{eff}}{n_{eff} - n_s sin\theta_i^*} \right\rfloor = 2$. This indicates that there could be a -2nd order. Taking $j = -2$ and equation (3.13) into equation (3.11), it can be found that $k_{i,-2z} = -\beta_w + 2k_s sin\theta_i^*$.

So, the θ_i^* cannot be zero to prevent the -2nd order from being coupled into a backward mode.

In the substrate region, $|j|_{max} = \left\lfloor \frac{n_{eff} + n_i}{n_{eff} - n_s sin\theta_i^*} \right\rfloor = \left\lfloor 1 + \frac{n_i + n_s sin\theta_i^*}{n_{eff} - n_s sin\theta_i^*} \right\rfloor$. When $\theta_i^* > \arcsin\left(\frac{n_{eff} - n_i}{2n_s}\right)$, the -2nd order will be radiated into medium i.

To avoid the -2nd-order radiation in the forward coupling, the in-coupling angle should lie within

$$0 < \theta_i^* < \arcsin\left(\frac{n_{eff} - n_s}{2n_s}\right) \quad (3.15)$$

When the index contrast of the waveguide is small or the core is thin, n_{eff} will be very close to n_s, so the $k_{i,-2z}$ will be very close to $-\beta_w$. According to the coupled mode theory, it is possible to be coupled into the back-propagating mode, which reduces the in-coupling efficiency.

For backward in-coupling, $sin\theta_i^* < 0$, and there will always be the -1st order in the substrate.

3.1.3 Radiation loss coefficients of a rectangular-teeth grating

The loss coefficient of the waveguide grating has been investigated by Kiselev's [128], [129] and Tishchenko's [130]. The model was built for a waveguide grating with a sinusoidal boundary and was theoretically solved from the Maxwell equation and the two continuous boundary conditions.

The radiation loss coefficients of the -1st diffraction order into the superstrate ($\alpha_{\mathrm{a}}^{(-1)}$) and substrate ($\alpha_{\mathrm{s}}^{(-1)}$) cladding are given by Sychugov et al. [123] in 1997. This is a rare theoretical analytical result that explicitly gives the scattering into upper cladding and substrate,

$$\alpha_{\mathrm{a}}^{(-1)} = \left(\frac{k\sigma^*}{2}\right)^2 \frac{n_{\mathrm{f}}^2 - n_{\mathrm{eff}}^2}{n_{\mathrm{e}} h_{\mathrm{e}}} \times \frac{\left(n_{\mathrm{f}}^2 - n_{\mathrm{a}}^2\right) N_{\mathrm{a}} \left[N_{\mathrm{s}}^2 + \left(n_{\mathrm{f}}^2 - n_{\mathrm{s}}^2\right)\cos^2\psi\right]}{\left(N_{\mathrm{f}}^2 + N_{\mathrm{s}} N_a\right)^2 - \left(n_{\mathrm{f}}^2 - n_{\mathrm{s}}^2\right)\left(n_{\mathrm{f}}^2 - n_{\mathrm{a}}^2\right)\cos^2\psi} \quad (a)$$

(3.16)

$$\alpha_{\mathrm{s}}^{(-1)} = \left(\frac{k\sigma^*}{2}\right)^2 \frac{n_{\mathrm{f}}^2 - n_{\mathrm{eff}}^2}{n_{\mathrm{e}} h_{\mathrm{e}}} \times \frac{\left(n_{\mathrm{f}}^2 - n_{\mathrm{a}}^2\right) N_{\mathrm{s}} N_{\mathrm{f}}^2}{\left(N_{\mathrm{f}}^2 + N_{\mathrm{s}} N_{\mathrm{a}}\right)^2 - \left(n_{\mathrm{f}}^2 - n_{\mathrm{s}}^2\right)\left(n_{\mathrm{f}}^2 - n_{\mathrm{a}}^2\right)\cos^2\psi} \quad (b)$$

where $N_j = \sqrt{n_j^2 - \left(n_{\mathrm{eff}} - \frac{\lambda}{d}\right)^2}$, $\psi = kN_{\mathrm{f}}d, j = \mathrm{a,f,s}$, and d is the thickness of the grating coupler. h_{e} is the effective thickness of the waveguide and n_{e} is the mode effective index. The original formula of equation (3.16) is derived for grating with the sinusoidal profile. Therefore, for a rectangular grating profile, the parameter $\sigma^* = \frac{2}{\pi}\sigma$ is the equivalent half depth of the sinusoidal grating, which corresponds to the first-order Fourier component of the rectangular grating corrugation. For the small-angle incidence, the -3^{rd} order will not be generated and -2^{nd} order will not be generated when the grating profile is an even function because the Fourier coefficient of even order is zero for an even function, which is the case for rectangular grating.

Thus, the in-coupling efficiency (CE) is purely dependent on α_{-1}^{a} and α_{-1}^{s} [123]

$$\eta_{-1} = \eta_{max}\frac{\alpha_{\mathrm{s}}^{(-1)}}{\alpha_{\mathrm{tot}}} = \eta_{max}\frac{\alpha_{\mathrm{s}}^{(-1)}}{\alpha_{\mathrm{s}}^{(-1)}+\alpha_{\mathrm{c}}^{(-1)}} = \eta_{max}\frac{N_{\mathrm{s}}N_{\mathrm{f}}^2}{N_{\mathrm{s}}N_{\mathrm{f}}^2+N_{\mathrm{a}}\left[N_{\mathrm{s}}^2+\left(n_{\mathrm{f}}^2-n_{\mathrm{s}}^2\right)\cos^2\psi\right]} \qquad (3.17)$$

where $\alpha_{\mathrm{tot}} = \alpha_{-1}^{a} + \alpha_{-1}^{s}$ The required grating pitch size (period) can be calculated from the grating functions (3.11) ~ (3.13), and the grating depth can be found by fulfilling the optimal overlap condition (3.8), $\alpha_{tot}w_0 = 0.68$.

Equation (3.17) shows that the coupling efficiency does not depend explicitly on the grating amplitude σ^*. However, the grating amplitude affects η_{max} through the total radiation loss coefficient α_{tot} (see Equations (3.7) and (3.16)), which defines the extent of the exponentially decaying function $g(z)$ (see equation (3.4)). In case the function $g(z)$ exceeds the grating length *L*, the maximum efficiency drops below the ultimate value ($\eta_{max} < 0.81$).

3.2 Results of Sychugov et al.'s model

To validate the model established in section 2.2, the analytical theoretical model is compared with our numerical results[42]. The numerical model is simulated with finite-deference time domain (FDTD) method-based commercial software $Lumerical^{TM}$ for following WGC configuration shown in Figure 3-5. The period of grating, marked as pitch P, and etch depth d are the parameters optimized by the swamp particle algorithm. The substrate is $\mathrm{SiO_2}$ and the cladding is air. The beam width is 25 μm and the grating length is fixed to 50μm. The Gaussian source is set inside the $\mathrm{SiO_2}$ substrate with a launching angle $\theta_{sub} = 5.5^o$, which is equivalent to a lunching angle in air of $\theta_{air} = 8^o$. The horizontal position of the input source, marked as the offset position z_c, is obtained by parameter sweep after obtaining the P and d. The reason for not optimizing z_c with P and d at the same time together because the three-dimensional optimization for WGC with FDTD is enormously time-consuming, and it is even too difficult for a supercomputer to conduct such computational workload. Meanwhile, as shown in Figure 3-3, the z_c does not change the CE coupling efficiency dramatically. Therefore, the results from the numerical model that simulated with the state-of-art commercial software are trustworthy and have been published recently[42], and are used as benchmarks for validating our theoretical models in the following sections.

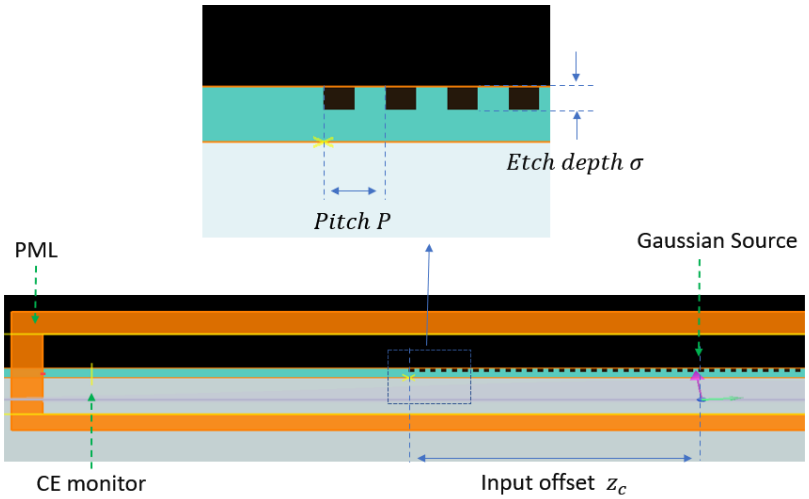

Figure 3-5. The numerical model for WGC, including the whole simulation area confined by the perfect match layer (PML) and the grating structure at the grating-waveguide joint.

As to the theoretical model, it is confirmed that there is only one scattering order in cladding and substrates, so equation(3.16) is used to calculate the scattering strength into cladding and substrate. We assume that, as a global optimization method, the swamp particle method could be able to find the optimal P and d pair. Therefore, the CE is calculated by Equation (3.17) with $\eta_{max} = 0.81$. The effective index n_{eff} is obtained by solving the dispersion equation of a typical three-layer waveguide, which is a function of the indices of each layer and the thickness of the core layer. The parameters of $N_{f(s)}$ are readily known with n_{eff}.

Firstly, the model is applied to the thin waveguide with a core thickness that maximizes the core-cladding surface intensity. The maximum surface intensity and its corresponding core thickness as a function of the core index are shown in Figure 3-6. The maximum surface intensity increases and optimal core thickness decreases with the core index.

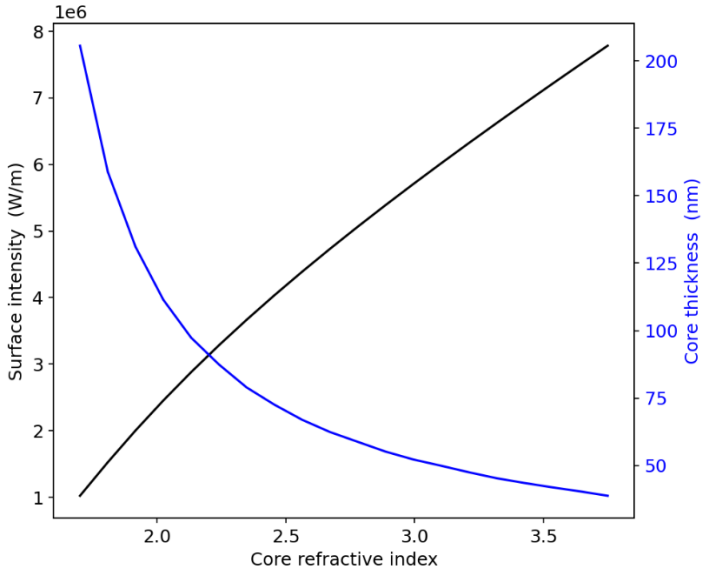

Figure 3-6. The maximum surface intensity and optimal core thickness that maximizes the surface intensify as a function of the core index.

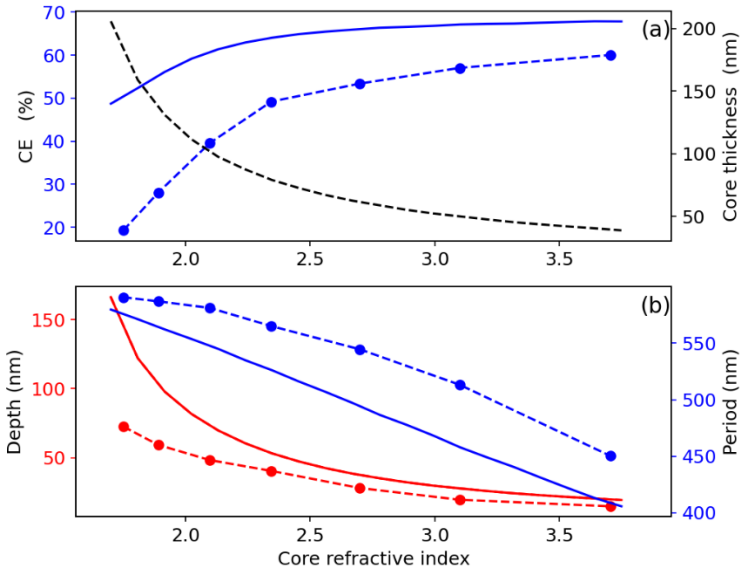

Figure 3-7. The theoretical and numerical results of coupling efficiency (CE) and the corresponding grating parameters as a function of core refractive index for the optimal thickness that maximizes the surface intensity. The beam radius is set to 25μm. Incident wavelength is 785nm(a) The theoretical CE was calculated with the optimal condition and the numerical results in the dot-dash line from[42]. (b) The optimal grating depth and period as a function of core refractive index. The solid lines are the theoretical result and dot-dash lines are the numerical results.

Here, we can assume that the numerical optimization results the optimal WGC parameters so that the optimal condition is satisfied. With the known beam width, w_0, the total

scattering coefficient α_{tot} can be calculated straightforwardly by the optimal condition defined in equation(3.8). Then the grating depth can be found with equation (3.16) by letting $\alpha_{-1}^{a} + \alpha_{-1}^{s} = \alpha_{tot}$. And the grating period can be easily obtained with equation (3.13). The CE for WGC as a function of the core refractive index is shown in Figure 3-7(a). For each core refractive index, we consider the optimal thickness, which maximizes the surface intensity as shown in Figure 3-7(a) on the right y-axis. The corresponding grating parameters, including grating depth and grating pitch size, are shown in Figure 3-7(b).

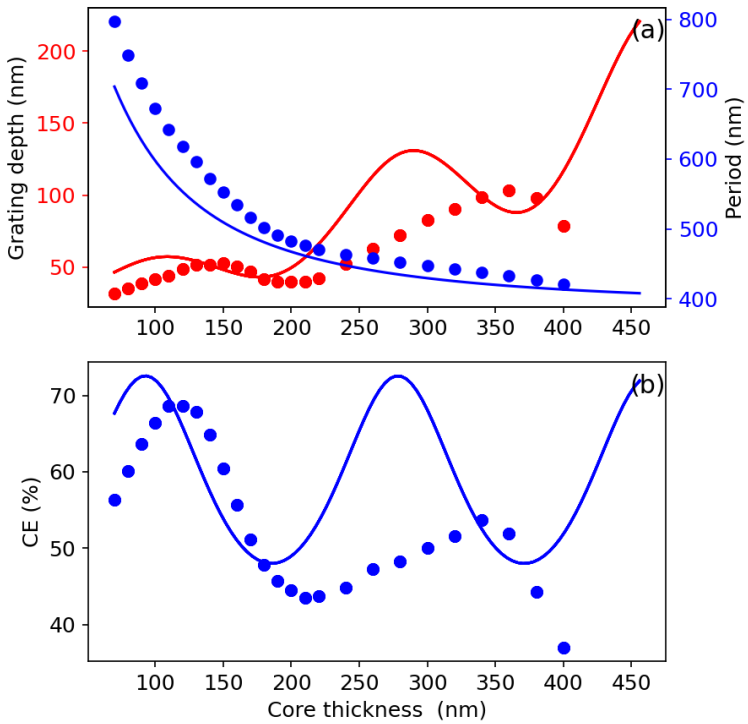

Figure 3-8. The theoretical and numerical results of CE and the corresponding grating parameters as a function of core thickness for Si-core WGC working at 1300nm. (a) The theoretical CE (solid line) is calculated with the optimal condition and the same applies to numerical results (dotted line). (b) The optimal grating depth and period as a function of core refractive index. The solid lines are the theoretical result and dotted lines are the numerical results.

As shown in Figure 3-7 (a), it is clear that the theoretical model cannot be practically applied due to poor accuracy. It is also obvious that the accuracy of the model becomes increasingly inaccurate with the reduction of the core index, although the trends for both CE and parameters of WGC predicted by the theoretical model roughly agree with the numerical results. It can be observed from Figure 3-6 and Figure 3-7 that the accuracy drops with the core index, surface intensity increases with the core index and grating depth increases with the core index. Considering the fact that Sychugov's model is derived from the theoretical analysis of corrugated waveguide, which is based on the assumption that the fluctuation of

the corrugation is shallow[130], the observed trends in efficiency can be easily understood When the surface intensity is weaker, the grating needs to be deeper to create enough scattering to fulfil the optimal condition (equation(3.8)), thus the accuracy drops because the shallow grating assumption is further violated.

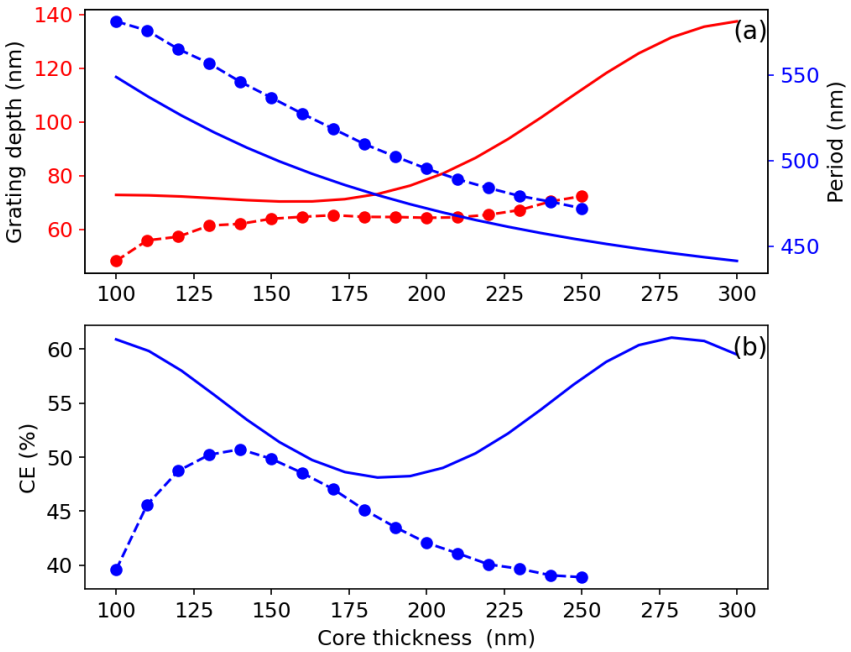

Figure 3-9. The theoretical and numerical results of CE and the corresponding grating parameters as a function of core thickness for Ta_2O_5-core WGC working at 785nm. (a) The theoretical CE is calculated with the optimal condition and the numerical results in the dot dash line. (b) The optimal grating depth and period as a function of core refractive index. The solid lines are the theoretical result and dot-dash lines are the numerical results.

To verify this reasoning, the CE and the corresponding WGC parameters are calculated for WGC as a function of core thickness with a fixed core index. Figure 3-8 shows the results for Si-core WGC with the core index of 3.505 working at the wavelength at 1.3μm. It is clearly shown that the accuracy for CE drops with the core thickness. The thicker core creates weaker core-cladding surface intensity, so deeper grating is needed for creating sufficient scattering, therefore departing from the shallow grating assumption. A similar comparison is made for Ta_2O_5-core WGC, which has a lower core index of 2.12 working at a wavelength of 785nm, as shown in Figure 3-9. The discrepancy between numerical and theoretical results is much more significant than that of the Si core WGC.

3.3 Modification to Sychugov et al.'s model

The previous study shows that the Sychugov et al.'s model demonstrate the trend of the behaviour of WGCs, although the accuracy is problematic for practical application. Observing the comparison with numerical results, besides the deteriorating accuracy of CE with grating depth, due to the shallow grating assumption, the following features can be found:

1. The simulated grating period is always bigger than the theoretical results.
2. There is a sinusoidal feature for the CE-core index curve, as shown in Figure 3-7 and Figure 3-8, which is attributed to the interference effect of the reflection from core-cladding and core-substrate interfaces. However, the numerical results always show a longer sinusoidal period.

3.3.1 Four-layer model assumption of WGC and modifications

3.3.1.1 Four-layer model assumption of WGC

Considering feature 1 together with the universal grating equation (3.13), a hypothesis can be made as follows. The value of propagating constant β_w is smaller than the value that we obtained by solving the dispersion equation for the three-layer structured waveguide. What causes the β_w become smaller might be the optical thickness in the grating area is smaller than the core thickness times the core refractive index. What causes the optical thickness in the grating area to be thinner might attribute to the grating teeth area being less optically dense than the core.

If this hypothesis is valid, feature 2 can also be explained readily—the ratio between the scattered power into cladding and substrate is given by [131]

$$\frac{\alpha_{\mathrm{s}}^{(-1)}}{\alpha_{\mathrm{a}}^{(-1)}} = \left|\frac{b_{-1}^{-}}{b_{-1}^{+}}\exp\left(iN_f^{-1}kd\right)\right|\frac{N_s^{-1}}{N_f^{-1}} \tag{3.18}$$

where $b_{-1}^{\pm}$ is the amplitude of the up and down going electric field at the core-cladding interface. Equation (3.18) clearly shows that smaller $N_f^{-1} = \sqrt{n_f^2 - \left(n_{\mathrm{eff}} - \frac{\lambda}{\mathrm{d}}\right)^2}$ produces the slower varying ratio between the scattered power into cladding and substrate. And the smaller N_f^{-1} could be a result of the less optically dense grating teeth area. Therefore, the

approximation made by almost all the literature that the propagating constant in the grating and waveguide are the same [132] is inaccurate.

To obtain the accurate propagating constant, the grating area must be treated more carefully. We assume that the horizontal periodic grating structure does not impact the propagating constant and we introduce the following two simplifications as shown in Figure 3-10. The first simplification is replacing the grating with a three-layer waveguide with an effective thickness. The effective thickness is calculated as the full thickness of the core minus the half-depth of the grating with a 50% duty cycle. The second simplification is treating the grating teeth area as a special layer with a volume-averaged permittivity, which has been widely applied for similar simplification [53,132]. In fact, the two methods have both been adopted in our modelling and have both shown better results but the four-layer simplification is slightly more accurate for calculating the propagating constant.

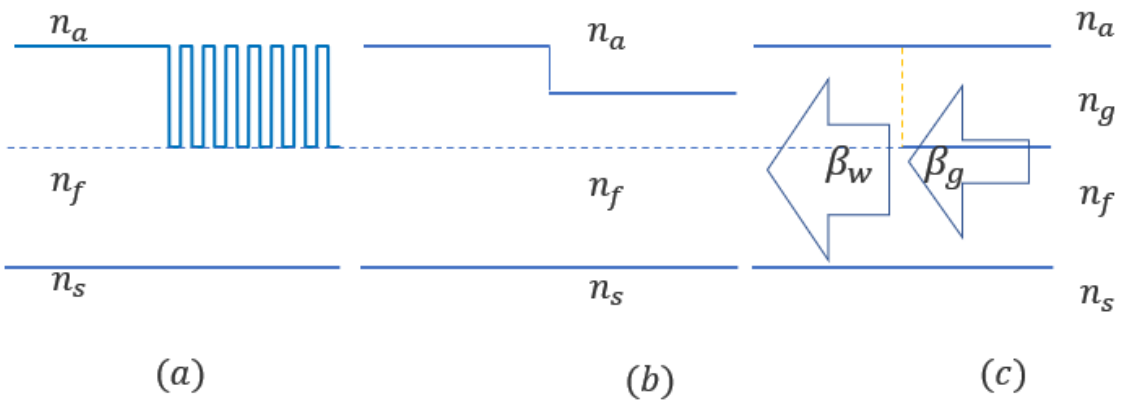

Figure 3-10. Simplification of WGC for calculating the propagating constant in the grating area (a) WGC (b) Three-layer with effective thickness structure (b) four-layer structure.

To obtain the propagating constant in the four-layer structure, a general transfer matrix method is applied [103,104], which has been applied for obtaining the propagating constant for the Ta_2O_5 waveguide with an extra SiO_2 layer in chapter 2.

For the four-layer structure, the expression for the scattering coefficient is modified accordingly,

$$\alpha_{\mathrm{a}}^{(-1)} = \left(\frac{k\sigma^*}{2}\right)^2 \frac{n_{\mathrm{f}}^2 - n_{\mathrm{eff}}^2}{n_{\mathrm{e}} h_{\mathrm{e}}} \times \frac{\left(n_{\mathrm{g}}^2 - n_{\mathrm{a}}^2\right) N_{\mathrm{a}} \left[N_{\mathrm{s}}^2 + \left(n_{\mathrm{f}}^2 - n_{\mathrm{s}}^2\right)\cos^2\psi\right]}{(N_{\mathrm{f}}^2 + N_{\mathrm{s}} N_a)^2 - (n_{\mathrm{f}}^2 - n_{\mathrm{s}}^2)(n_{\mathrm{f}}^2 - n_{\mathrm{a}}^2)\cos^2\psi} \quad (3.19\text{ a})$$

$$\alpha_{\mathrm{s}}^{(-1)} = \left(\frac{k\sigma^*}{2}\right)^2 \frac{n_{\mathrm{f}}^2 - n_{\mathrm{eff}}^2}{n_{\mathrm{e}} h_{\mathrm{e}}} \times \frac{\left(n_{\mathrm{g}}^2 - n_{\mathrm{a}}^2\right) N_{\mathrm{s}} N_{\mathrm{f}}^2}{(N_{\mathrm{f}}^2 + N_{\mathrm{s}} N_{\mathrm{a}})^2 - (n_{\mathrm{f}}^2 - n_{\mathrm{s}}^2)(n_{\mathrm{f}}^2 - n_{\mathrm{a}}^2)\cos^2\psi} \quad (3.19\text{ b})$$

in which the permittivity contrast is $\left(n_{\mathrm{g}}^2 - n_{\mathrm{a}}^2\right)$ different to $\left(n_{\mathrm{f}}^2 - n_{\mathrm{a}}^2\right)$ and the phase shift term ψ is no longer expressed as $\psi = kN_{\mathrm{f}}d$, but as $\psi = kN_{\mathrm{f}}(d-\sigma) + kN_{\mathrm{g}}(\sigma)$.

In addition to above modifications, another issue is raised. Since the propagating constant in the grating region is different, there must be a mode conversion at the joint between the grating and waveguide. Therefore, joint loss at the joint must be considered for the mode conversion. The joint loss in the WGC has not been discussed before, but similar connection loss in fibres [133–135] has been thoroughly studied and is similar to our three-to-four-layer structure mode conversion.

3.3.1.2 Joint loss assumption of WGC

An analytical expression for the conversion efficiency can be derived from the reciprocal theorem. At the coupling joint interface, the mode converted into grating can be obtained using the transverse field decomposition of the modes in the waveguide, namely

$$E_{\mathrm{g},i} = \sum_j a_j E_{\mathrm{w},j}$$
$$H_{\mathrm{g},i} = \sum_j b_j H_{\mathrm{w},j} \quad (3.20)$$

where *i* and *j* denote the i^{th} and j^{th} mode in waveguide and grating, a and b are the coupler coefficients for electric and magnetic field respectively. g and w denote the grating and waveguide.

According to the reciprocal theorem, expressed by equation (3.2), the coupling coefficient can be found as

$$a_j = \frac{\int E_{\mathrm{w},i} \times H^*_{g,i} \cdot dy}{\int E_{\mathrm{w},j} \times H^*_{\mathrm{w},j} \cdot dy}$$
$$b_j = \frac{\int E_{\mathrm{g},i} \times H^*_{\mathrm{w},j} \cdot dy}{\int E_{\mathrm{w},j} \times H^*_{\mathrm{w},j} \cdot dy} \tag{3.21}$$

Therefore, the coupling efficiency between $i^{th} - j^{th}$ mode is expressed as;

$$\eta_{i,j} = \frac{P_{\mathrm{g},i}}{P_{\mathrm{w},j}} = \frac{\frac{1}{2}\mathrm{Re}\left\{\int a_j E_{\mathrm{g},j} \times \left(b_j H_{\mathrm{g},j}\right)^* \cdot dy\right\}}{\frac{1}{2}\mathrm{Re}\left\{\int E_{\mathrm{w},i} \times H_{\mathrm{w},i}{}^* \cdot dy\right\}} \tag{3.22}$$

when $i = j$, it is the same order mode conversion with relatively good conversion efficiency because of similar propagation constants and field distributions. However, when the waveguide core becomes thicker, it is possible that a higher-order mode can be excited, although the conversion efficiency is very small according to our calculation and cannot be observed from numerical simulations. A single mode propagating in the grating region because the propagating constant is fixed by the launching angle and grating period (see equation (3.12)), while the waveguide could be multimoded as the mode from the grating region can be coupled to any mode supported by the waveguide with different mode conversion efficiency (although the efficiency for higher mode is minimal for the WGC designed for fundamental mode). So, the overall coupling efficiency is the sum $\Sigma_{\mathrm{j}}\eta_{i,j}$. For each mode in waveguide, the conversion efficiency is simplified to

$$\eta = \mathrm{Re}\left\{\frac{\int E_{\mathrm{g}} \times H^*_{\mathrm{w}} \cdot dS \int E_{\mathrm{w}} \times H^*_g \cdot dy}{\int E_{\mathrm{g}} \times H^*_g \cdot dy}\right\} \frac{1}{\mathrm{Re}\left\{\int E_{\mathrm{w}} \times H_{\mathrm{w}}{}^* \cdot dy\right\}} \tag{3.23}$$

And for TE mode, it is further simplified to

$$\eta = \frac{\left|\int E_g E_w dy\right|^2}{\int |E_g|^2 dy \int |E_w|^2 dy} \tag{3.24}$$

In addition to mode conversion, as marked in Figure 3-10 (c), there is refractive index contrast which means a reflection takes place reducing the coupling efficiency. The transmittance T can easily be calculated with the Fresnel equation, thus, the total loss at the joint interface is

$$L_j = \left(1 - \Sigma_{\mathrm{j}} \eta_{i,j}\right) T \tag{3.25}$$

3.3.1.3 Unfulfilled optimal condition

If the assumption of joint loss is valid, it is possible that the optimal condition, $\eta_{max} = 0.81$, will not be satisfied. To obtain $\eta_{max} = 0.81$, sufficient scattering strength needs to be generated, which might require a deep grating, resulting in stronger joint loss. In this case, there could be a compromise between optimal η_{max} and joint loss.

3.3.1.4 The effective depth of WGC

Besides modifying Sychugov's model with the four-layer assumption for WGC, another point that should be addressed is that Sychugov's model is derived based on the shallow grating approximation. One of the key points in this assumption is that the modal electric field is a constant E_0 over the grating region. The value E_0 is approximated with the surface modal field of the three-layer waveguide with a core thickness of $d - 0.5\sigma$, i.e. the average thickness. However, when the depth is not negligible compared to the core thickness, the electric field is no longer a constant. To mitigate the limits of the shallow grating assumption in Sychugov's approach, the effective depth is introduced in the modified model as

$$\sigma^* = \frac{2}{\pi} \frac{|\int_{-\sigma}^{0} E(y) dy|}{|E_0|} \tag{3.26}$$

Where $2/\pi$ is the Fourier coefficient for rectangular wave. With the modifications that have been made in section 3.3.1, the overall CE is calculated with

$$\eta_{-1} = \left(1 - L_j\right) \eta_{max} \frac{\alpha_{\mathrm{s}}^{(-1)}}{\alpha_{\mathrm{s}}^{(-1)} + \alpha_{\mathrm{c}}^{(-1)}} \tag{3.27}$$

Where η_{max} is no longer 0.81 but calculated with equation (3.7) and the scattering coefficients are calculated by equation (3.19 a)) with the effective depth.

3.3.2 Results of modified Sychugov et al.'s model

Figure 3-7, Figure 3-8 and Figure 3-9 are re-calculated results for CE and corresponding parameters with the modified model. Due to the optimal condition is no longer fulfilled

according to our analysis, and the grating period needs to be calculated with the propagating constant in the grating region, the optimization method is applied to search the optimal grating parameters and the optimal CE. The simple Brent algorithm [136] provided by the SciPy package [137] as a default searching method is utilized. Results calculated with the modified model are shown in Figure 3-11, Figure 3-12 and Figure 3-13. Theoretical results with and without considering joint loss have both been presented in the figures.

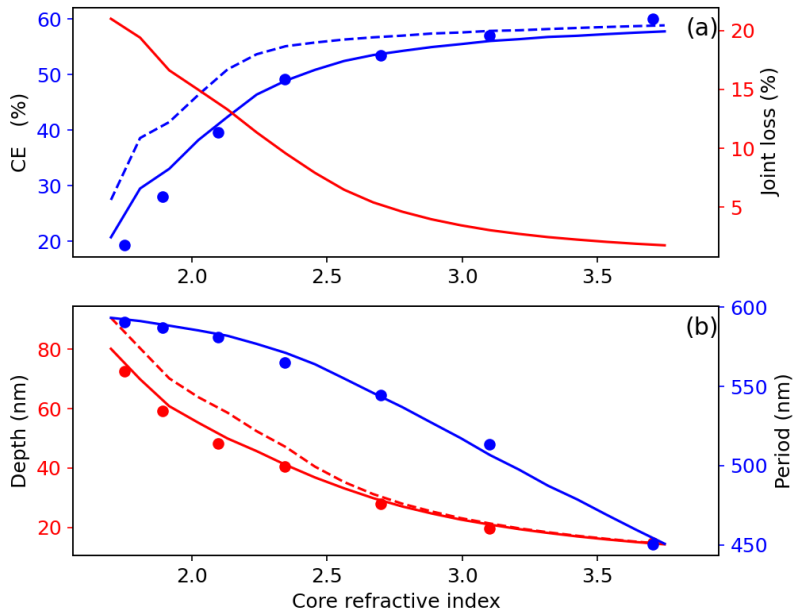

Figure 3-11. Reproduction of Figure 3-7 with the modified model.The theoretical and numerical results of CE and the corresponding grating parameters as a function of core refractive index for the optimal thickness that maximizes the surface intensity. Dash lines are the results without considering the joint loss considered; the solid line is the theoretical results and the dot for numerical results (a) The CE and theoretical joint loss. (b) The optimal grating depth and period.

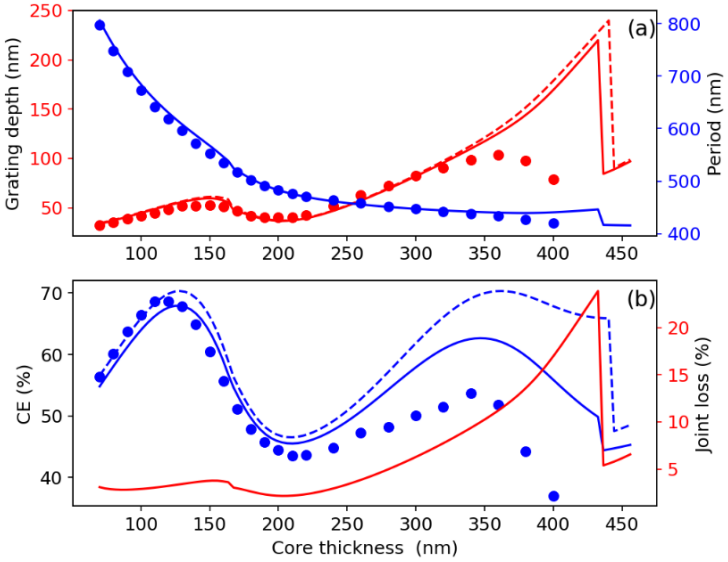

Figure 3-12. Reproduction of Figure 3-8, CE and the corresponding grating parameters as a function of core thickness for Si core WGC, with the modified model. Dash lines are the results without considering the joint loss; the solid line is the theoretical results and the dot for numerical results (a) The optimal grating depth and period. (b) The CE and theoretical joint loss.

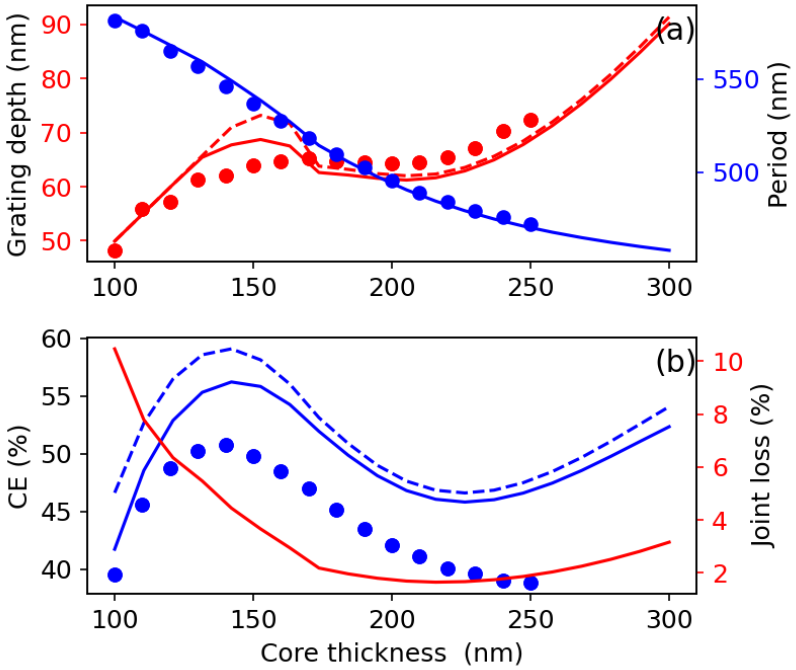

Figure 3-13. Reproduction of Figure 3-9, CE and the corresponding grating parameters as a function of core thickness for Ta_2O_5-core WGC, with the modified model. Dash lines are the results without considering the joint loss; the solid line is the theoretical results and the dot for numerical results (a) The optimal grating depth and period. (b) The CE and theoretical joint loss.

Figure 3-11 shows the CE and the corresponding grating parameters as a function of the core refractive index for thin-core WGC--the optimal thickness that maximizes the surface intensity. It is clearly showing that, compared to the original model, the discrepancy between the theoretical model becomes smaller. The theoretical results of CE and parameters agree with the numerical results relatively well, especially for the higher-index thinner core with shallower grating depth.

It is also notable that the consideration of joint loss makes a significant difference, and improves the accuracy significantly, especially for the lower-index thicker core as shown in Figure 3-11. The well-agreed grating period shows that the propagating constant in the grating region, which is different from the waveguide region, has been correctly calculated. In this sense, our four-layer model for grating should be a valid assumption.

To further validate our modified model, WGCs with thicker cores are also calculated. The results for the Si d Ta_2O_5 core with different core thicknesses are shown in Figure 3-12 and Figure 3-13. For both Ta_2O_5 core and Si core, considering the joint loss improves the agreement between theoretical and numerical results. The accuracy for both Ta_2O_5 core and Si core is relatively good when the core thickness is below 200nm. However, the accuracy is still getting worse with the increase of the core thickness. The most likely reason is the shallow grating assumption, although the effective grating depth has been applied to ease the limits of the shallow grating assumption in Sychugov et al.'s model, however, the limits of the shallow grating assumption can not be fully resolved. If the grating teeth should be theoretically treated as a special optical layer, there must be a multilayer reflection and multilayer interference effect in the cladding-grating layer-waveguide core-substrate structure. However, Sychugov et al.'s model has been exhausted to cover this fact.

Another point that needs attention is the breaking point in Figure 3-12. It seems like a failure of our model but will be discussed and clarified in the next chapter.

3.4 Conclusion

Optimizing the WGC is dependent on multiple factors. There are three parameters, including pitch size (period), thickness, and depth, for the grating, and three parameters, including beam width, angle and offset, for the input beam. Therefore, WGC optimization

with numerical method is extremely time-consuming and could produce local optima. This chapter developed a theoretical model based on the reciprocity principle of the grating coupler. The complicated in-coupling CE is converted to the easier calculation of the inverse grating-assisted out-scattering waveguide loss and the overlapping of the in/out beams.

The original Sychugov et al.'s model, an analytical theoretical model, is able to calculate the scattering coefficient from the cladding and substrate and has been first adopted into our theoretical model to calculate the CE and corresponding WGC parameters. Theoretical results have been compared to numerical results. The comparison shows that the theoretical model can only predict the trend but lacks accuracy for practical WGC design.

By analysing the comparison between theoretical and numerical results, modifications to the original Sychugov et al.'s model have been made. At first, the propagating constant in the grating region β_g for calculating the grating period is smaller than that in the waveguide region β_w. For more precisely calculating the propagating constant β_g, the grating teeth region is modelled as an optical layer with volume-averaged permittivity, which is lower than the waveguide core and therefore produces a smaller propagating constant. Secondly, the joint between the waveguide and grating creates joint loss because of the mode conversion and interface reflection. The limit of shallow grating assumption in the original Sychugov et al.'s model is alleviated with the introduced effective depth to account for the fact that the modal field in the grating teeth area is not evenly distributed. The modifications have been incorporated into the improved model for CE and WGC parameters calculation.

The modified Sychugov et al.'s model has shown an improved accuracy for both CE and WGC parameters calculation, which has been validified with numerical results. The modified model can precisely reproduce the results from the numerical model with very good agreement for WGC with a thin core, although the accuracy for a thicker core is still problematic. However, the physical mechanism confirmed with the model is valuable for investigating WGC. A more accurate and flexible model that refers to the mechanism revealed in this chapter is developed and presented in the next chapter.

Chapter 4 Modelling Waveguide grating coupler for WERS II—accurate dipole radiation model

4.1 Introduction

The potential of the Sychugov et al.'s model has been exhausted for calculating the WGC even with extensive modifications including comprehensive physics factors. However, the following mechanism verified with Sychugov et al.'s model is valid and useful for more advanced models. The grating can be theoretically treated as a four-layer 1D structure for calculating the propagating constant and field distribution. Therefore, the key point for a more accurate theoretical analytical model is to calculate the scattering coefficient from the four-layer structure.

The random scattering from the rough waveguide surface, which can be treated as a grating with a wide range of spatial frequencies, has been modelled as a thin dipole sheet and the scattering field is successfully solved with Green's function [53], showing good accuracy. Given that the discrepancy is obvious between the scattering from grating which has a scale much bigger than the sub-nanometre roughness, however, it is possible to model the grating teeth as an optical layer filled with radiating dipoles for calculating the scattered field. Following this thought, the dipole radiation model for WGC is explored in this chapter.

4.2 Accurate dipole radiation model

4.2.1 Theoretical model

4.2.1.1 Dipole radiation generated field

The starting point of the model is the same as the previous chapter—calculating the CE via its reciprocal problem out coupling. The period of the grating is adjusted so that no more than one order is scattered into the substrate and the cladding. Therefore, the CE is calculated as [125],

$$\frac{P_0^*}{P_i^*} = \left[\int_{-\infty}^{\infty} g(z)h(z)dz\right]^2 \frac{P_{-1}^S}{P_0} \tag{4.1}$$

where $g(z)$ and $h(z)$ are the profiles of the normalized field distribution of the in-coupling P_i^* and out-coupled P_{-1}^s fields, Power conservation requires that the total power P_0 is the summation of all the scattered powers, so $P_{-1}^{s(a)}/P_0 = \frac{\alpha_{-1}^{s(a)}}{\alpha_{tot}}$, where $\alpha_{tot} = \alpha_{-1}^s + \alpha_{-1}^a$[123], and $g(z) = g_0 \exp(-\alpha_{tot} z)$. $g(z)$ is the aperture function [125] normalized such that $\int_{-\infty}^{+\infty} [g(z)]^2 dz = 1$.

The scattering strength is calculated following the

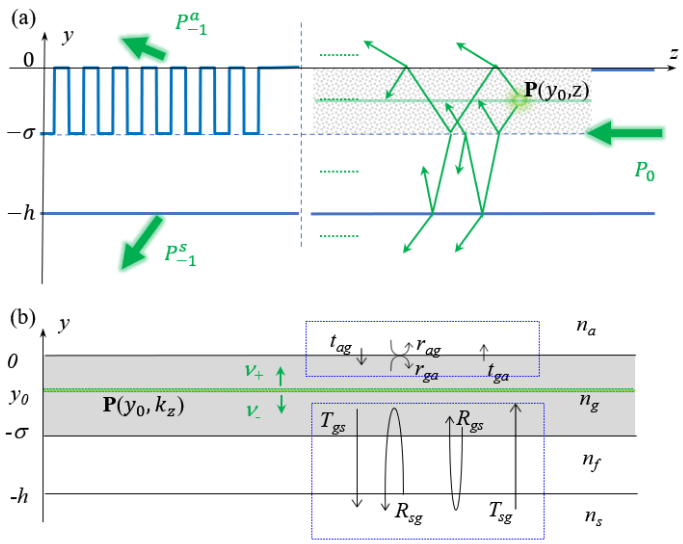

Figure 4-1. Dipole radiation model of WGC: (a)The grating is replaced by a special layer of average index n_g filled with infinitesimal scattering-dipole sheets, with radiation exiting the multilayer structure after multiple reflections. (b) Illustration of multilayer reflection/transmission with the dipole source in the grating layer, in terms of transfer matrix method.

To make it applicable to "thick" gratings, we replace the grating region with a uniform layer with a volume-averaged refractive index n_g and thickness that equals the grating depth σ, following the approach in Ref.[53], as shown in Figure 4-1. To be specific, the average index n_g for the grating layer is calculated as $n_g = \left(\int_V \frac{n^2(r)}{Vdr}\right)^{\frac{1}{2}}$ within a single period, i.e. $V \in (0 > z > -d, 0 > y > \sigma)$, following Ref.[132]. The grating layer is typically much thicker than the random roughness in Ref.[53], so the reflection in the grating layer must be considered. We next consider the fields generated by infinitesimal scattering-dipole sheets within the "average index" layer and take into account multiple reflections from the adjacent layer interfaces, utilizing the transfer matrix method[104] (see Figure 4-1(b)), which will be detailed below. We finally integrate the entire "average index" grating layer thickness to obtain the total scattered fields.

To calculate the scattering coefficients α_{-1}^{s} and α_{-1}^{a}, we consider an infinitesimal scattering-dipole sheet within the grating region (n_g) (see Figure 4-1). A similar approach has been adopted by Schmid et al. for modelling scattering loss from the rough surface of waveguides[53]. In this case, the wave equation takes the form:

$$\nabla^2 \boldsymbol{E} + n^2(y) k_0^2 \boldsymbol{E} = 4\pi k_0^2 \boldsymbol{P}, \tag{4.2}$$

where k_0 is the wavevector in free space and $n(y)$ is the refractive index in each layer shown in Figure 4-1, and $\mathbf{P}$ is the additional polarization source due to the infinitesimal scattering-dipole sheet at y_0. The modal field for the four-layer structure can be found from the homogeneous solution of equation(4.2). The simple transfer matrix method [104] is applied to obtain the propagating constant and modal field distribution in this work. Using the shift theorem[138], the polarization $\mathbf{P}$ induced by the modal electric field $\mathbf{E}(y_0, z) = \mathbf{E}(y_0)\exp(\beta z)$ at the point y_0 is Fourier transformed and takes the form[53]:

$$\boldsymbol{P}(y_0, k_z) = \frac{\left(n_g^2 - n_a^2\right)}{4\pi\sigma} \tilde{s}(k_z - \beta)\left(\hat{x}\hat{x} + \frac{n_a^2}{n_g^2}\hat{y}\hat{y} + \hat{z}\hat{z}\right)\boldsymbol{E}(y_0), \tag{4.3}$$

where $\tilde{s}(k)$ is the Fourier transform of the grating surface profile and $\hat{x}, \hat{y}, \hat{z}$ are the coordinate unit vectors. This study focuses only on the TE mode, so only the x component of the electric field is considered. Fourier transforming the square-wave grating profile, $\tilde{s}$ is expressed as $\tilde{s}(k_z - \beta) = \sigma \Sigma_{\mathrm{n}=-\infty}^{+\infty} c_n 2\pi\delta\left(k - \frac{2\pi n}{d}\right)$, where $c_n = Dsinc(nD)$ is the Fourier coefficient for the square wave with duty cycle D. Therefore, the polarization in a TE mode WGC is simplified to

$$\boldsymbol{P}(y_0, k_z) = \frac{\left(n_g^2 - n_a^2\right)}{2} \sum_{n=-\infty}^{+\infty} c_n \delta\left(k + \frac{2\pi n}{d}\right)\boldsymbol{E}_x(y_0). \tag{4.4}$$

The scattered field due to the source polarization inside an optical layer is found with the use of surface Green's functions, the introduction of vectors $e_q(y) = \left[E_{q+}e^{+ik_{yq}y},\ E_{q-}e^{-ik_{yq}y}\right]^T$, and the use of the discontinuity-source vector in terms of transfer matrix method [139]:

$$e_a(0) = M_{ag}M_g(-y_0)e_g(y_0^+)$$
$$e_g(y_0^-) = M_g(y_0+\sigma)\boldsymbol{M_{gs}}e_s(-h). \tag{4.5}$$

where the vector $e_a(0) = [\mathbf{E}_a^+, 0]^T$ and $e_s(-h) = \left[0, \mathbf{E}_s^- \mathrm{e}^{ik_{ys}h}\right]^T$, in which the $\mathbf{E}_a^+$ and $\mathbf{E}_s^-$ are the electric field in the cladding and substrate generated by the source polarization $P(y_0, k_z)$ at point y_0 (no incident fields from $\pm\infty$ are assumed). The matrices M_{xy} and M_x are the regular transfer matrices (detailed in Ref. [104,139] and section 2.2). Specifically, the subscript xy denotes transfer from layer x to y, while the single-lettered subscript indicates transfer in a specific layer, where there is only the phase shift due to propagation. $\boldsymbol{M_{gs}} = M_{gf}M_fM_{fs}$ is the composite *g-s-f* layer transfer matrix, which can also be expressed in terms of the *g-f-s* layer total reflectivity (R_{gs}, R_{sg}) and transmissivities (T_{sg}, T_{gs}) [139] (see also Figure 4-1 (b)) The field at the upper limit $e_g(y_0^+)$ and lower limit $e_g(y_0^-)$ of the source have the relation expressed as, $e_g(y_0^+) = [v_+, -v_-]^T + e_g(y_0^-)$, where $v_\pm$ is given as[53,139]:

$$v_\pm = \frac{2\pi i k_0^2}{k_{\mathrm{yg}}} \cdot \boldsymbol{P}(y_0, k_z), \tag{4.6}$$

where k_{yg} is the y component of the wavevector and given as $k_{yg}^2 = k_0^2 n_g^2 - k_z^2$. The transfer matrix method for calculating the transmitted scattered field into upper cladding and lower substrate from the grating layer is deducted below.

4.2.1.2 Transfer matrices with a polarization layer

Transmittance/reflectivity across the multilayer is calculated with the transfer matrix method, including the standard case that the source wave is incident from outside the optical layers and the case that the source is dipole radiation from inside the optical layers, and detailed as follows.

Ref. [139] has given the solution for field distribution of the dipole radiated from a three-layer system, however, here, the solution for the four-layer system that was discussed in the main text will be found from a generalized n-layer system.

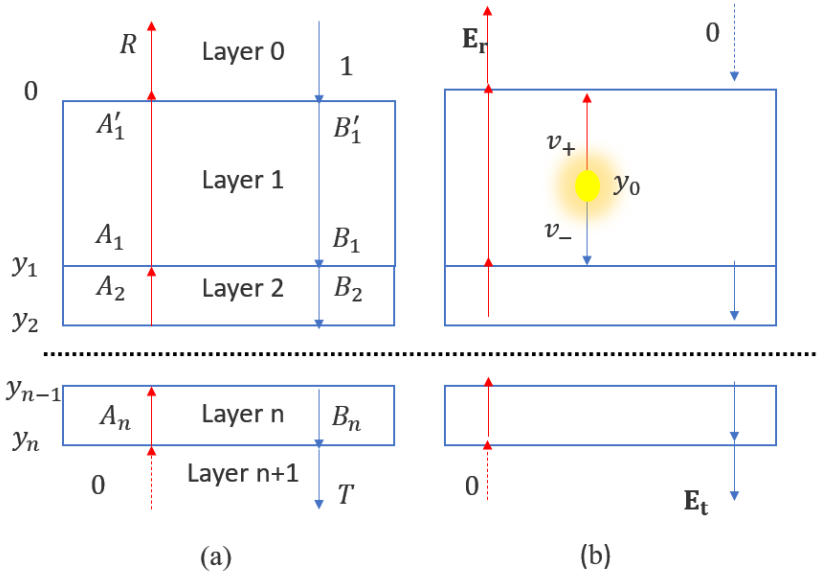

Figure 4-2. Diagram for the transfer matrix for (a) standard multilayer structure with input source from outside layers (b) multilayer structure with a special layer with a source in the first layer.

Standard transfer matrix method for calculating the multilayer structure illustrated in Figure 4-2(a) is given by Ref. [103,104], and has been discussed in subsection 4.2.1.2.

R and T, the reflectivity and transmissivity for the whole structure from layer 0 to layer n is expressed with the elements of the total transfer matrix $\widetilde{\mathbf{M}}$ (c.f. subsection 4.2.1.2),

$$T = \frac{1}{\widetilde{M}_{11}},\ R = \frac{\widetilde{M}_{01}}{\widetilde{M}_{11}} \tag{4.7}$$

The transfer matrix can also be applied for calculating the propagating constant β of the guided mode of the surface wave in the layers simply by find the singular z component of the wavevector letting [10]:

$$\widetilde{M}_{11} = 0. \tag{4.8}$$

For the structure with the input source from the dipole located within the layers, the electric field in the first and the last layer are $e_0 = [\mathbf{E_r}, 0]^T$ and $e_n = [0, \mathbf{E_t}]^T$, so the transfer matrix is,

$$e_0 = \widetilde{\boldsymbol{M}}\ e_n. \tag{4.9}$$

Assuming the dipole is located at y_0 in the grating layer, the electric field is not continuous and can be expressed as for the lower (y_0^-) and upper limits (y_0^+) around the dipole is [139],

$$e_1(y_0^+) = \begin{bmatrix} v_+ \\ -v_- \end{bmatrix} + e_1(y_0^-) \tag{4.10}$$

The dipole induced electric field through the multilayer after multiple reflection is, in terms of transfer matrix, resambling Ref. [139]

$$\begin{aligned} e_0(0) &= M_{01}M_1(-y_0)e_1(y_0^+) \\ e_1(y_0^-) &= M_1(y_0 + y_1)\boldsymbol{M_{1n}}e_n(y_n) \end{aligned} \tag{4.11}$$

Where $\mathbf{M}_{1n}$ is the transfer matrix for layer 1 to layer n, as can be inferred from the definition of transfer matrix in subsection 4.2.1.2. Then we can use equation (4.10) and (4.11) to derive the connection between the dipole generated field in the upper and lower side of the multilayer

$$\begin{bmatrix} \boldsymbol{E_r} \\ 0 \end{bmatrix} = M_{\mathbf{01}}M_{\mathbf{1}}(-y_0) \begin{bmatrix} v_+ \\ -v_- \end{bmatrix} + \widetilde{\boldsymbol{M}} \begin{bmatrix} 0 \\ \boldsymbol{E_t} \end{bmatrix} \tag{4.12}$$

Thus, the up and down going electric field can be found as

$$\begin{aligned} \boldsymbol{E_t} &= \frac{\frac{1}{t_{01}}\left(v_- e^{-i\phi_0} - r_{01}v_+ e^{i\phi_0}\right)}{\widetilde{M}_{11}} \\ \boldsymbol{E_r} &= \frac{1}{t_{01}}\left(-r_{01}v_- e^{-i\phi_0} + v_+ e^{i\phi_0}\right) + \widetilde{M}_{01}\boldsymbol{E_t} \end{aligned} \tag{4.13}$$

where $\pm i\phi_0$ is the phase shift for the up and down going wave from y_0 to lower 0-1 interface. Equation (4.13) is the general expression for the dipole at y_0 induced electric filed at both upper and lower sides of the n-layer structure. For the four-layer structure in our main text, the solution can be simply found by replacing the subscribe of 0~3 with a, g, f and s for the transfer matrices. Since $\boldsymbol{M}_{ij}\boldsymbol{M}_{ji} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the 2x2 unit diagonal matrix, so equation (4.12) can be rewritten as

$$M_g(y_0)M_{ga}\begin{bmatrix} \boldsymbol{E}_a^+ \\ 0 \end{bmatrix} = \begin{bmatrix} v_+ \\ -v_- \end{bmatrix} + \boldsymbol{M}_{gs}\begin{bmatrix} 0 \\ \boldsymbol{E}_s^- \end{bmatrix} \tag{4.14}$$

By replacing the matrix element with corresponding reflectivity/transmissivity and phase shift indicated in equation (4.4) and (4.7), an expression with clearer physical meaning is found as,

$$\mathbf{E}_\mathbf{s}^-(y_0, k_z) = \frac{T_{gs}e^{+ik_{yg}\sigma}\left(v_+ r_{ga}e^{-i\Phi_0} + v_- e^{+i\Phi_0}\right)}{1 - r_{ga}R_{gs}e^{+i2k_{yg}\sigma}} \tag{4.15}$$

$$\mathbf{E}_a^+(y_0,k_z) = \frac{t_{ga}e^{+ik_{yg}\sigma}\left(v_+e^{-ik_{yg}\sigma}e^{-i\Phi_0}+v_-R_{gs}e^{+ik_{yg}}\right.}{1-r_{ga}R_{gs}e^{+i2k_{yg}\sigma}}$$

By substituting equation (4.4) and equation (4.6) into equation (4.15), $\mathbf{E}_{\boldsymbol{a}}^+$ and $\mathbf{E}_{\boldsymbol{s}}^-$ can be written as function of modal field $\mathbf{E}_{\mathbf{x}}(y_0)$ in the wavevector domain. The scattered field due to source polarization at y_0 can be found by Fourier transforming $\mathbf{E}_{\boldsymbol{a}}^+(y_0,k_z)$ and $\mathbf{E}_{\boldsymbol{s}}^-(y_0,k_z)$ back into the real space. Similar approach has been followed by Payne and Lacey when modelling the random scattering from waveguides with rough surfaces[96].

The total scattered electric fields in the substrate (s) and cladding (a) areas are obtained by integrating the contributions of the infinitesimal source polarization sheet over the entire grating layer, and expressed as:

$$\begin{aligned} E_{ns}^- &= \frac{c_n k_0^2\left(n_g^2-n_a^2\right)e^{+ik_{yg}\sigma}}{2k_{yg}\left(1-r_{ga}R_{gs}e^{+i2k_{yg}\sigma}\right)}\left[\int_{-\sigma}^{0} f_s(y_0)\boldsymbol{E_x}(y_0)dy_0\right] \\ E_{na}^+ &= \frac{c_n k_0^2\left(n_g^2-n_a^2\right)e^{+ik_{yg}\sigma}}{2k_{yg}\left(1-r_{ga}R_{gs}e^{+i2k_{yg}\sigma}\right)}\left[\int_{-\sigma}^{0} f_a(y_0)\boldsymbol{E_x}(y_0)dy_0\right], \end{aligned} \tag{4.16}$$

where $f_{\mathrm{s}}(y_0)=\left(r_{ga}e^{+i\Phi_0}+e^{i\Phi_0}\right)$, $f_{\mathrm{a}}(y_0)=\left(e^{-ik_{yg}\sigma}e^{-i\Phi_0}+R_{gs}e^{+ik_{yg}\sigma}e^{+i\Phi_0}\right)$. It should be stressed that the terms $f_{\mathrm{s}}(y_0)$, $f_{\mathrm{a}}(y_0)$ and $F(\sigma)=\left[1-r_{gs}R_{gs}\exp\left(i2k_{yg}\sigma\right)\right]$ express the effects of multiple reflections from the interfaces adjacent to the thick grating layer. The impact of multiple reflections on scattering has also been discussed in Ref.[123] following a different approach applicable to "shallow" gratings without treating the grating as a specific layer.

The scattered powers into the substrate and cladding can now be calculated from the corresponding scattered fields given by equation (4.16). The guided power in the grating region is calculated by $P_0=\frac{\beta}{\omega\mu\int_\infty|E_x|^2dy}$. Note that β and E_x refer to the propagation constant and the modal field distribution in the four-layer (n_a-n_g-n_f-n_s) grating region. Calculation of the scattering coefficients and consequently of the aperture function g(z) results in the calculation of the grating coupling efficiency $\eta_g=P_0^*/P_i^*$, through equation (4.1). Considering the joint loss that has been shown in section 3.3.1.2, the overall efficiency

is $\eta_g^t = \eta_g(1 - L_j) = (P_0^*/P_i^*)(1 - L_j)$, where L_j is the joint loss defined in equation(3.25).

The other grating parameter, grating period, has been discussed in section 3.1.2 and simply calculated with (3.11).

4.2.2 Results and discussion

To validate the theoretical model, we compared it with our previously published numerical results based on vectorial 2D-Finite-Difference Time-Domain simulations using Lumerical FDTD Solutions [42]. The WGC is configured as follows: the grating is etched on the waveguide core with SiO_2 substrate and air cladding. The grating profile is regular uniform square-wave-shaped with a 50% duty cycle and 50μm length. The incident light with 785nm wavelength is launched from the substrate with the angle in the substrate $\theta_{sub} = 5.5^o$. The incident light has a Gaussian profile with a beam width of 25μm. The etching depth and period of the grating are optimized for maximum CE. In the numerical model, the CE is calculated by solving the Maxwell equation for the whole structure when searching the optimal grating etch depth (e) and period (p). Particle swarm optimization, combined with a nested sweep of etch depth (e) and period (p), is used for this purpose. The numerical optimization is conducted on a supercomputer and takes hours to produce the results shown in Figure 4-3, Figure 4-4 and Figure 4-5. For the theoretical model, the running time is a few tens of seconds on a personal computer to produce much more data points, using the Brent algorithm [136] provided by the SciPy [137] package as a default searching method.

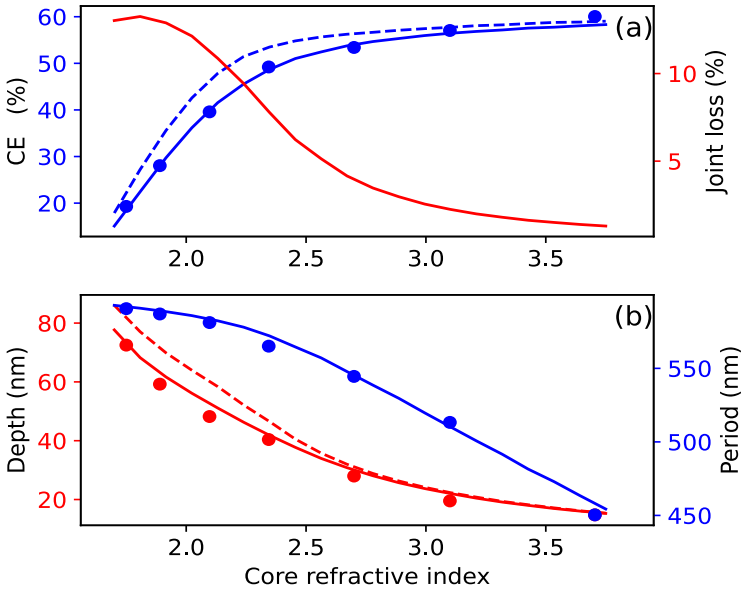

Figure 4-3.: CE and WGC parameters for different core indices, with core thickness selected to maximize the surface intensity; (a) CE on the left y-axis in blue and joint loss on the right y-axis (b) WGC etch depth (left axis) and grating period (right axis). Dots: numerical from Ref.[42]), solid line: theoretical, dashed line: theoretical without joint loss.

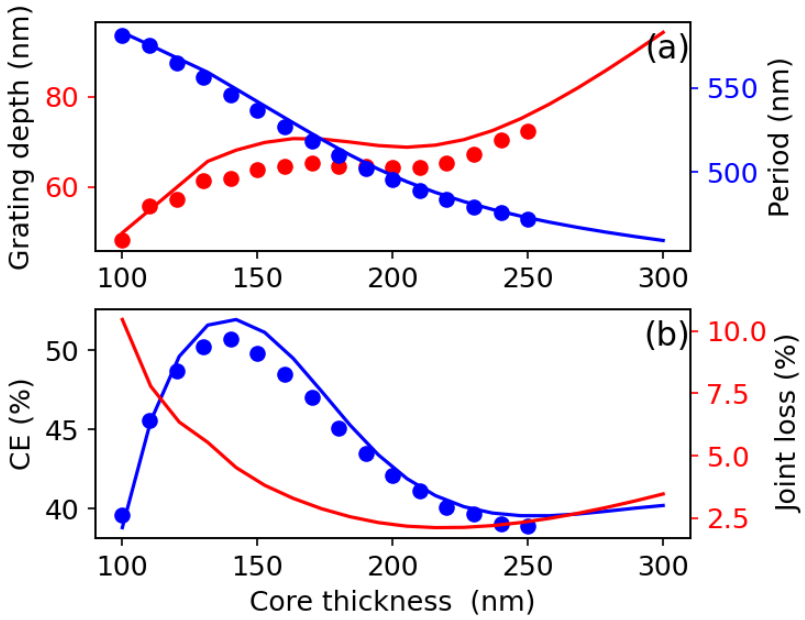

Figure 4-4. Total CE and WGC parameters for different Si core thicknesses; (a) WGC etch depth (left axis-red) and grating period (right axis-blue), (b) CE (left axis–blue) and joint loss (right axis-red). Solid lines: theoretical result; dots: numerical results from Ref.[42]

To further validate the theoretical model, we have compared the total CE and WGC parameter variation as a function of core thickness and compared them again to numerical data from Ref.[42]. The grating configuration and beam profile remain unchanged, except for the core material being Si at the wavelength of 1300nm. The results are summarized in

Figure 4-3 (a) plots the WGC etch depth (left axis-red) and grating period (right axis-blue), and Figure 4-3 (b) plots the total CE (left axis – solid blue line) and joint loss (right axis-red line). Also shown are the numerical results from Ref.[42] (dots).

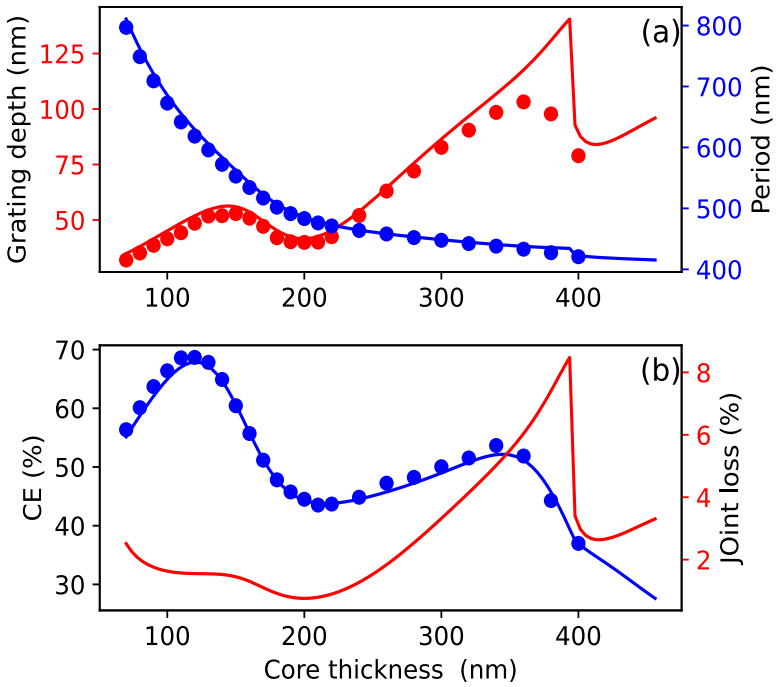

Figure 4-5. Total CE and WGC parameters for different Si core thicknesses; (a) WGC etch depth (left axis-red) and grating period (right axis-blue), (b) CE (left axis–blue) and joint loss (right axis-red). Solid lines: theoretical result; dots: numerical results from Ref.[42]

In Figure 4-3, the requirement for maximum surface intensity results in relatively thin waveguides and a monotonic increase of total CE along with a monotonic decrease of grating depth with the core refractive index. In sharp contrast, Figure 4-5 shows that both the total CE and the grating depth vary non-monotonically as the core thickness increases. This is a direct result of the multiple-scattering interference effects discussed in section 4.2.1.2. In addition, unlike Figure 4-3, there is a significant discrepancy of the grating depth between the numerical and theoretical results for a core thickness around ~390nm, although the total CEs are similar.

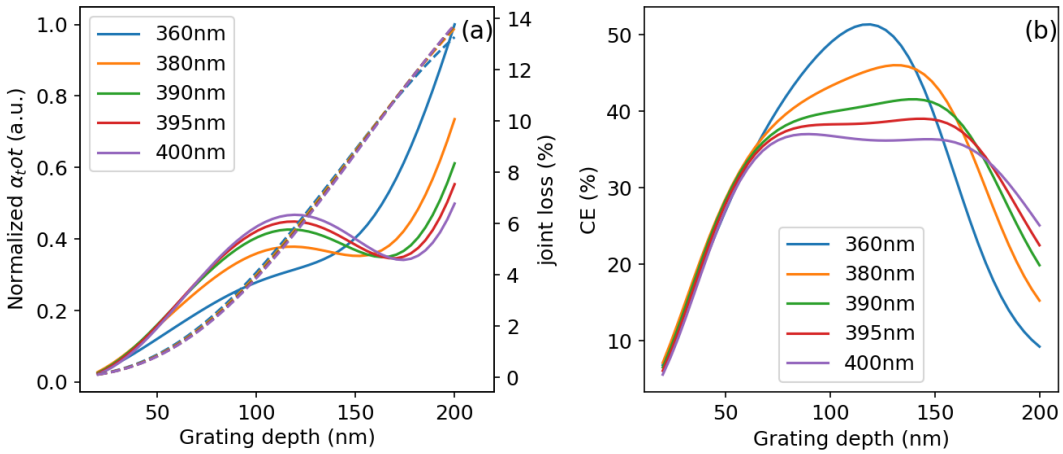

Figure 4-6. (a) Normalized attenuation coefficient (left axis - solid lines), and joint loss (right axis - dashed lines) as a function of etch depth; (b) total CE as a function of grating depth, for different core thicknesses.

To clarify this discrepancy, a more detailed study on grating depth for core thickness around 390nm is shown in Figure 4-6 (a) shows that the scattering strength does not increase monotonically with the etch depth, due to multilayer interference effects. At the same time, joint loss is exponentially growing. These two factors combine to “flatten” the total CE dependence on grating depth and make it multi-peaked around the core thickness of 390nm. This renders the WGC optimization a non-convex problem, making the numerical calculation of the exact optimum grating depth very difficult. However, the “flatness” of the total CE dependence on grating depth around this core thickness results in minute differences in the returned total CE.

After the validation of the theoretical model with three sets of numerical results, the good agreements have been reached, verifying the accuracy of the theoretical model. The hypothesis such as the joint loss and reflection/interference within multiple optical layers including the modelled grating teeth layer are valid.

4.3 Limits and breakthrough of multilayer interference effect in WGC

As can be observed from Figure 4-3 to Figure 4-6, the CEs vs core thickness are showing a sinusoidal shape, which is evidently a result of the multilayer interference effect. In other words, the CE is limited by the multilayer interference effect once the optimal condition (discussed in Section 3.1.1) is fulfilled. As a result, the CE can be over 70%, reaching the theoretical upper limit 80%, for some specific core thicknesses. Meanwhile, optimal CE is

far from the upper limits for some core thickness. Therefore, the limitation of multilayer interference effect on CE is explored, and one of the possible methods to break such limits is presented. Si on SiO_2 is currently the most popular photonics platform, so the following study is conducted on a Si core waveguide on SiO_2 substrate working at 1300nm.

4.3.1 Theoretical and numerical model

To begin with, the input angle, which has been fixed to 8 degrees (in the air) for all previous studies, is investigated. As it is known, the multilayer interference effect can be impacted by the input angle because the reflectivity and transmissivity are functions of the incident angle at optical interfaces (Fresnel equation) and the input angle also changes the optical path in each optical layer (phase shift due to propagation).

To study the impact of input angle the shallow grating (10nm etch depth) is assumed for different core thicknesses to eliminate the joint loss, although shallow grating needs impractical grating length to create enough scattering. CE and corresponding grating length as function input angle (in the substrate) and core thickness are shown in Figure 4-7. In Figure 4-7 (a), it's evident that the highest coupling efficiency (CE) is achieved when the input angle is significantly large, almost parallel to the waveguide. However, it's important to note that it is the propagation angle in the substrate, which means that a prism is required to achieve this angle in the substrate when launching the laser from beneath the substrate. Another obvious point clearly shown is the interference effect between the core interfaces defined by the core thickness. Figure 4-7 (a) also shows a shadowed corner on the contour map. The reason for the CE drops for the large angle input for thin waveguide core is higher scattering order will be generated once the -1^{st} order is scattered with a large forward angle while the propagating constant is small, which can be inferred from section 3.1.2. Figure 4-7 (a) indicates that a large backward input is optimal, however, practically, the laser is launched outside the substrate. Therefore, the critical refraction angle and refraction from the substrate back to free space must be considered, just as shown in Figure 4-7(b). When considering the substrate reflection, the small input angle is more favourable, justifying our choice of input angle in our previous publication [42].

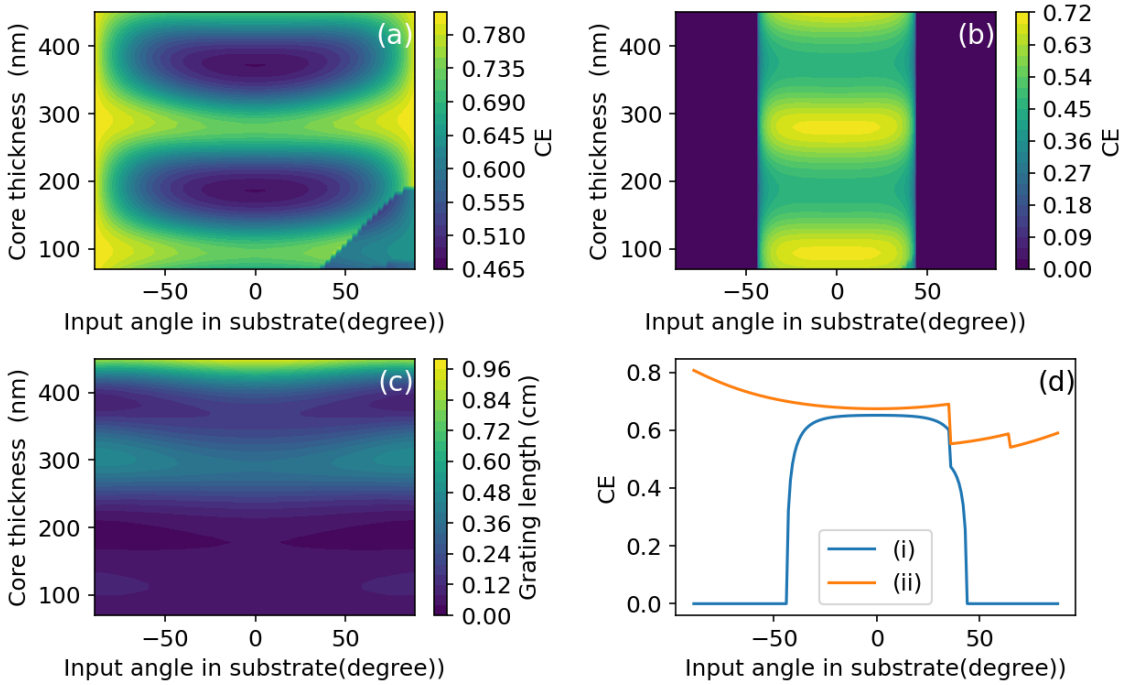

Figure 4-7. Impact of input (in substrate) angle for different core thickness. (a) CE as a function of input angle and core thickness without considering the reflection from substrate back to air when launching the laser beneath the SiO_2 substrate (b) CE as a function of input angle and core thickness with considering the reflection from substrate back to air (c) corresponding optimal grating length as a function of input angle and core thickness for shallow grating teeth (10nm) (d) CE as a function of input angle for 70nm core thickness (i) substrate reflection considered (ii) substrate reflection not considered

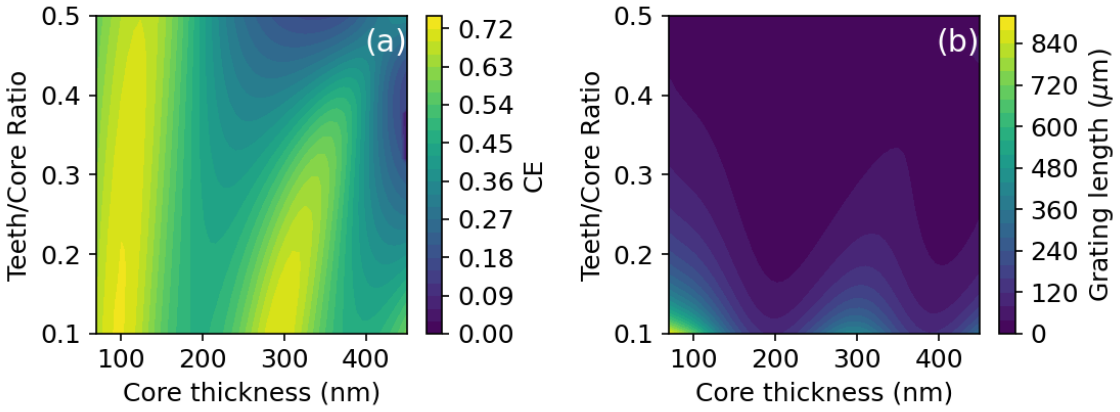

Figure 4-8. (a) CE with corresponding (b) optimal grating length as a function of grating depth and core thickness for the grating

Figure 4-7(c) shows the corresponding grating length calculated in accordance with the optimal condition equation (3.8) for the shallow grating teeth. It is clearly showing that the shallow grating is impractical due to the long grating length. Therefore, the input angle is

fixed to 8 degrees (in air) again to study the CE as a function of grating depth and core thickness.

CE and grating length as a function of grating depth and core thickness are shown in Figure 4-8 (a) and (b), respectively. The grating depth is shown in terms of the ratio between grating teeth and total core thickness for clarity. The contour map gives a full picture of how parameters, especially the grating depth of WGC, impact the CE, from which some mechanism can be inferred and refer to more general WGC design. Firstly, the deeper grating creates stronger scattering so the corresponding grating length can be shorter, while, in general, the CE decreases with grating depth due to increased joint loss. Secondly, there are two ridges on the CE contour map, which is, evidently, a result of multilayer interference. The ridges, especially the ridge for the thicker core, are bending towards the thicker core with the increase of the grating depth. This bending behaviour can be easily explained as follows. For a specific WGC, the optical path through the grating layer (n_g) and core layer (n_c) decrease with the increment of the grating layer due to its smaller refractive index, although the total thickness of the grating layer and core layer stay the same. Thirdly, the bending of the ridges makes the CE-grating depth function multi-peaked for some certain core thickness, e.g., 360 nm. This is again proof of the non-convex nature of CE optimization of WGC, and a supplementary explanation for the overshot in Figure 4-5 (more clearly shown in Figure 4-9). Last but not least, some specific core thickness suffers from low CE, i.e., core thickness around 200nm and 400nm due to the limits of interference effect.

Regarding the low CE limits of specific core thickness, it is possible to manipulate the interference effect. The interference could be impacted by intentionally manipulating the phase of the scattered light in the grating teeth region. Recalling the theoretical model in section 4.2.1., the grating teeth are modelled as a special layer filled with polarized dipole source so the phase of the scattered light can be changed via changing the horizontal position of the dipole sheet to add an extra phase shift—tilting the grating teeth. As aforementioned, the deeper grating offers more space for manipulation (considering the ridge in Figure 4-8 (a) around core thickness is 400nm).

For investigating the tilted grating, the grating length is fixed at 50μm, which is a practical value for fabrication and application, following our previous setup [42]. CE as a function of

core thickness and grating depth with the fixed length is shown in Figure 4-9, including the numerically optimized grating depth for each core thickness. Expectedly, the thicker core suffers from even lower CE with limited grating length. With the increase of core thickness, the modal field at the grating teeth region decreases, which weakens the scattering strength. The scattering can be increased with grating depth, while the dilemma is the consequent increased joint loss.

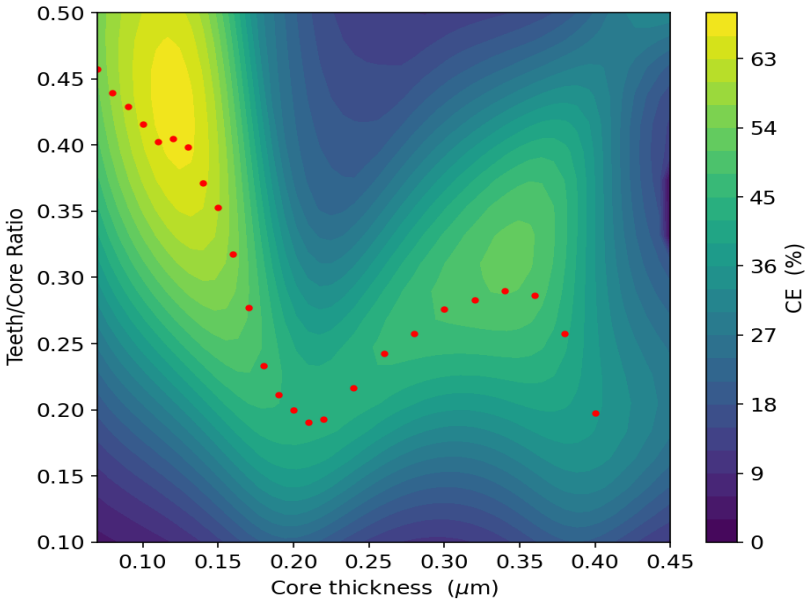

Figure 4-9. CE as a function of grating depth and core thickness for grating length fixed at $50\mu m$. Red dots are the results from numerical optimization.

Since the thicker core suffers low CE, while allowing deeper etching that enables more space for manipulating the phase shift, i.e., by tilting the same angle, more phase shift can be obtained with deeper grating. A structure of numerical model for tilted grating built with Lumerical FDTD solution is shown in Figure 4-10.

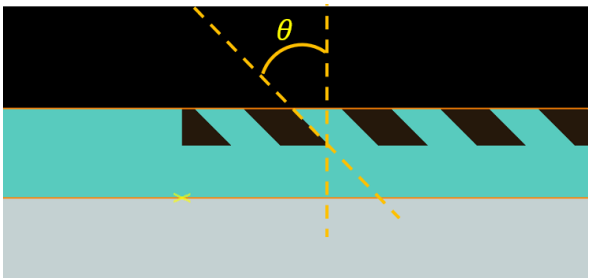

Figure 4-10. Diagram of tilted grating. The tilted angle θ denotes the angle between the sidewall of the grating teeth and the norm.

Following the theoretical model built in section 4.2. The grating in teeth region is still modelled as a dipole-filled optical layer. The infinitesimal dipole sheet at y_0, due to the tilting, has a horizontal position shift $\Delta z = y_0 \sin\theta$, equivalent to a phase shift in Fourier transformed wavevector domain, as shown in Figure 4-11. The shift in the space domain and in the wavevector domain is linked by the shift theorem [138], given as

$$s(z-\Delta z)e^{\beta z} \leftrightarrow \tilde{s}(k-\beta)e^{ik_d \Delta z} \tag{4.17}$$

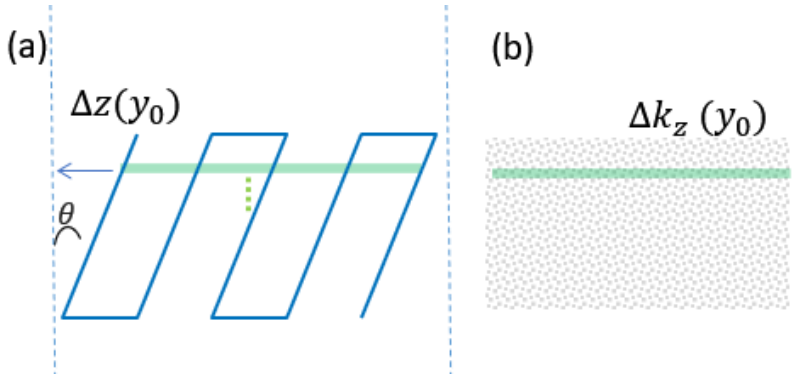

Figure 4-11. Illustration of the shift of infinitesimal dipole sheet due to tilting (a) in space domain and (b) in wavevector domain.

where $k_d = \frac{2\pi}{d}$ is the spatial frequency of the grating, and $\tilde{s}(k_z-\beta) = \sigma \Sigma_{n=-\infty}^{+\infty} c_n 2\pi\delta\left(k - \frac{2\pi n}{d}\right)$, where $c_n = Dsinc(nD)$. Therefore, the expression for dipole radiation at position y_0 in tilted grating, can be expressed by substituting (4.17) into the expression for dipole polarization (4.3),

$$\mathbf{P}(y_0, k_z) = c_n(y_0)\frac{\left(n_g^2 - n_a^2\right)}{2}\sum_{n=-\infty}^{+\infty} c_n \delta(k + nk_d))\boldsymbol{E}_x(y_0). \tag{4.18}$$

Fourier coefficient $c_n(y_0)$ is no longer a constant but a function of its vertical position due to additional shift is given by

$$c_n(y_0) = c_n(0)\exp(k_d y_0 \sin\theta) \tag{4.19}$$

where $c_n(0) = Dsinc(nD)$.by substitute the polarization **P** that considers the phase shift into (4.6) and then (4.15), the final scattered field is found as

$$E_{ns}^{-} = \frac{c_n(0)k_0^2\left(n_g^2 - n_a^2\right)e^{+ik_{yg}\sigma}}{2k_{yg}\left(1 - r_{ga}R_{gs}e^{+i2k_{yg}\sigma}\right)}\left[\int_{-\sigma}^{0} f_s(y_0)\boldsymbol{E_x}(y_0)e^{k_d y_0 \sin\theta}dy_0\right]$$
$$E_{na}^{+} = \frac{c_n(0)k_0^2\left(n_g^2 - n_a^2\right)e^{+ik_{yg}\sigma}}{2k_{yg}\left(1 - r_{ga}R_{gs}e^{+i2k_{yg}\sigma}\right)}\left[\int_{-\sigma}^{0} f_a(y_0)\boldsymbol{E_x}(y_0)e^{k_d y_0 \sin\theta}dy_0\right] \quad (4.20)$$

where $f_s(y_0) = \left(r_{ga}e^{+i\Phi_0} + e^{i\Phi_0}\right)$, $f_a(y_0) = \left(e^{-ik_{yg}\sigma}e^{-i\Phi_0} + R_{gs}e^{+ik_{yg}\sigma}e^{+i\Phi_0}\right)$, same to the definition in 4.2.1.3, expressing the effects of multiple reflections from the interfaces adjacent to the thick grating layer. After presenting the analytical expression for calculating the scattered field for tilted grating, CE of WGC with core thickness of 350nm and 400nm is calculated with the procedure clarified in section 4.2.1.3. The core thickness of 350nm and 400nm are selected for investigation is that these thicknesses suffer from low CE and meanwhile allow deeper grating for sufficient phase shift by tilting the grating teeth.

4.3.2 Theoretical results and numerical validation

The CEs as a function of tilted angle and grating teeth for 350nm and 400nm are firstly calculated with analytical expression and shown in Figure 4-12 and Figure 4-13, respectively.

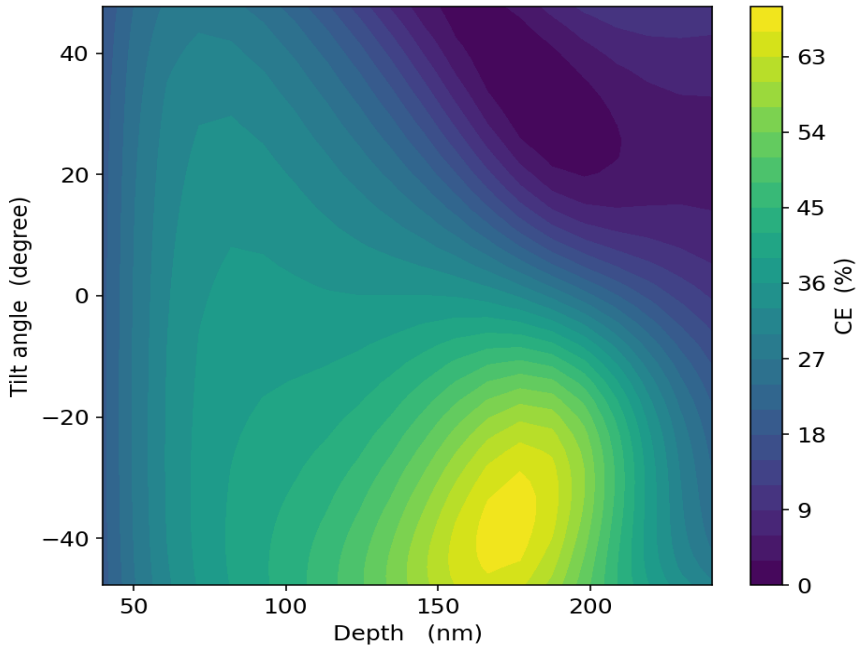

Figure 4-12. CE as a function of tilted angle of grating teeth and depth of grating for core thickness is 400 nm.

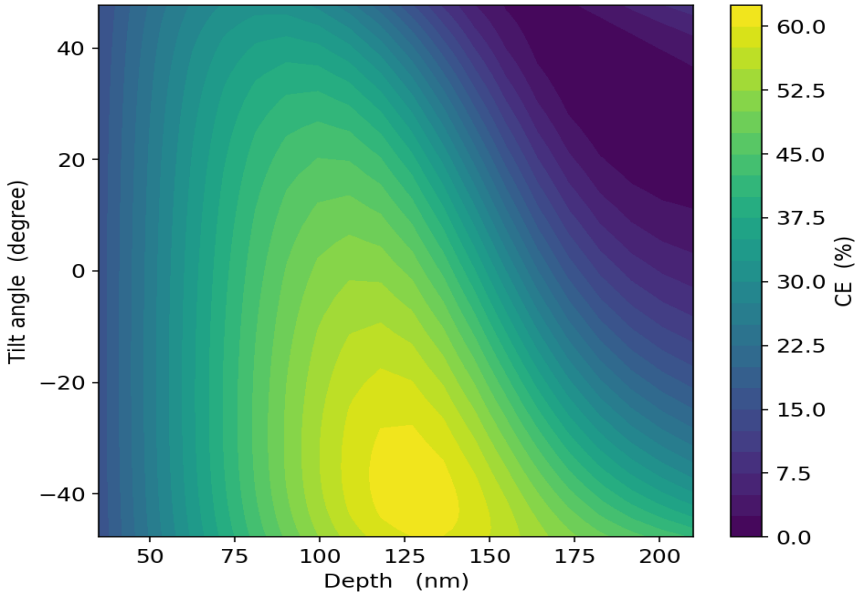

Figure 4-13. CE as a function of the tilted angle of grating teeth and depth of grating for core thickness is 350 nm.

For both core thicknesses of 400nm and 350nm, as shown, tilting the grating teeth ranging from -45^o to 45^o can significantly impact the CE. For regular rectangular square-like grating teeth, the optimal CEs for WGC with core thicknesses of 350nm and 400nm are around 50% and 40%, which is clearly shown in Figure 4-5. However, tilting the grating teeth to -40^o could increase the CEs for the core thickness of 350nm and 400nm to 60% and 67%, respectively.

Notably, the optimal grating depth for tilted grating is much deeper than the regular grating, e.g., the optimal depth for regular WGC with 400nm core is only around 80nm while for the optimal tilted grating is 175nm. Such a big discrimination is a result of multifactor impacts by the grating depth. As discussed, grating depth impact the CE via the following mechanisms. At first, obviously, deeper grating creates stronger scattering. Secondly, the interference effect, grating that is modelled as an optical layer lessens the total optical path in the whole WGC structure for the scattered light because it is less optically dense. Thirdly, deeper grating enables more phase shifts with the same tilted angle. Lastly, the deeper grating creates more joint loss which is unfavourable but ineligibly unavoidable when grating does deep.

Contour maps shown in Figure 4-12 and Figure 4-13 are unlikely to be able to be reproduced by numerical methods due to the huge computation workload even for a

supercomputer. As mentioned in section 4.2, two factors (grating pitch and depth) optimization could take hours on a supercomputer. Therefore, the numerical validation in this section is only conducted for the optimal grating depth for each core thickness, presented as CEs as a function of tilted angle. The numerical validation for the core thickness of 400nm and 350nm are shown in Figure 4-14.

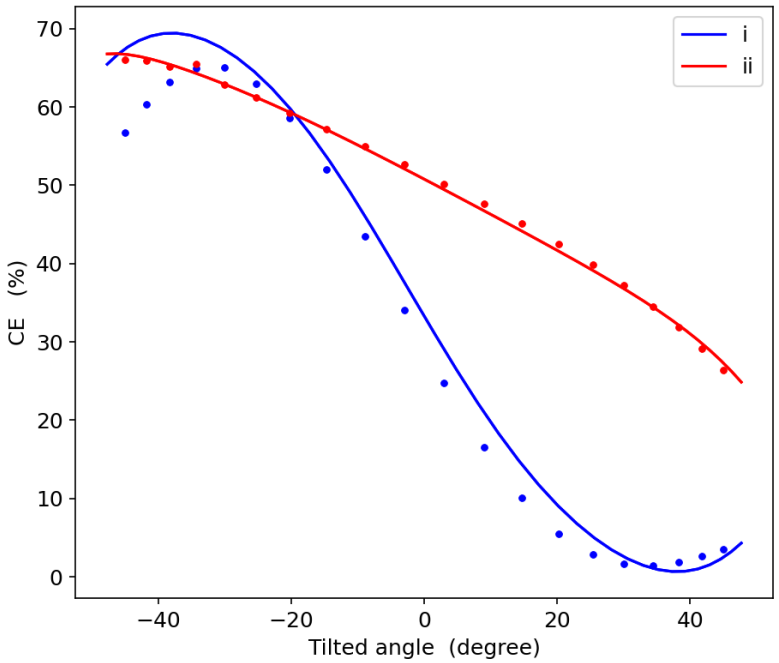

Figure 4-14. CE as a function of tilted angle for optimal grating depth for WGC with a core thickness of (i) 400nm and (ii) 350nm. The optimal grating depth are 175nm and 123nm for (i) and (ii), respectively. Solid lines are theoretical results while dots are numerical results.

Figure 4-14 indicates that our simple theoretical model on the titled grating is valid. It confirms our assertion that when the grating length and grating depth are compatible to fulfil the optimal condition (detailed in Section 3.1.1), the multilayer interference effect limits CE. Tilting the grating is a simple method to manipulate the multilayer interference effect by intentionally changing the phase of scattered light. Regular WGC with a thick core, in our case, suffers from low CE, while higher CE could be achieved via tilting the grating teeth as a thicker core allows deeper grating to enable more phase shift.

Figure 4-15 is showing the normalized electric field for the grating with different tilting angles. It is clear that tilting the grating teeth results in differentiating the power portion of transmitted/reflected and coupled.

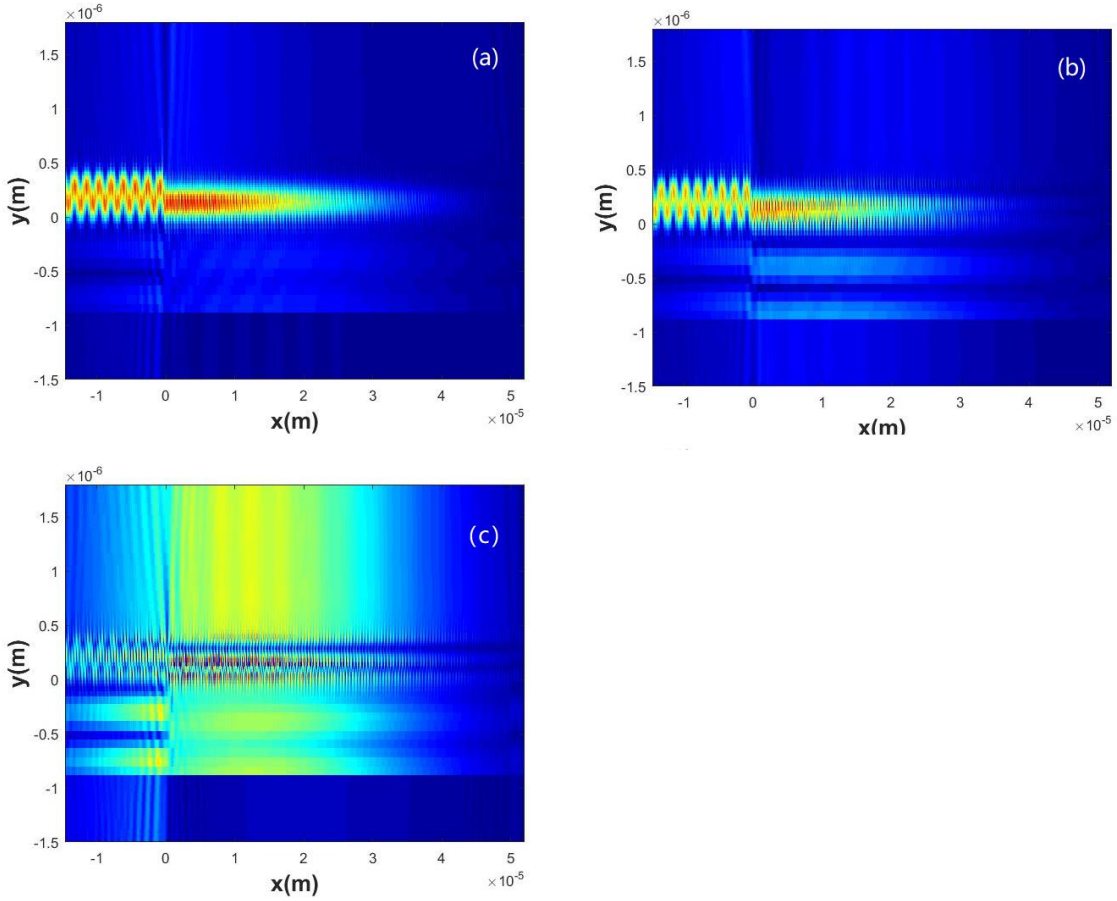

Figure 4-15. Normalized distribution of electrical field for WGC of 400nm core thickness and tilted grating teeth with different angles (a) -40° (b)0° (c) 40° (unlike theoretical model, the y=0 is at the core-substrate interface in the numerical model)

Another method, including adding multiple interfaces to the substrate for increasing CE via manipulating the multilayer interference, could also explained and modelled by our theoretical model but will remain as a future work. Meanwhile, there is a significant discrepancy in Figure 4-14 for the 400nm core when its tilted angle is small. The reason for this discrepancy has not been fully understood but will be discussed in the next section.

4.4 Effect of the grating profile

The previous study on WGC is for 50% duty cycle rectangular grating, although it is the common grating profile for WGC design. However, in practical grating fabrication, such a shape can be difficult to be achieved, especially when the grating period is small. Moreover, the grating profile does not have to be rectangular. Other than the standard rectangular grating profile, a saw-shape[140] shape and even more complicated stair-case or double-etched shaped grating profile has been presented [141,142], showing an increment of CE

while without clarifying the mechanism. A systematic study on how the grating profile impacts the CE has not been conducted so far. Therefore, it is valuable to explore how the grating profile impacts the CE with our theoretical model.

4.4.1 Theoretical and numerical model

Trapezoid-shaped grating profile with variable upper and lower edges, as shown in Figure 4-16, is adopted for investigating the impact of the grating profile on CE. The upper and lower edge of the trapezoid is r and 1-r of the period of the grating, and tilted with an angle of θ, as specified in the figure caption. The profile can therefore be defined by the ratio between the upper edge and period, r, i.e., when $r = 0$, the profile is triangular while $r = 0.5$ is rectangular.

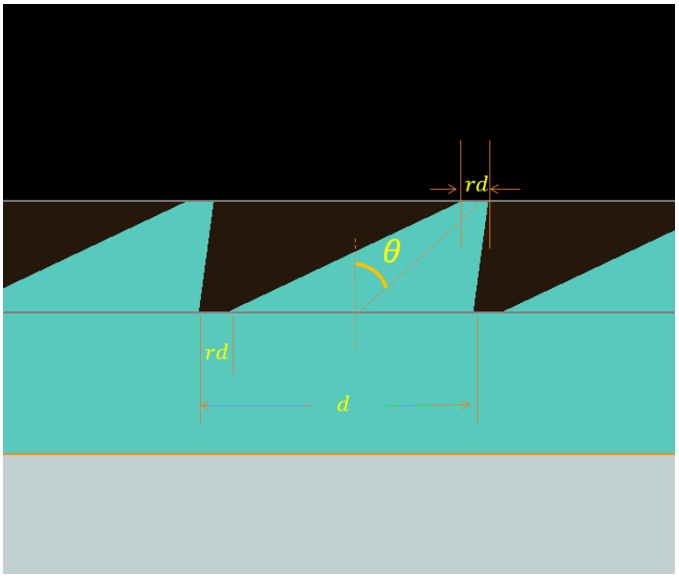

Figure 4-16. Grating profile in the numerical model. The length of the upper and lower edge is rd and $(1-r)d$, respectively, while the d is grating pitch. The tilting angle θ is the angle between central line and norm of the trapezoid.

For the grating profile defined above, a significant discrepancy from the previous rectangular grating is that, if the grating is still to be modelled as a dipole-filled source layer, the effective refractive index of the "dipole layer" is no longer uniform because, at each vertical position y_0 in the grating layer, the portion of high index (n_c) and low index (n_a) is different. Therefore, the grating layer should be treated as an optical layer with a gradient refractive index and not a constant index. Direct evidence is simply shown as follows. If the grating shown in Figure 4-16 is still to be treated as a constant-index layer, the value of the index should be the same to the rectangular grating studied in the section 4.2 and 4.3. The grating period for the grating structure shown in Figure 4-16 should stay the same for the

1300nm coupling with an input angle of 8^o. However, the numerical simulation shows that the central wavelength can shift up to 2% for $r = 0.5$ (rectangular grating) to $r = 1$ (triangular grating), which will be detailed in the next section. The reason for the central wavelength shift can be understood as follows. The upper edge ratio r impacts the profile of grating, so, equivalently, the refractive index of the modelled grating layer is changed. The propagating constant for 1300nm wavelength in the four-layer structure is then impacted by the changing of the index in the grating layer. However, according to the grating function for WGC (3.12), the fixed input angle and grating period can only couple light with a specific propagating constant, so the wavelength can be coupled into WGC is then shifted. Based on the analysis and numerical simulation, the WGC is modelled again as a four-layer optical structure, but the grating is modelled as a dipole-filled optical layer with a gradient index.

The theoretical model for the trapezoid grating is illustrated as follows in Figure 4-17.

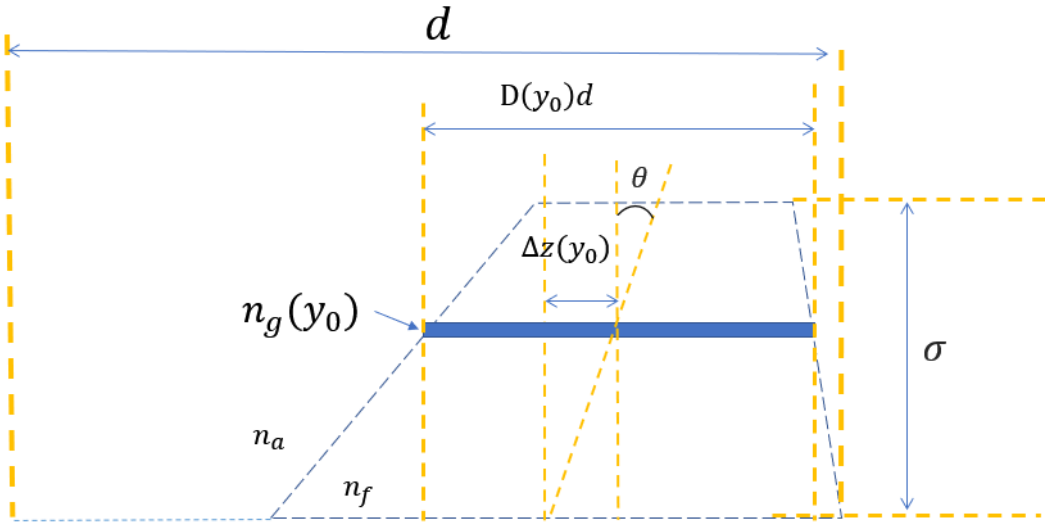

Figure 4-17. Illustration of Theoretical model for trapezoid grating in a single grating period. n_g is the volume averaged index while n_a and n_f are the indices of cladding and core. The tilting angle θ is the angle between central line and norm of the upper and lower edges of the trapezoid, while Δz is the position shift due to tilting.

The effective volume averaged index is a function of y_0 due to the varying indices portion, and can be found as

$$n_g(y_0) = \sqrt{D(y_0)n_f^2 + \big(1 - D(y_0)\big)n_a^2} \tag{4.21}$$

where $D(y_0)$ is the duty cycle of the infinitesimal periodic rectangular shape at y_0 (assuming the trapezoid shape is a composition of a pile of infinitesimal rectangular sheets), as shown in Figure 4-17. The "duty cycle" of the infinitesimal square wave can easily be found with is simple geometry of the trapezoid:

$$D(y_0) = 1 - r - \frac{(1-2r)y_0}{\sigma} \tag{4.22}$$

Apparently, the Fourier component in equation (4.3) for the dipole polarization is now also a function of vertical position y_0 other than a constant, i.e., the Fourier transform of the infinitesimal periodic rectangular wave with a shift is,

$$\tilde{s}(k-\beta)e^{ik_d\Delta z} = \mathrm{d}\sigma\Sigma_{\mathrm{n}=-\infty}^{+\infty} c_n(\mathrm{y}_0)2\pi\delta\left(k - \frac{2\pi n}{d}\right), \tag{4.23}$$

where,

$$c_n(\mathrm{y}_0) = D(y_0)sinc\big(nD(y_0)\big), \tag{4.24}$$

due to the variation of volume averaged index in the grating layer, the y component of the scattered wavevector, k_{yg}, is also a function of vertical position, making the scattered field expressed as

$$\begin{aligned} E_{ns}^{-} &= \int_{-\sigma}^{0} \frac{c_n(y_0)k_0^2\big(n_g^2(y_0) - n_a^2\big)e^{+ik_{yg}(y_0)\sigma}}{2k_{yg}(y_0)\big(1 - r_{ga}R_{gs}e^{+i2k_{yg}(y_0)\sigma}\big)} f_s(y_0)\boldsymbol{E_x}(y_0)e^{k_d \mathrm{y}_0 \sin\theta}dy_0 \\ E_{na}^{+} &= \int_{-\sigma}^{0} \frac{c_n(y_0)k_0^2\big(n_g^2(y_0) - n_a^2\big)e^{+ik_{yg}\sigma}}{2k_{yg}(y_0)\big(1 - r_{ga}R_{gs}e^{+i2k_{yg}(y_0)\sigma}\big)} f_a(y_0)\boldsymbol{E_x}(y_0)e^{k_d \mathrm{y}_0 \sin\theta}dy_0 \end{aligned} \tag{4.25}$$

where $f_{\mathrm{s}}(y_0) = \left(r_{ga}e^{+i\Phi_0} + e^{i\Phi_0}\right)$, $f_{\mathrm{a}}(y_0) = \left(e^{-ik_{yg}\sigma}e^{-i\Phi_0} + R_{gs}e^{+ik_{yg}\sigma}e^{+i\Phi_0}\right)$, same as the definition in 4.2.1.3, expressing the effects of multiple reflections from the interfaces adjacent to the thick grating layer.

For calculating the reflectivity transmissivity and propagating constant in such a four-layer structure with a gradient-indexed layer, the grating layer with continuous gradient index is approximated by discrete index distribution as shown in Figure 4-18, due to the difficulty of including the gradient-index layer.

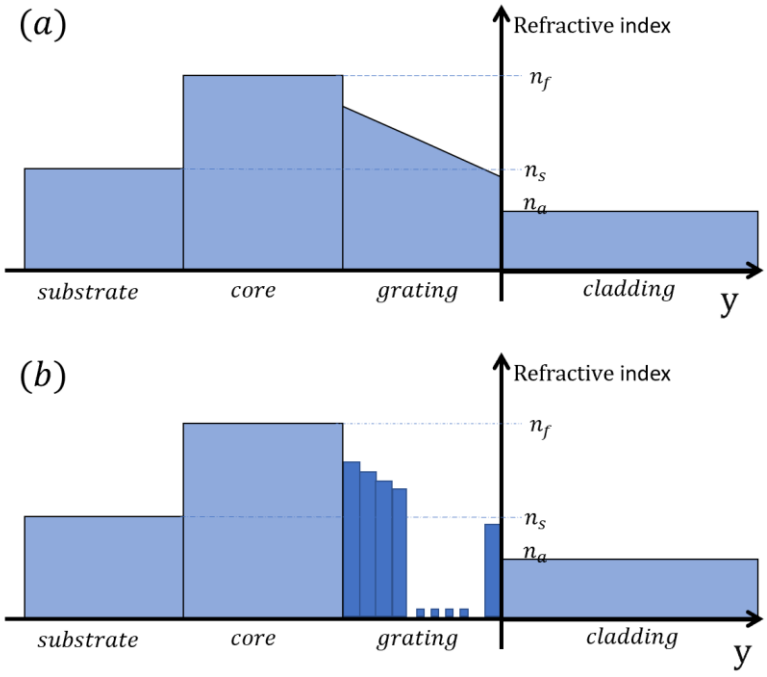

Figure 4-18. Illustration of (a) refractive index distribution in the modelled four-layer structure and (b) approximated index distribution of the four-layer structure for theoretical calculations

After replacing the grating layer gradient index with discrete layers with a constant index that calculated with (4.21) based on position y_0, the approximated structure as shown in Figure 4-18 (b) can be calculated with the transfer matrix method—each discrete layer is mathematically expressed as a matrix as described in sections 2.2.2.2 and 4.2.1.3. With the propagating constant calculated with the approximation made above, the propagating constant related grating period is ready to be obtained with equation (3.11).

4.4.2 Theoretical results and numerical validation

The theoretical calculation for CE as a function of tilted angle and upper edge ratio to study the impact of grating profile to CE. The WGC with 400nm core thickness and 175nm grating depth is adopted for the investigation. Such combination of the WGC parameters, thick core and deep grating, has shown a significant impact of tilting the grating to CE in the last section, while it is also expected to show and to study the impact of grating profile with the deep grating in thick core in this section for improving the CE.

The variation of the grating profile is determined by two factors, upper edge ratio r and tilt angle θ. The CE as a function of $r \in [0,0.5]$ and $\theta \in [-45^o, 45^o]$ is calculated. The value for $r > 0.5$ is not calculated due to the fact that the inverse trapezoid is unlikely to be

feasible. The theoretical result is shown in Figure 4-19. Accordingly, as analysed before, unlike tilting the grating, the upper ratio r causes a non-uniform distribution of volume averaged refractive index that changes the propagating constant in the four-layer structure, although the thickness of the modelled layer has not changed, which makes the grating period to change accordingly. Therefore, the corresponding grating period is calculated and shown in Figure 4-20.

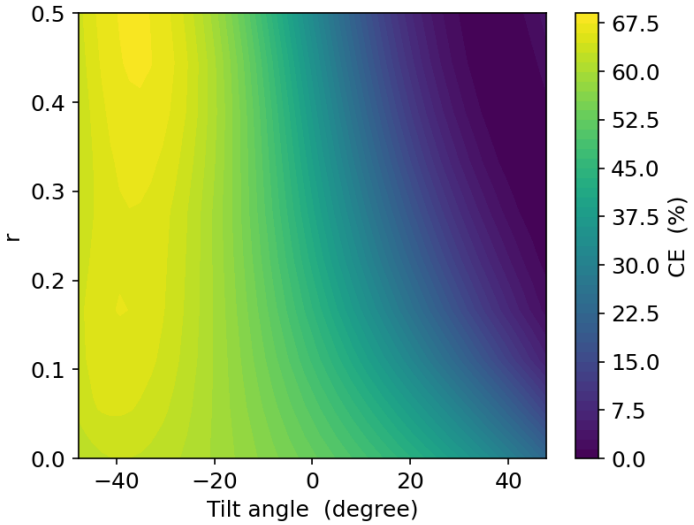

Figure 4-19. CE as a function of tilted angle and r, the offset ratio of the of the upper edged of trapezoid shape defined in Figure 4-17 for the core thickness of 400nm and grating teeth of 175nm.

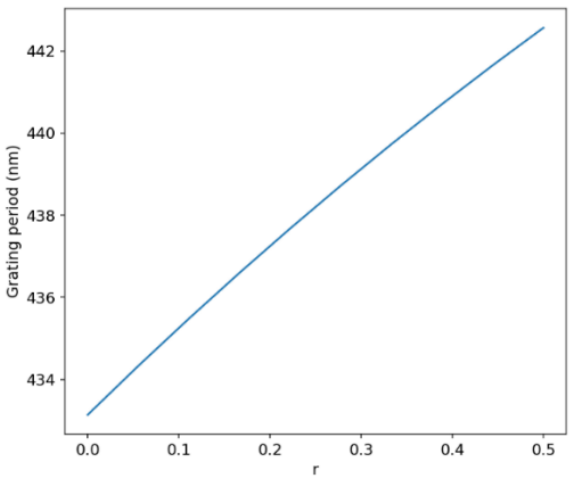

Figure 4-20. Grating period as a function of r, the offset ratio of the of the upper edged of trapezoid shape defined in Figure 4-17 for the core thickness of 400nm and grating teeth of 175nm.

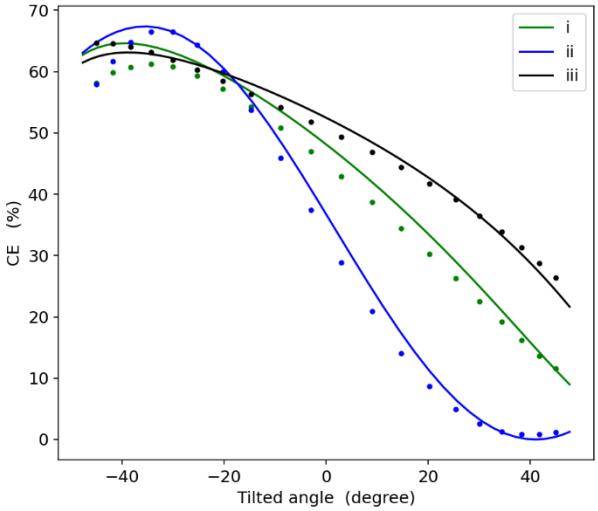

Figure 4-21. CE as a function of tilted angle for (i) r = 0.8, (ii) r = 0.6 and (iii) r = 1. Solid lines are the theoretical results while dots are for numerical simulation.

Due to the same reason as in section 4.3, it is relatively difficult to reproduce the result shown in Figure 4-20 with the numerical method, so the CE for $r = 1, 0.8$ and 0.6 has been selected and reproduced with numerical simulation and is shown in Figure 4-21. Notably, the grating period, although the maximum difference is only around 8nm for different r values, is very sensitive to the CE. The numerical simulation confirms that if the period keeps same as in section 4.3 (r=0.5), the central wavelength of the pump that can be coupled into the WGC shifts by more than 20nm for r=0, resulting in the CE for 1300nm below 10%. Therefore, the grating period has been changed accordingly in the numerical model, adopting the theoretical grating period values for simulation. The numerical simulations with the theoretical grating period produce exact the central wavelength of 1300nm for the in-coupling and produce the results shown in Figure 4-21. The agreement on the grating period from theoretical and numerical results, at least, validified the gradient index hypothesis.

Figure 4-20 indicates that the impact of tilting increases with the ratio of upper edge r, while, for small r, the CE is less sensitive to a tilt angle but stays at a relatively good value. Figure 4-20 shows that the longer upper edge of the trapezoid grating profile needs a longer grating period, which means that the propagating constant decreases with the length of the upper edge (refer to equation (3.11)), although the difference is minimal.

Figure 4-21 shows the trend of the exact CE values from the numerical and theoretical model are the same, while the agreement is still quite good after the approximation for the gradient index layer although slightly worse than the previous sections.

4.4.3 Accuracy analysis

Regarding the discrepancies in Figure 4-21 and Figure 4-14 for the numerical and theoretical results, the factors that impact the numerical accuracy, such as the thickness of the perfect match layer and size of the discrete grade, have been optimized for accuracy. Besides any underlying mechanisms that haven't been covered by this model, there are some possible defects in both the numerical and theoretical models, which are explained as follows.

Firstly, as illustrated in Figure 4-22, in the numerical model, it is impossible to keep the interface of two different materials always on the numerical discrete grid, which means the effective geometry after numerical discretization is distorted. When the discrete grade crosses the non-horizontal or non-vertical interfaces, the smooth interface could be discrete as a staircase shape. This nature of the numerical method is unfavourable to fine structures, including the subwavelength grating teeth, especially the accuracy could be worsened when the sidewall of the grating is not perpendicular to the core.

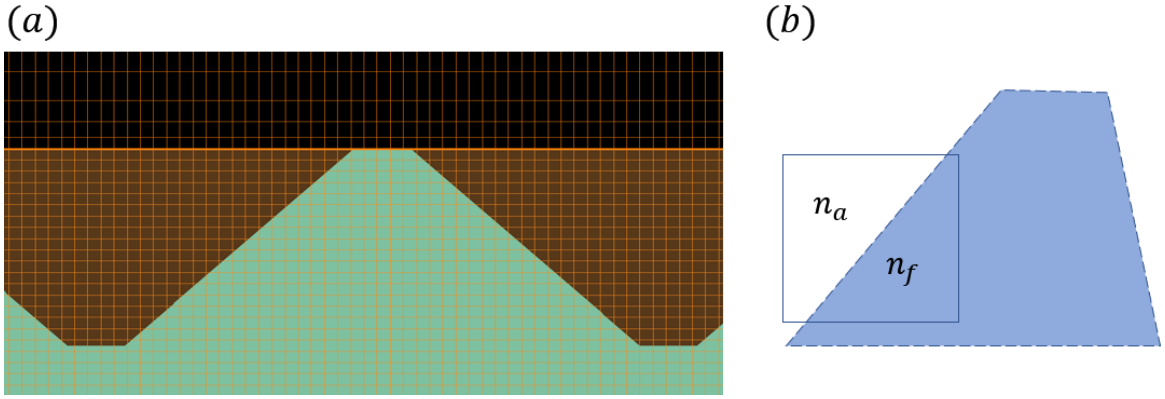

Figure 4-22. Illustration of discretization across interfaces of different material (a) numerical model with the discrete grid. (b) illustration of a single grid crossing two materials.

On the other hand, the analytical theoretical model also suffers from numerical inaccuracies when conducting calculations with computers. Noting that the theoretical model is based on the transfer matrix method, which has been shown to be unstable [143]. One of the main reasons for the instability of the transfer matrix method is that intensive

exponential calculation with extreme values is involved for calculating the phase shift. The phase shift due to propagation is defined as the product of the wavevector and thickness in a specific layer in the transfer matrix. Both the wavevector and thickness of the optical layer are either extremely big or small values for exponential calculations in the transfer matrix method. Especially, when the gradient-index layer is sliced into infinitesimal layers for approximation, the value is more extreme, and inaccuracy is accumulating with an increment of the number of layers. This could result in the inaccuracy of the trapezoid grating profile. On the other hand, reducing the layer will reduce the theoretical accuracy.

Other than the two possible inaccuracies in numerical computation, theoretically there could be some physical mechanisms that were not included in the model. One possibility is the reflection of the sidewalls of the grating profile. However, no publication was found to address this issue and cannot be verified with numerical simulation, and there is no such theoretical approach to include this factor in our model. However, the discrepancy between numerical and theoretical results remains at a limited level.

4.5 Conclusion

This chapter first, following Chapter 3, establishes an analytical theoretical model that accurately calculates the CE of WGC in section 4.2. Modelling of the grating as a modal field polarization source layer and the inclusion of joint loss between grating and waveguide are valid assumptions confirmed by numerical simulation with very good agreement. Important physical mechanisms of WGC operation are revealed with our model. The model is applicable to deep gratings and explicitly shows the impact of multiple reflections from adjacent layers in the grating region on the variation of the total CE with grating depth and core thickness. It also shows that the inclusion of the joint loss is important in calculating the total WGC CE. Such physical insights are lost when using numerical models for WGC optimization. Moreover, the analytical theoretical model is much faster than numerical models, requiring only a few minutes on a standard laptop to cover a much more extended parameter space, which makes it ideal for fast, large-scale WGC optimization.

Based on the clarified physical mechanism of multilayer interference effect that limits the CE of WGC, section 4.3 investigated a method for overcoming such limitation. Manipulating the phase of scattered light by tilting the grating profile has been shown to be a valid

method to enhance the CE for WGC of certain core thickness that suffers low efficiency. The thicker-cored WGC with a regular rectangular grating profile might suffer low CE because of the interference effect weakened scattering order in the substrate, while the thicker core could bear deeper grating, creating space for phase shifting. Grating tilting-induced phase shifting has been shown to have a significant impact on the CE, which has been theoretically calculated and numerically verified with good agreement. The lately successful fabrication of 2D grating with a precisely controlled slanted angle of the grating teeth via ion etch [144,145] intimates the practical value of our approach in this section. The success of modelling the tilted grating with our theory implies that the CE could be optimized by methods that impact the interference effect of scattered light. A simple example is adding extra interfaces in the substrate, just like our experiments to validate the scattering loss model, which could be an easy approach for implementation and can be modelled with our approach.

Section 4.4 focuses on studying the effect of the grating profile, taking variable trapezoids as examples. By approximating the trapezoid shape with discrete infinitesimal rectangles, the complicated grating shapes have been included in our model. The effectiveness of the hypothesis that the grating can be treated as a layer of the non-uniform refractive index has been validified with the accurate prediction of the grating period. The CEs of WGC as a function of grating profile are theoretically calculated and compared to the numerical simulations. The agreement between theoretical and numerical results is good, although not as good as in sections 4.2 and 4.3 where numerical dots are sitting on theoretical curves and possible accuracy-impacting factors are discussed. The work in Section 4.4 is not only valuable for further understanding the nature of WGC but also of value for practical fabrication—the shape of the grating might not be kept rectangular as designed. Moreover, the multilayer approximation is also applicable to more complicated grating shapes, such as the staircase-shaped grating profile that has been proposed and numerically optimized recently [141,142].

In summary, this chapter has established the physical mechanisms that play critical role in our WERS system WGC CE. The model holds the potential to be expanded for more complicated WGC structure/grating profile, and TM modes coupler which has not been fully finished due to the theoretical and coding-wise complexity and traffic in a supercomputer. However, an assertion can be made at this point that the optimal core

thickness of the WERS needs not to be much compromised for the CE of WGC as our previous publication [42]. The CE can be enhanced by tilting the grating teeth with an optimized grating profile. This is particularly important as our new FOM indicates that the thicker core WERS is more favourable, while this chapter proves that the thicker core does not necessarily suffer low CE. Moreover, upon finalizing modelling the TM mode and higher mode in-coupling and combining our new FOM, optimized and detailed WGC design for multimode pump in-coupling for WERS can be produced with the model established in this chapter.

Chapter 5 Post signal processing of WERS spectra

Raman signals are usually weak and, due to its small excitation and emission cross-section, are often obscured by background noise generated by various sources. Signal degradation is particularly severe in WERS due to the core/substrate generated background the wide point spread function of the system. This chapter presents new techniques in denoising, baseline removal and deconvolution. The signal processing has first validified with simulated data, and then applied to the real Raman spectra from our WERS system[119]. In the end, the final processed WERS signal is presented in comparison with Raman microscope spectra, and the detailed middle steps are also presented.

5.1 Multi-frame non-local means (MNLM) based denoising for WERS

5.1.1 Background

As mentioned in the introduction, standard filtering methods suffer from low accuracy and are of a blurring nature, while recent advanced methods can produce good denoising results but mostly suffer from sensitive parameter settings. Savitzky-Golay (SG) filter can potentially remove genuine Raman peaks with low full width at half maximum (FWHM) if the parameters are not appropriately set [63]. For the wavelet transform filtering (WTF), sensitive thresholds for every decomposed frequency component have to be set. In addition, proper wavelets need to be chosen from the large family of wavelets, and complicated optimization algorithms have been developed just for choosing appropriate parameters for WTF [68].

In contrast to the aforementioned neighbour-pixel-based filtering methods, the non-local means (NLM) based denoising method was proposed for image recovery in 2005 [146]. It extracts signals by comparing the similarity of the pixel patch across the whole image and replacing the pixels with similarity-weighted averaging. NLM has demonstrated effective performance in noise reduction without complicated multi-parameter settings. It has been described as “parameter-free” [147] and is known for its ability to preserve high-frequency signals while cancelling out noise in the same high-frequency domain, thus acting as an "edge-preserving denoising method" [148]. Recently, the NLM has also been successfully applied to a periodic 1D signal for fault diagnosis of rolling bearings [149].

A new algorithm, the multi-frame NLM (MNLM) based method, to address current challenges for Raman signal denoising is proposed in this section. In practice, the Raman spectrum for a specific analyte can be easily acquired multiple times, and each frame of the spectrum is identical but polluted with random noise. This study assumes that the noise pattern in different frames acquired from the spectrometer with the same acquisition parameters is similar. Therefore, we expect that the signal can be restored by averaging the pixel blocks across the entire signal range in different frames based on similarity. MNLM is expected to preserve the outstanding features of NLM, such as 'edge-preserving' and 'parameter-free,' to address the weaknesses of current denoising methods.

5.1.2 Mathematic model and implementation

MNLM is demonstrated to restore the Raman signal that is degraded by additive noise, expressed as follows,

$$Y(I,j) = X(I,j) + n(I,j), \tag{5.1}$$

$\boldsymbol{Y} = \{y_i^T \mid i \in N\}$is the 2D matrix of raw data, consisting of N frames spectra in a 1D vector, where $y_i = \{y(j) | \, j \in V\}$ is the i^{th} frame Raman spectrum sub-vector, and $\boldsymbol{Y}(i,j)$ is the observed signal at j^{th} position in i^{th} frame. X is the unpolluted signal and n is the additive noise.

Similarly, to standard NLM, the MNLM filter recovers the estimated signal $\hat{x}$ by weighted averaging of all pixel values in the searching window. However, unlike standard NLM used in imaging, in MNLM, the searching window is extended to other frames that are sampled with the same acquisition parameters, and the weight for averaging is calculated based on the 1D sub-vectors as shown in Figure 5-1 (it is worth noting that the calculation in standard NLM is performed on 2D sub-matrices. The recovered j_1^{th} pixel block in i_1^{th} frame is expressed as

$$\hat{x}_{i_1}(j_1) = \sum_{j_2 \in V} \sum_{i_2 \in V} w_{i_1,j_1}(i_2,j_2) y_{i_2}(j_2) \ , \tag{5.2}$$

where $w_{i_1,j_1}(i_2,j_2)$is the normalized weight, and the value is determined by the similarity between the j_1^{th} pixel block M_{j_1}, i.e. $y_{i_1}(M_{j_1}) = \{y(j) | j \in M_{j_1}\}$, in i_1^{th} frame and the j_2^{th}

pixel block M_{j_2} in the i_2^{th} frame. The similarity between M_{j_1} and M_{j_2} is measured by the Gaussian-weighted Euclidean distance between the two vectors, namely,

$$d^2_{i_1,j_1}(i_2,j_2) = ||y_{i_1}(M_{j_1}) - y_{i_2}(M_{j_2})||^2_{2,a}, \tag{5.3}$$

where a is the standard deviation of the Gaussian kernel, $\|*\|^2_{2,a}$ L^2 norm with Gaussian convolution. The purpose of applying Gaussian weighting to the Euclidean distance between the two vectors is to emphasize the significance of the central pixel in the neighborhood. As the similarity between the neighborhoods M_{j_1} and M_{j_2} increases, the Gaussian-weighted Euclidean distance $d_{i_1,j_1}(i_2,j_2)$ between the corresponding vectors $y_{i_1}(M_{j_1})$ and $y_{i_2}(M_{j_2})$ decreases, resulting in a higher weight of a pixel j_1 during the calculation of the average value

$$w_{i_1,j_1}(i_2,j_2) = \frac{1}{Z_{i_1,j_1}} e^{-\frac{d^2_{i_1,j_1}(i_2,j_2)}{h^2}}, \tag{5.4}$$

where Z_{i_1,j_1} is a normalization parameter to ensure $\Sigma w_{i_1,j_1} = 1$, and h is the smoothing factor.

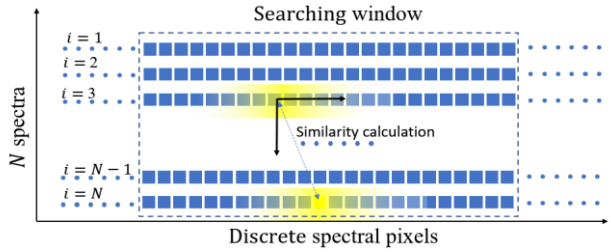

Figure 5-1. Diagram of MNLM searching performed on N spectra. The Gaussian-weighted Euclidean distance is calculated in the whole searching window in N frames.

The key function of the algorithm is in calculating the weights for each pixel, which depend on the similarity calculated by Gaussian-weighted Euclidean distance in equation (5.3). Figure 5-1 illustrates the processing of searching similarity between each two pixel-blocks. With the Gaussian-weighted Euclidean distance, the weight for each pixel can be obtained with equation (5.4). Finally, the pixel value is restored by weighted averaging of all the pixels in the searching area across different frames.

To validate the algorithm, an artificial Raman spectrum is generated by applying the following procedure. A frame of the Raman spectrum is simulated with multiple Lorentz peaks, as usually utilized for Raman peak fitting [150]. Each frame contains 1000 data points, including random 7-15 Lorentzian peaks with random positions, random heights within 0

to 1, and random half-widths between 1 and 7 pixels. The noise is an artificially generated random matrix with a size of $N \times 1000$, namely N frame noise vectors. Each signal vector is added with random vectors in the noise matrix to obtain N frames data set, simulating N frames of noisy Raman signal acquired from the same spectrometer. This procedure has been repeated 30 times, so a dataset with $N \times 30$ frames of the simulated signal is prepared for validating the algorithm.

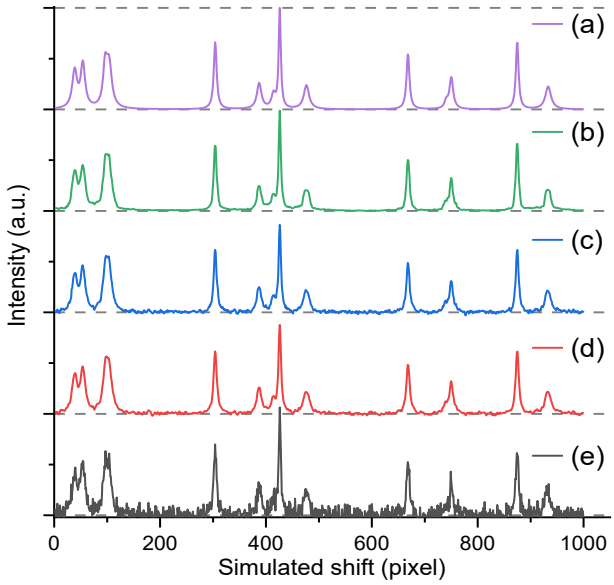

Figure 5-2. An example of simulated Raman spectra including the (a) original spectrum, the (e) noise-degraded spectrum, and the restored spectrum using different methods, including (b) MNLM filtering (c) Gaussian filtering (d) median filtering. The noise level is σ_n =0.05, and the frame number is 5, indicating that the original signal is added to 5 different noise vectors, simulating the acquisition of 5 frames of signals with the same parameters. The spectra shown are all 5-frame averaged. The smoothing factor $h = 0.2\ \sigma_n$, and the length of searching window is 30 pixels.

To start the validation process, the standard deviation, σ_n, of the random matrix is set to 0.05. A 5-frame averaged original, degraded, MNLM processed signal is shown in Figure 5-2. The Gaussian filtering and median filtering as the two standard denoising approaches are also conducted for comparison. The results show that the fine structure of the original signal, e.g., the overlapping double-peaks around 50, 420 and 750 are preserved by all the methods, however, the MNLM method shows evidence of a much better performance in removing noise without over flattening the signal peaks (e.g., the main peak at 420), which will be quantified below.

To validate the hypothesis made at the beginning of this section that applying NLM within multiple frames could help with cancelling unwanted noise, the mean squared error (MSE) against the frame number N is calculated. After generating 30 sets of the original signals, a group of N frames of noise is generated and added to each set of the original signal. So, for each specific original signal, there are N frames of corresponding degraded signal, simulating the Raman spectrum acquired *N* times from a specific sample. For each set of signal data, averaging, averaging after Gaussian filtering and averaging after MNLM are conducted, and then compared to the original signal to obtain the MSE. After repeating this process for all 30 sets of data, the average MSE is obtained and shown in Figure 5-3, vs the number of frames, N, used. MNLM consistently outperforms the other methods when the frame number is greater than 1, which clearly indicates that our assumption is valid. By restoring pixels through averaging based on similarity across different frames, noise cancellation performance is enhanced. Figure 5-3 shows that more frames create smaller MSE, but cost more time in practice. Figure 5-3 also shows 5 frames are already enough for MNLM to produce good results so frame number *N* is fixed to 5 for further validation.

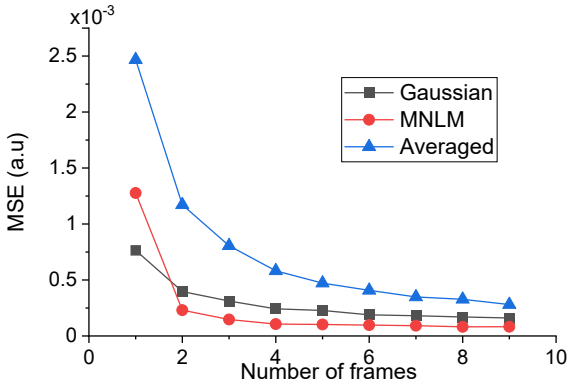

Figure 5-3. Average of mean square error of restored data with increasing number of frames used. The smoothing factor $h = 0.2\ \sigma_n$, and the length of searching window in each frame, w, is 30 pixels.

As mentioned in the introduction, some advanced filters, such as SG and WTF, suffer from complicated and sensitive parameter setting which limits their application. In contrast, only two parameters: the length of searching window w and smoothing factor h need to be set for NMLM. Ref. [151] indicated that the denoising results are weakly sensitive to the values of these parameters for NLM. There, the length of the searching window is around 20~30 pixels and smoothing factor is related to the noise, i.e., $h = k\sigma_n$, where the k value is around 0.5 [151]. It is expected that the performance of the MNLM will also be tolerant to

variation in these parameters. Therefore, parameter-sweeping for k and w was conducted, and the result is shown in Figure 5-4. The k value is within the range from 0.1 to 1 while the w is from 3 to 30 pixels. As shown in Figure 5-4, The MSE is for the denoised result is within a range of 0.8× 10^{-4} to 1.7× 10^{-4}, and for the most of the cases the MSE is around 1×10^{-4}, besides some extreme $k-w$ combinations, e.g. (k~0.1, w~3) and (k~1, w~30). It is evident that the weak sensitivity to parameters has been inherited by MNLM, making it a potential parameter-free denoising method for wider application.

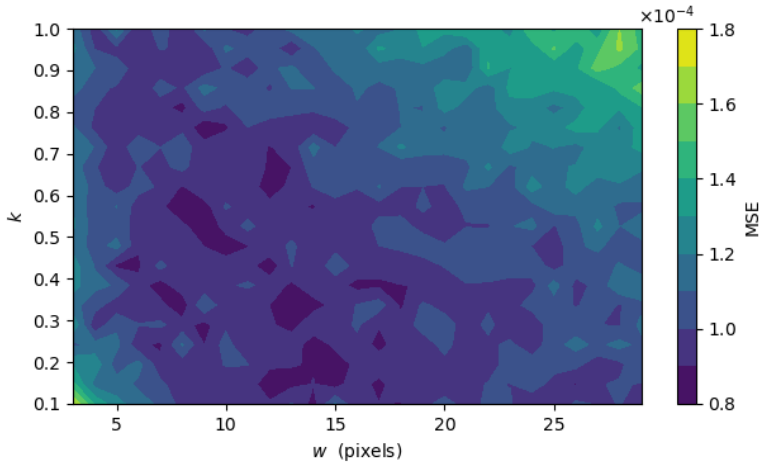

Figure 5-4. MSE as function of NMLM parameter k and the length of the search window, w. The frame number N is 5; σ_n=0.05

Filter	$*k$	$\sigma_n = 0.01$			$\sigma_n = 0.05$			$\sigma_n = 0.1$			$\sigma_n = 0.2$		
		(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
Gaussian	0.1	27.4	0.021	46.8	13.5	0.5	33.1	7.4	1.8	27.4	0.5	8.2	20.9
	0.5	**27.9**	**0.019**	**47.3**	15.3	0.3	35	9.3	1.2	29.3	2.4	5.3	22.7
	1	22.0	0.15	41.4	**16.6**	**0.2**	**36.3**	**12.2**	**0.6**	**32.1**	5.9	2.4	26.2
	2	14.7	0.41	34.1	12.9	0.6	32.5	11.4	0.8	31.3	**7.9**	**1.5**	**28.2**
Median	3	**19.0**	**0.15**	**38.4**	**15.1**	**0.3**	**34.8**	**11.3**	**0.7**	**31.3**	5.8	2.5	26.1
	5	10.9	0.96	30.3	10.0	1.1	29.6	9.0	1.3	29	**6.7**	**2.1**	**27**
	7	7.6	2.0	27	6.2	2.7	25.9	5.4	3.0	25.4	3.7	4.2	24
	9	4.2	4.6	23.6	3.2	5.5	22.9	2.3	6.3	22.3	0.9	8.3	21.2
MNLM	/	**30.5**	**0.011**	**49.9**	**20.4**	**0.1**	**40**	**15.0**	**0.3**	**35**	**9.1**	**1.2**	**29.5**
avg. signal	/	27.4	0.021	46.8	13.5	0.5	33.1	7.4	1.8	27.4	0.5	8.2	20.9

Table 5-1 Indicators including (a) signal noise ratio (b) Mean square error (Value shown in the table is 1000 times MSE) and (c) peak signal noise ratio for the recovered signal with the best performance for each filter and noise level in bold (*Value shown in the table is 1000 times MSE). *Filter parameter k are standard deviation and filer window for Gaussian and Median filter, respectively. Last line shows the averaged signal.

A detailed comparison is presented in Table 5-1 to thoroughly validate the performance of denoising using different denoising methods, including Gaussian and median filtering, under various levels of noise. The noise matrix is generated with different standard deviations (σ_n=0.01, 0.05, 0.1, and 0.2) and denoising is processed using Gaussian filtering, median filtering, and MNLM filtering. The parameters for the MNLM filtering are kept unchanged (w is 30 pixels, and h is fixed to 0.01, simulating that the user inaccurately estimates the noise level) to test the robustness of the method without tailoring the parameters to the specific data, while other two filters are tested with four different filtering parameters each. Specifically, for the Gaussian filtering, the parameter k (i.e., the standard deviation for the Gaussian kernel) is tested for k= 0.1, 0.5, 1, and 2, and for the median filter, the parameter k (i.e., the size of the filter window) is tested for k= 3, 5, 7, and 9. Therefore, 30 sets of signal data were generated for each noise level, with each set of data containing 5 frames of the same signal polluted with random noise of the same standard deviations σ_n. Each set of the signal spectrum is then processed with the aforementioned filters, and the restored signal is averaged for comparison with the original signal, to obtain indicators including single-noise-ratio (SNR), peak signal noise ratio (PSNR), and MSE. After obtaining these indicators for the 30 sets, the value of each indicator for each filter and averaged signal with noise is averaged and recorded in Table 5-1.

Table 5-1 shows the indicators of restored signal from Gaussian filtering and median filtering with different parameters for different-level noise, meanwhile the indicators from MNLM with fixed parameters. It is clearly shown that the performance of each filter varies with noise level and filtering parameters. However, the MNLM with fixed parameters outperforms all other filters, regardless of the noise level. Thus, after validation with different indicators and comparisons, MNLM has demonstrated outstanding performance and robustness, making it ready for application to practical experimental Raman spectra.

5.2 Baseline constrained blind deconvolution

5.2.1 Background

The baseline removal techniques summarized in the introduction chapter can be roughly categorified into two terms. The first method is termed filtering. The baseline is supposed to be the lower frequency part of the signal, thus, by filtering out the high-frequency

component, the baseline can be found. These methods include the Fourier transform [76] [77] and wavelet transform [78] [79] approaches. The other method can be classified as fitting, which fit a smooth curve to closely match the spectra but always underneath the steep Raman peaks. Asymmetric least square [73], spline smoothing [74], and polynomial fitting [75] are the most studied methods. Due to the inherent accuracy limitations of the polynomial fitting, it is not very suitable for complicated curves. The latest asymmetric least squares (ALS) method, which is essentially a fitting method, has shown state-of-the-art results, and the model can be expressed as [73],[152], $\mathrm{b} = \arg\min_{W,b} W||s-b||^2 + \lambda||\nabla b||^2$, where b is the baseline for fitting, s is the spectra vector and $\lambda||\nabla b||^2$ is the smoother and W is the weight for the least square. Mathematically, the vector b is of same dimension of the spectra vector $\mathbf{s}$, which is the number of pixels. So, at first, it might be facing the risk of overfitting. And secondly, the smoother item could be problematic—when the Raman peak has a similar gradient with noise or the baseline itself, the operator ∇ could provide an incorrect regularization.

To address the limitations of the current ALS method, we propose employing a radial basis function (RBF) [153,154] to model and fit the baseline. The RBF, which is based on kernel methods [153], enables the fitting to be conducted in a reduced-dimensional space. In theory, lower dimension fitting could reduce the risk of overfitting and practically the curvature of the fitted baseline can be controlled with the width of the RBF.

Moreover, regarding the fact that the baseline should always be below the Raman peaks, the baseline applied as a non-negative constraint to blind deconvolution is also implemented aiming at reducing the ringing artefact. The blind deconvolution that is integrated with the baseline removal is aiming at reducing the impact of resolution degradation due to multimode-collection induced wide PSF. The deconvolution is developed from the Wiener filter-based imaging deconvolution proposed in [84] by implementing a total variation (TV) regularization [83] for suppressing ringing artefacts.

5.2.2 Mathematic model and implementation

Besides the noise, the unpolluted signal x that is defined in equation (5.1) is, however, not the “true” signal but a broadened signal on a baseline due to the point spread function (PSF) instrument function. WERS often suffers from wider PSF as it commonly uses highly

multimode fibre to collect weak Raman signal as much as possible. The baseline, as discussed in the introduction chapter, typically cannot be completely removed via normalization and background subtraction. The unpolluted signal is therefore expressed as

$$x(j) = s(j) \otimes h(j), s(j) = f(j) + b(j) \tag{5.5}$$

where $x(j)$, $s(j)$, $h(j)$, $f(j)$ and $b(j)$ are the unpolluted spectra, PSF of the instrument, true spectra $f(j)$ with baseline $b(j)$added respect to Raman shift j. The operator $\otimes$ is the convolution operator. Usually, the PSF, baseline, and noise are all unknown.

To obtain the "true" spectra $f(j)$, deconvolution and baseline removal need to be conducted to the unpolluted signal x. The unpolluted signal x is replaced with a denoised signal $\hat{x}$, as it is practically impossible to obtain noise-free Raman spectra. In this sense, denoising is a critical procedure for removing noise as much as possible because deconvolution is a typical ill-conditioned processing. The notorious ringing artefacts are the most disturbing artefact in deconvolution, creating ripples that appear near the strong peaks (1D signals) and strong edges (2D images). The ill-conditioning has been explained with many theories, among which the Gibbs effect in the numerical algorithms for deconvolution is mostly attributed to [155,156]. However, other than specifically focusing on the mechanism of ringing artefact, a brief explanation is given as follows. According to the convolution theorem, convolution is equivalent to multiplication in Fourier transformed domain, namely:

$$x(j) = s(j) \otimes h(j) \leftrightarrow \hat{x}(\omega) = \hat{s}(\omega) * \hat{h}(\omega) \tag{5.6}$$

So, the deconvolution is simply,

$$\hat{s}(\omega) = \frac{\hat{x}(\omega)}{\hat{h}(\omega)} \tag{5.7}$$

while it is possible at some point $\exists h(\omega) = 0$, which makes $\hat{s}(\omega) = \frac{\hat{x}(\omega)}{0}$. This would most likely happen in any numerical deconvolution model. The PSF typically contains zero or close-to-zero components (which will be truncated to zero in the numerical model) in Fourier transformed domain, resulting $\frac{\hat{x}(\omega)}{0}$.

Due to the possible ringing artefact that could amplify the noise (high-frequency peaks), more than precisely removing the noise with the dedicated method as a must pre-

processing, including regularization and constrain into the deconvolution is a common method that reduces the impact of the ringing artefact.

Given the current challenges in accurately measuring the PSF directly from the system, the blind-deconvolution method, by which the PSF and signal can be obtained at the same time with an iterative method, follows Wiener's filter-based deconvolution proposed by Ref. [84]. The detailed steps are presented after clarifying the mathematics. As discussed, an improvement in this work is that the total variation (TV) regularization, which has been shown to have a significant effect on suppressing the ringing artefacts has been incorporated [83]. The mathematical formulation is,

$$\underset{s}{argmin} f(s) = \underset{s}{argmin}\ ||s \otimes h - y||^2 + \lambda_1 ||\boldsymbol{s}||^2 + \lambda_2 \int_\Omega |\mathrm{diff}\, \boldsymbol{s}| dj \qquad (5.8)$$

Where λ_1 and λ_2 are the strength of the regularizations, and $\mathrm{diff}\, \boldsymbol{s}$ is the difference of the spectra. Replacing difference item (diff) the with Laplacian operator∇, which returns the second-order difference, and turning the expression in the frequential domain, the deconvolution turns to:

$$\underset{s}{argmin} f(\hat{s}) = \underset{s}{argmin}\ ||\hat{s} * \hat{h} - \hat{y}||^2 + \lambda_1 ||\hat{s}||^2 + \lambda_2 ||\widehat{\nabla} * \hat{s}||^2 \qquad (5.9)$$

Where $\widehat{\nabla}$ is the Fourier-transformed Laplacian operator, and other variables with the hat sign denote the corresponding Fourier-transformed variables in (5.8). The convexity of (5.9) can be easily proved with $\frac{\partial^2 f(\hat{s})}{\partial \hat{s}^2} < 0$, which means that, unlike the optimization from WGC in previous chapter, there is no local optima. Then $\hat{s}$ can be found by letting $\frac{\partial f(\hat{s})}{\partial \hat{s}}$=0, so

$$\hat{s} = \frac{\hat{h}}{\hat{h}^2 + \lambda_1 + \lambda_2 \widehat{\nabla}^2} \hat{y} \qquad (5.10)$$

Then the optimization of signal and PSF is following the steps described in Ref. [84], but replace Wiener's filter with (5.10). The signal and PSF can be obtained with a fast convergence rate by cross iterating $\hat{s}$ and $\hat{h}$ with (5.10) (i.e. exchange $\hat{s}$ and $\hat{h}$ in the equation every other iteration, following the method provided by Ref. [84]). The initial width of PSF can be approximated by measuring the thinnest Raman peak in the spectrum.

The baseline, as can be inferred from equation (5.5), is added to the real signal and broadened by the PSF. And, as mentioned in the introduction, the asymmetric least square (ALS) is a powerful algorithm commonly used for baseline removal in Raman spectroscopy data analysis [73,157]. The main idea of ALS can be simply concluded as follow. As known, typically, the useful Raman peaks in a spectrum varies much faster than the baseline, while the much smoother baseline can be fitted by least square method. Differed from the standard least square, there is a penalty term as a smoother added in the least square and also a weight to the least square, and can be expressed as

$$\mathrm{Arg}\min_{W,b} g(b) = \arg\min_{W,b} W||s-b||^2 + \lambda||\nabla b||^2 \tag{5.11}$$

where b is the estimated baseline, and $||\nabla b||^2$ is the smoother with the strength of λ, and W is the weight of the ALS: ideally, after the optimization weight is a bigger value, e.g., 1, for the non-peak pixels and a smaller value, e.g., 0, for the pixels in the peak area.

As can be inferred from the expression of ALS, the premise of fitting out the baseline is that the gradient-based smoother $||\nabla b||^2$ could successfully identify the peak so that the weight can be successfully calculated. However, in WERS, the baseline is generated with a complicated source. Except for the residual pump laser that most Raman techniques suffer from, waveguide material background and waveguide dispersion can both create a complicated baseline. Moreover, in WERS, the Raman peaks can be broadened with the wider PSF to have a similar gradient, creating difficulties for ALS in accurately identifying the baseline. As a result, the fitted baseline could bend to the Raman peaks (c.f. Figure 5-5(b)).

Here, a simple method is proposed to solve this problem. The main idea of the improvement is using ALS to train a basic radial basis function (RBF) interpolation, which is essentially an RBF neuron network (NN) [153,154], to approximate the baseline, other than fitting the baseline directly with ALS. The advantages of using the RBFNN is specified after showing the mathematical expression.

The mathematical formulation for baseline is modelled as

$$b(j) = U(c)G(c,j) \tag{5.12}$$

$$\mathrm{Arg}\min_{U,W} g(U) = \arg\min_{U,W} W||s-UG||^2 + \lambda||U||^2 \tag{5.13}$$

As shown in (5.13) the baseline $b(j)$ is approached via $UG(c,j)$, where $G(c,j)$ is the Green's matrix [153,154] and U is the weight vector for the RBF regression. And U is obtained from the optimization of (5.13). Here the Green's matrix is calculated with the Gaussian function as the RBF kernel. The parameter c is the number of RBF centres, which can be much smaller than the dimension of original signal s. And the vector W in (5.13) is the weight vector defined as same as in ref [73,152]—takes big values when $s < UG(c,j)$ and small value otherwise. The item $||U||^2$ is the L2 regulator with strength λ, which provides smoothing effect similar to the smoother in ALS with simpler expression.

The difference compared to standard RBF NN is that the training is based on the ALS. According to the theory of RBF NN, the accuracy can be tuned just by controlling the number of the RBF centres, while the smoothness of the output can be controlled via the width of the RBF and the L2 regulator. Thus, compared to ALS, the dimension has been reduced via applying RBF, which means, theoretically, the risk of overfitting is reduced. The bending to Raman peaks of fitted baseline can be simply avoided by setting the width of RBF wider than the Raman peaks. Figure 5-5 provides an example of with and without applying RBF NN for the baseline extraction.

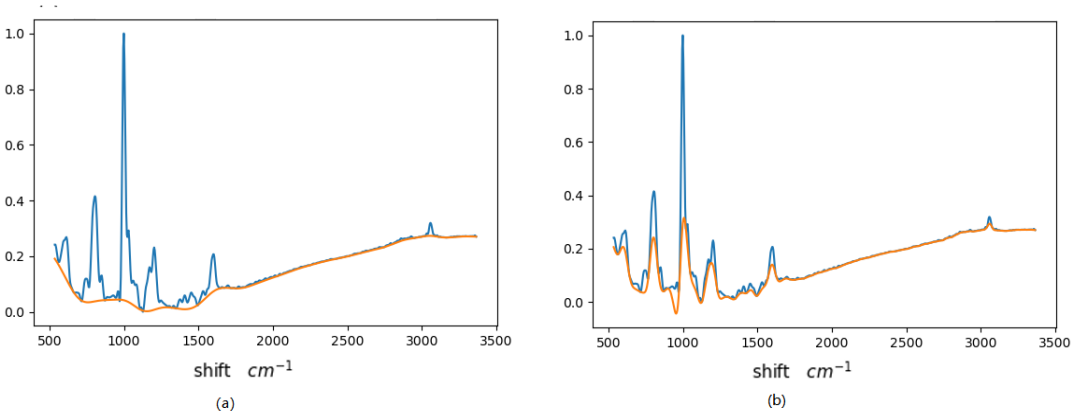

Figure 5-5. ALS extracted baseline (a) with and (b) without RBF NN applied. The strength of smoother is the same for the two sub-figures. The baseline is intentionally made complicated by conducting min-max normalization out of region of interest (ROI), to show the effect of our approach.

Another physical factor mentioned in the introduction is the non-negative constraint for the deconvolution – the physical fact that has not been considered in current Raman spectra deconvolution is that the baseline should be always beneath the Raman signals. This sets a non-negative constraint for the deconvolution. As aforementioned, constraints could help with reducing the ringing artefact. So, the mathematical formulation for

deconvolution, which integrates the algorithms of deconvolution and baseline removal, is converted to,

$$\begin{cases} \underset{s}{\text{Arg}}\, minf(s) = \underset{s}{argmin}\ ||s \otimes h - y||^2 + \lambda_1||s||^2 + \lambda_2||\nabla \otimes s||^2 \\ s.t. \quad s - UG > 0 \end{cases} \tag{5.14}$$

Due to the appearance of the inequality, unlike (5.10), it is not easy, for the moment, to find the one-step results for the optimization problem. So, the traditional but robust gradient descendent method is applied to tackle the optimization of the equation (5.14).

Figure 5-6 shows the significant impact of including the baseline non-negative constraint for the deconvolution. The inverse peak around 1000 cm^{-1}, which is clearly an artifact, has been suppressed. It is worth noting that the spectra shown in Figure 5-6 is from our early-stage system, which suffers a lower signal-noise ratio, to show the significance of the approach of non-negative constraint.

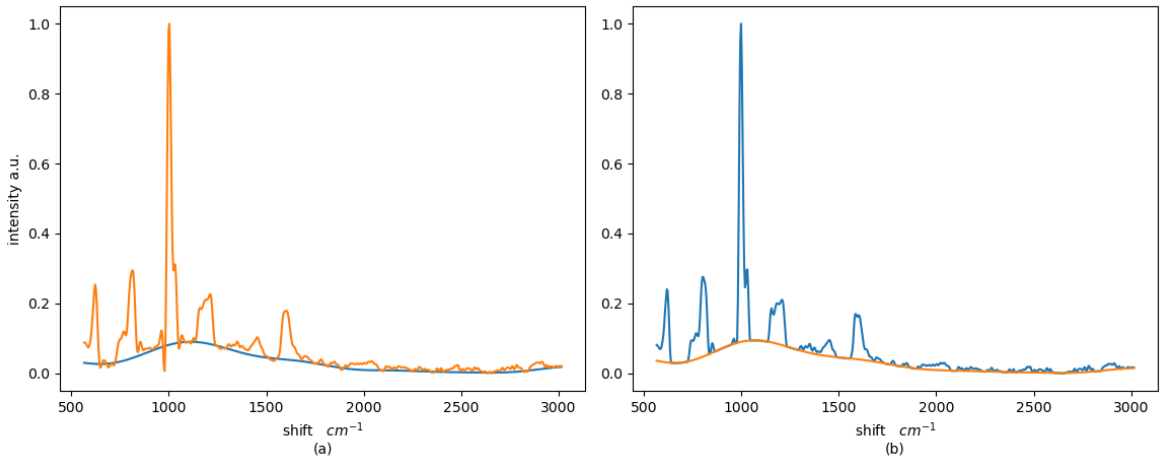

Figure 5-6. Results of blind deconvolution, baseline extraction. (a) Results of blind deconvolution, initial baseline extraction. (b) Non-negative constrained blind deconvolution and refined baseline.

5.3 Results and discussion

Therefore, after clarifying the mathematics, the steps of processing the spectra are detailed as follows.

Procedure 1:

Step 1: Normalization

- *min-max normalize the multi-frame spectra, with and without analyte, at region of interest (ROI).*

Step 2: Background subtraction

- *Normalized spectra with analyte minus normalized spectra without analyte*

Step 3: denoising

- Roughly extract and remove the residual baseline via low-pass filter
- Denoising with the method detailed in section 5.1
- Add the baseline extract in step 3.1 back to the denoised signal as unpolluted signal for processing in Procedure 2

Procedure 2:

Step 1: Optimize the $\hat{s}$ and $\hat{h}$ via cross iteration with (5.10). Fourier transform them back to s and h.

Step 2: Feed s into RBF NN (5.12)(5.13) to obtain the baseline b:

- 2.0 Initialize the weight $W = [1,1,1 \dots]^T$
- 2.1 Obtain U with (5.13)
- 2.2 Update W
- 2.3 Repeat 2.1 - 2.2 until expected $g(U)$ or iterations reached. Numerical experiments have shown that 20 times iteration seems to be an economic setting.
- obtain the baseline b with (5.12)

Step 3: Feed b and the h to (5.14); update the s

Step 4: Repeat Step 3- Step 4 until iterations are reached. Typically, this iteration stops within 5 times.

It worth noting that in procedure 1 the baseline is roughly extracted and removed with a simple low-pass filter is to reduce the impact of the baseline on the similarity calculation in MNLM. Adding the roughly extracted baseline back is for accurately deconvolving the spectra – a slight distortion of the spectra could result in amplified distortion with deconvolution. Therefore, the accurate baseline removal is conducted with deconvolution as detailed in procedure 2.

Unlike denoising, which can be characterized with simulated data, the baseline can hardly be simulated and characterized. To be specific, the baseline varies with multiple factors and no mathematical approach has been found to simulate the baseline that could cover all the factors that contribute to the baseline in WERS, and there are no mathematical indicators to characterize the method. Moreover, our approach incorporates baseline removal and

deconvolution as integrated processing. Therefore, the baseline removal and deconvolution are conducted to the denoised experimental data. The processed data, after procedure 2, is compared to the high-quality spectrum obtained with an expensive Raman microscope for the same analyte, to show the effectiveness of our processing.

To validate the MNLM and following processing on real data, Raman spectra for benzyl alcohol were obtained using our experimental waveguide-enhanced Raman spectroscopy (WERS) system [119]. The WERS system is a implementation of our previous design[42]. The benzyl alcohol as the testing analyte was provided by our partner from chemistry school along with the Raman spectrum from a conventional Raman microscope. The WERS signal in the paper is obtained by Dr. Ettabib and processed with the algorithms developed in this chapter. The WERS The system has been optimized and finely tuned to generate better quality Raman spectra than that shown in Figure 5-6. A measured baseline Raman spectrum for benzyl alcohol was obtained using a conventional Raman microscope.

Firstly, denoising of the WERS spectrum was carried out with MNLM following the steps described in procedure 1 maintaining the same parameters used in section 5.1. After MNLM denoising, the baseline-constrained blind deconvolution was conducted to increase the resolution that was degraded by the instrument function of the WERS system following the steps detailed in procedure 2. The results for each step is shown in Figure 5-7.

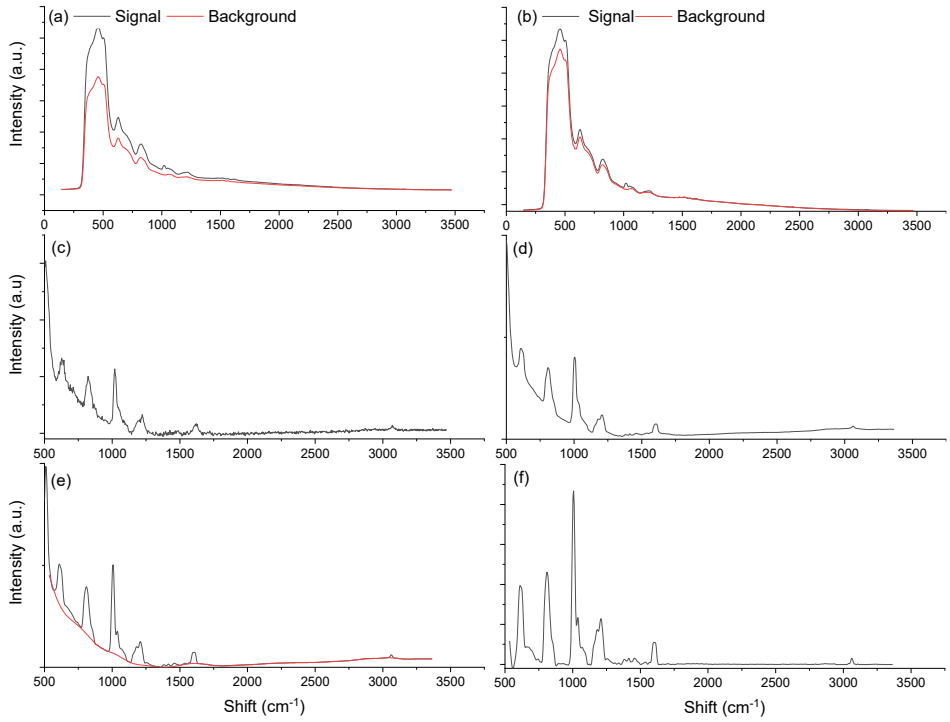

Figure 5-7. Results for each step (a) Raw background and signal; (b) Normalized; (c) Background subtracted; (d) Denoised; (e) Deconvoluted with extracted baseline; (f) Processed signal.

Figure 5-7 is showing a typical processing of the raw Raman spectrum from a slab WERS system. (a) is the raw average spectra with (marked as signal) and without analyte (marked as background) on the top of the waveguide. The spectra are obtained with the exact same parameters, while the shift of the spectra could be caused by the refractive index changing with and without the analyte applied. Therefore, a min-max normalization in ROI (500~1500 cm^{-1}) is applied, as shown in (b). (c) shows the spectrum after background subtraction, where the residual background, known as baseline, is not completely cancelled out. The denoising is then applied to the Raman spectra, and the result is shown in (d). After denoising, a blind deconvolution is conducted. And lastly, the baseline is fitted after deconvolution, as shown in (e). The restored Raman spectrum is finally obtained after subtracting the baseline, as shown in (f).

Lastly, the processed result is taken out for comparison with the Raman microscope. As mentioned, deconvolution is an ill-conditioned process, and noise can easily cause artefacts or distortion when performing deconvolution. However, as shown in Figure 5-8, some fine structures measured using the Raman microscope have been recovered, such as the multi-peaks around 1000 cm^{-1} and 1300 $cm^{-1}$and the very small peak around 3100 cm^{-1}.

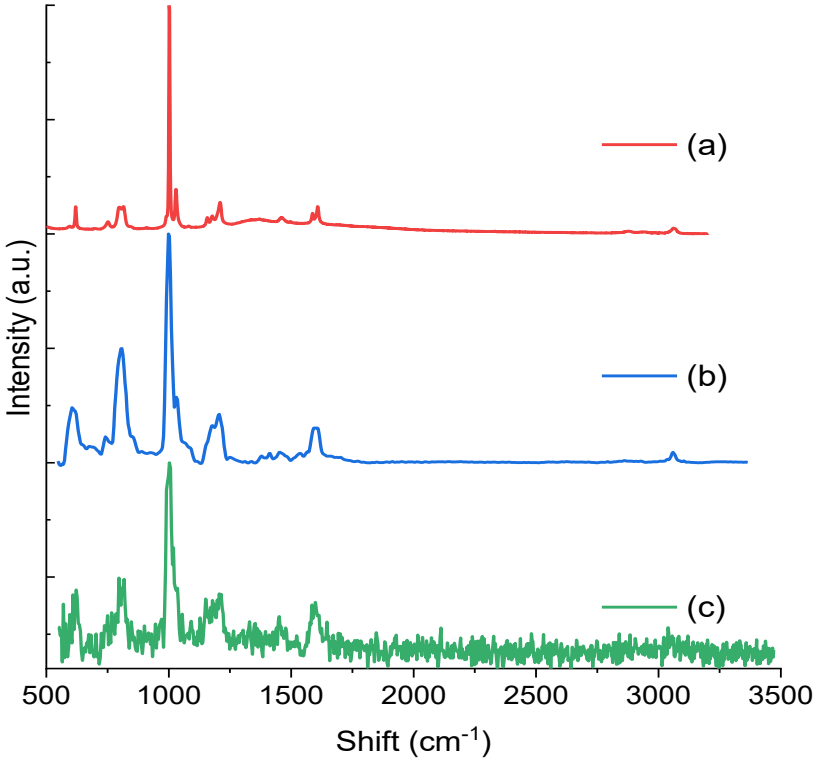

Figure 5-8. Raman spectra from (a) a Raman microscope, (b) the processed experimental WERS spectrum, and (c) the experimental WERS measurement for benzyl alcohol (baseline removed for clarity). Spectra are normalized and shifted for clarity.

5.4 Conclusion

In summary, the NLM denoising method has been developed and extended to a novel MNLM filter, enabling the use of information across different frames and applicable to 1D signals. Non-local filtering is more robust than classical conventional neighbour pixels-based approaches such as Gaussian and median filtering. This feature has been extended to multi-frame based 1D signal denoising. Both statistical indicators for the quality of restored artificial signal and experimental data processing have proven the good performance of our MNLM filter and the insensitivity to the filter parameters. This can significantly benefit non-expert users and potentially enable fully-automized Raman signal processing.

ALS has been developed by incorporating the RBF NN for baseline extraction, which has been illustrated to have a good performance on avoiding the baseline bending to Raman peaks when the baseline is complicated and Raman peaks are wide.

The baseline removal has been mathematically integrated into the deconvolution as a non-negative constraint, following the physical fact that the Raman peaks are always above the baseline. The non-negative constraint has been shown to have a significant impact on reductions of ringing artifacts, especially when the SNR of the signal is low.

After building up the mathematical model of data processing, and clarifying each step of implementation of the algorithms bundle, the steps of processing experimental data have been shown in detail. The final result of processed experimental data is presented and compared to the Raman spectrum from an expensive conventional Raman microscope. Fine structures and subtle Raman features have been restored after the processing. Disadvantages, such as strong background due to the waveguide material and wide PSF, have been alleviated to a great extent via the computational method with minimal cost.

Chapter 6 Conclusion and outlook

6.1 Conclusion

After a brief introduction on WERS and the relevant challenges, the optimization of waveguides for advanced Waveguide Enhanced Raman Spectroscopy (WERS) sensor applications is conducted in the first part of the thesis. Conventional optimization methods focus on maximizing the surface intensity of the pump and signal under single-mode operation, but this approach leads to increased surface-induced scattering losses, propagation losses, and reduced Raman collection efficiency. The impact of planar waveguide thickness on surface scattering loss is studied theoretically and experimentally. The results show that surface scattering loss is influenced by interference effects due to reflections at different interfaces, and, particularly, additional reflections from adjacent substrate layers also have an impact on scattering loss. This is valuable not only to WERS but also to benefit a wider range of on-chip photonics as a method for scattering loss reduction.

The study considers capturing Raman radiation from dipoles in different polarization and spatial modes. Experimental results confirm the collection of both co-polarized and cross-polarized signals from a pumped waveguide. A new Figure of Merit (FOM) is introduced to optimize WERS sensors considering multimode excitation/collection operation, waveguide thickness, and mode-dependent propagation losses. The optimized larger-core waveguides result in significant improvements in signal collection efficiency. The use of multimode waveguides and signal collection is discussed, highlighting that the spectroscopic measurements are not significantly affected when using a multimode fibre for signal collection and feeding it into the spectrometer. This relaxation of fabrication and experimental tolerances is expected to result in more sensitive, robust and cost-effective WERS sensors.

The second part, at first, developed a theoretical model based on the reciprocity principle of the grating coupler. The model aims to calculate the coupling efficiency (CE) and corresponding WGC parameters accurately and efficiently for better optimizing WERS. Initially, the theoretical model incorporates Sychugov et al.'s model, which calculates the

scattering coefficient and predicts the CE trend but lacks accuracy for practical WGC designs. To improve accuracy, modifications are made to Sychugov et al.'s model, based on several hypotheses made on WGC. The modified model demonstrates improved accuracy for CE and WGC parameter calculations, validated through comparison with numerical results. While the model performs well for thin-core WGCs, accuracy remains an issue for thicker cores. Nonetheless, the physical mechanism confirmed by the model provides valuable insights for WGC investigation, leading to the development of a more accurate and flexible model in the subsequent modelling.

The following model, built upon the verified hypotheses, accurately calculates the coupling efficiency (CE) and provides valuable physical insights into the operation of WGCs. It demonstrates good agreement with numerical simulations while offering faster computation times, making it ideal for full-parameters WGC optimization.

The theoretical model, which treats the grating as a modal field polarization source layer and considers joint loss between the grating and waveguide, is established with detailed mathematic deduction. The model reveals the impact of multiple reflections from adjacent layers in the grating region on the total CE, emphasizing the importance of including joint loss in CE calculations. The physical mechanisms uncovered by the model are often overlooked by numerical models used for WGC optimization.

Following up, the limitations imposed by multilayer interference effects on WGC CE are addressed and the principle for breaking such limits is elaborated and demonstrated with tilted gratings. The model investigates the use of tilting the grating profile to manipulate the phase of scattered light and enhance CE for thicker-cored WGCs. The theoretical calculations are validated through numerical simulations, demonstrating the potential practical value of this approach.

The effect of the grating profile is studied. The model approximates the trapezoidal shape by discrete infinitesimal rectangles and the “grating layer” as a gradient-indexed layer. The approximation has been validified with numerical simulation by precisely predicting the grating period, filling the 2% gap of inaccuracy. The CE as a function of the grating profile is theoretically calculated and compared to numerical simulations, providing insights for understanding WGC's nature and potential applications in practical fabrication.

The part on WGC modelling establishes an accurate analytical theoretical model for WGC, shedding light on the physical mechanisms and potential design improvements. The findings challenge the notion that thicker core WERS systems compromise CE, suggesting that CE can be enhanced through optimized grating profiles and tilting techniques.

In the signal processing section, a novel MNLM filter, allowing the use of information from different frames and applicable to 1D signals is first proposed and implemented. The MNLM filter has shown robustness compared to conventional neighbour-based approaches like Gaussian and median filtering. It has demonstrated good performance in both artificial signal restoration and experimental data processing, indicating its potential for automated Raman signal processing due to its effectiveness and potential of being parameter-free.

The ALS method was upgraded by incorporating the RBF neuron network for baseline extraction, effectively avoiding the bending of the baseline towards Raman peaks, especially when dealing with complex baselines and wide Raman peaks. The integration of baseline removal as a non-negative constraint in deconvolution has significantly reduced ringing artefacts, particularly in low SNR signals.

After establishing the mathematical model and explaining the algorithm implementation, the detailed steps for processing experimental data are presented. The processed experimental data is compared to the Raman spectrum obtained from expensive conventional Raman microscopes. The computational method has successfully restored fine structures and subtle Raman features, mitigating the limitations of strong background and wide point spread function (PSF) associated with the waveguide material.

Overall, the set of interdependent algorithms provides effective solutions for obtaining high-quality Raman spectra from WERS, overcoming limitations such as strong background signals and wide PSF, with minimal cost.

6.2 Future possibilities

Regarding waveguide optimization, the full potential of multimode cross-polarization Raman signal collection is currently limited by waveguide fabrication challenges, particularly the loss caused by the waveguide material. However, advancements in thin film deposition techniques are expected to overcome this limitation in the future, allowing for

the validation of our approach. Furthermore, the principles of addressing scattering loss through waveguide geometry design and manipulating interference effects with multilayer reflection to reduce losses are not limited to slab waveguides alone. Our theoretical model and experimental findings, along with our new FOM, have the potential to be extended to more complex 3D waveguide structures, such as ridge/channel waveguides. The scattering losses arising from core-cladding interface roughness in slab waveguides and sidewall roughness in 3D waveguides share theoretical similarities. Consequently, it becomes possible to mitigate scattering losses from the sidewalls by manipulating the multilayer interference effect. Similarly, in 3D waveguide-based WERS, the impact of sidewall scattering can be alleviated by widening the waveguide (reducing the lateral modal field at the sidewalls). Furthermore, the increased width of the 3D waveguide could facilitate Raman signal collection with multiple cross-polarization lateral modes, leading to enhanced efficiency in WERS. Therefore, our findings and strategies hold relevance and applicability beyond the specific context of slab waveguides.

The model developed in the second part for WGCs demonstrates high accuracy and possesses further potential for expansion. With the proven possibility that the CE of WGC is not necessarily to compromise for optimal waveguide core thickness, grating-in-coupled WERS is promising to be a flexible, compact, fault-tolerant label-free sensing method at low cost. The approach of considering grating teeth as dipoles significantly simplifies the analytical complexity and has the capability to handle more intricate structures, including non-uniform gratings and gratings with irregular grating profiles. In a more ambitious endeavour, if this modelling approach, which replaces complex optical structures with dipole layers, can be extended to a 2D scenario, it could significantly change current nano-scale photonics design. This would represent an intriguing undertaking once the model for TM modes is finalized.

For signal processing, mathematic-driven methods developed in this project have shown good performance with minimal cost to exploit the maximal potential of WERS systems. However, except for the denoising method, the other processing steps rely much on the parameter settings, and it could be improved by introducing physics-driven or data-driven methods into the signal processing method. For example, the PSF of the system can be characterized physically to reduce the error that blind deconvolution could produce due to inaccurately estimated PSF. Furthermore, data-driven methods, such as machine learning,

offer promising opportunities to develop comprehensive solutions for purifying and interpreting Raman spectra. It can be expected that, with the accumulation of standard data set for training the model, data-driven methods could produce more and more reliable accurate and robust processing. Meanwhile, it is also valuable to explore the incorporation of data-driven and mathematic driven method to exploit maximum physical potential of the instrument. In addition, the algorithms set developed in this project hold the possibility to apply to other spectroscopy, such as mid infrared absorption spectroscopy, for better sensing efficiency.

Appendix A Modal field calculation of waveguides

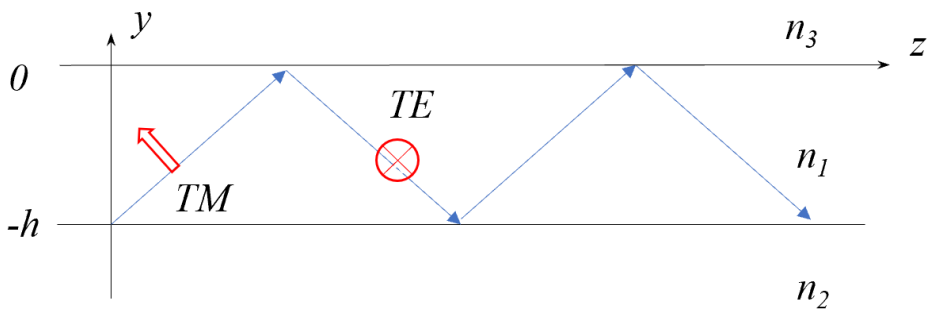

Figure A-1. Illustration of guided wave in a slab waveguide with refractive indices for cladding, core and substrate denoted as n_3, n_1 and n_2, respectively.

The electric field distribution in three-layer / multi-layer structured waveguide is calculated analytically with electromagnetic treatment. Figure A.1 illustrates a three-layer waveguide, with the refractive indices for cladding, core and substrate denoted as n_3, n_1 and n_2, respectively.

Analysing this problem electromagnetically is relatively straightforward because each segment of the dielectric structure is homogeneous. Furthermore, solutions to Maxwell's equations in homogeneous media take the form of plane waves. Consequently, our approach involves establishing plane-wave solutions for each segment and subsequently applying boundary conditions at the interfaces.

For the analysis, we are considering the propagation of monochromatic wave along the z-axis. Maxwell's equations can be expressed in the following form:

$$\nabla \times \mathrm{H} = i\omega\varepsilon_0 n(y)^2 \mathrm{E}, \nabla \times \mathrm{E} = -i\omega\mu \mathrm{H}, \tag{A.1}$$

where n is the refractive index profile. Since the whole structure is homogeneous along the z axis, solutions to the equations (A.1). can be taken as,

$$\begin{aligned} \mathrm{E}(y,t) &= \mathrm{E}_j(y)\exp[i(\omega t - \beta z)] \\ \mathrm{H}(y,t) &= \mathrm{H}_j(y)\exp[i(\omega t - \beta z)], \end{aligned} \tag{A.2}$$

where β is the z component of the wave vectors and is known as the propagation constant to be determined from Maxwell's equations. $\mathrm{E_j(y)}$ and $\mathrm{H_j(y)}$ are wavefunctions of the

guided modes, the subscript m denotes the mode number. Via applying curl operator to Eq. (A.1), H is eliminated from Eq. (A.1),

$$\nabla^2 \mathrm{E} = \mu_0 \varepsilon_0 \mathrm{n}^2 \frac{\mathrm{d}^2 \mathrm{E}}{\mathrm{d}t^2} \tag{A.3}$$

For layered dielectric structures that consist of homogeneous and isotropic materials, the wave equation can be obtained as follows,

$$\frac{d^2}{dy^2}\mathrm{E}(y) + (k_0^2 n^2 - \beta^2)\mathrm{E}(y) = 0, \tag{A.4}$$

where k_0 is the wavevector in air.

The guide modes can be classified as TE and TM modes, as shown in figure A.1. The TE modes have their electric field perpendicular to the y-z plane (plane of incidence) and thus have only the field components E_x, H_y and H_z. The TM modes have theH_x, E_y and E_z field components.

For TE modes, the electric field is written as $E_y(x,z,t) = E_j(x)\exp[i(\omega t - \beta z)]$. The solution for preceding wave equation for TE modes, which is a second order ordinary differential equation, can be easily found as,

$$\mathbf{E_j}(y) = \begin{cases} A \cdot e^{-qy} & (cladding) \\ \mathrm{B} \cdot \cos(ky) + \mathrm{C} \cdot \sin(ky) & (core) \\ \mathrm{D} \cdot \mathrm{e}^{py} & (substrate) \end{cases} \tag{A.5}$$

where A, B, C and D are amplitude coefficients that need to satisfy the boundary condition at interfaces. And

$$\begin{aligned} k &= [(k_0 n_1)^2 - \beta^2]^{\frac{1}{2}}, \\ q &= [\beta^2 - (k_0 n_3)^2]^{\frac{1}{2}} \\ p &= [\beta^2 - (k_0 n_2)^2]^{\frac{1}{2}}. \end{aligned} \tag{A.6}$$

The boundary condition requires that tangential components$\mathrm{E_x}$ and $\mathrm{H_z}$ (parallel to the interface) should be continuous at each interface.

The continuity of $\mathrm{E_y}$ results in the relation among A, B, C and D as follows,

$$\begin{aligned} \mathrm{B} &= \mathrm{A} \\ \mathrm{D} &= [\mathrm{B} \cdot \cos(kh) - C \cdot \sin(\mathrm{k}h)] \cdot \mathrm{e}^{ph} \end{aligned} \tag{A.7}$$

where h is the core thickness as shown in figure A-1. Since $\mathrm{H_z} = \left(\frac{i}{\omega\mu}\right)\left(\frac{\partial E_x}{\partial y}\right)$, the continuity of $\mathrm{H_z}$ results in

$$\begin{gathered} q \cdot \mathrm{A} = -k \cdot \mathrm{C} \\ k \cdot [\mathrm{A} \cdot \cos(kh) + \mathrm{C} \cdot \sin(kh)] = p \cdot [\mathrm{A} \cdot \cos(kh) - \mathrm{C} \cdot \sin(kh)] \end{gathered} \tag{A.8}$$

The amplitude coefficients can be eliminated after mathematical manipulation, resulting in the eigenvalue equation for TE modes:

$$\mathrm{Tan}(kh) = \frac{p+q}{k\left(1-\frac{pq}{k^2}\right)} \tag{A.9}$$

The propagation constant β can be obtained via numerically solving the eigenvalue equation. Consequently, the electric field can be expressed amplitude coefficients by substitute constant β to Eq. (A.5).

For TM modes, the magnetic field is perpendicular to the incidence plane. After similar approach, the eigenvalue equation for TM modes is found as follows,

$$\mathrm{Tan}(ht) = \frac{h(\bar{p}+\bar{q})}{h^2-\bar{p}\bar{q}}, \tag{A.10}$$

where,

$$\bar{p} = \frac{n_1^2}{n_2^2}p, \bar{q} = \frac{n_1^2}{n_3^2}q. \tag{A.11}$$

After clarifying the eigenvalue equations for three-layer waveguide and the consequent field distribution, the propagation constant and the following field distribution for waveguide with multilayer structure is specified as follows.

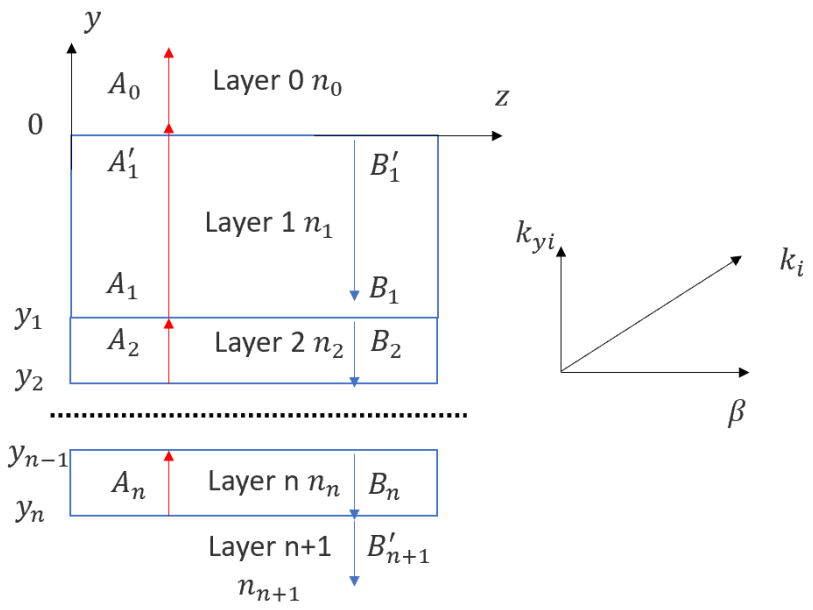

Figure A-2. Illustration of wave in multilayer structure.

A multilayer structure that guides wave along z direction is shown in figure A-2. In layer n, the electric field is written as

$$\mathrm{E_j}(y) = A_n e^{ik_{nx}(y-y_n)} + B_n e^{-ik_{lx}(y-y_n)}, \quad y_{n-1} < y < y_n, \tag{A.12}$$

where A_n and B_n are the amplitude coefficients for the up and down doing wave.

The amplitude coefficients are calculated with transfer matrix method that has been explicated in section 2.2.2.2, and expressed as

$$\begin{bmatrix} A_n \\ B_n \end{bmatrix} = \boldsymbol{M}_{nn+1} \begin{bmatrix} A_{n+1} \\ B_{n+1} \end{bmatrix}, \tag{A.13}$$

where $\boldsymbol{M}_{nn+1}$ is the transfer matrix for light travelling from layer n to layer $n+1$. Detail on calculation of the transfer matrix is explained in section 2.2.2.2, and simply expressed as,

$$\boldsymbol{M}_{nn+1} = \begin{bmatrix} e^{i\phi_n} & 0 \\ 0 & e^{-i\phi_n} \end{bmatrix} \frac{1}{t_{nn+1}} \begin{bmatrix} r_{nn+1} & 1 \\ 1 & r_{nn+1} \end{bmatrix}. \tag{A.14}$$

in which $\phi_n = k_{yn} h_n$ is the phase shift in layer n, where k_{yn} and h_n are the y component of the wave vector in layer n and thickness of layer n, respectively. The y component of the wave vector, for guided mode with propagation constant β, is given as,

$$k_{yn} = (k_n^2 - \beta^2)^{\frac{1}{2}}. \tag{A.15}$$

For the guide wave, which is different to the discussion in section 2.2.2.2 that discusses about the reflectivity and transmissivity, there is no input and the amplitude vanish at infinity so that $B_0 = A'_{n+1} = 0$. Therefore, the light travelling through the whole structure is,

$$\begin{bmatrix} A_0 \\ 0 \end{bmatrix} = \widetilde{\boldsymbol{M}} \begin{bmatrix} 0 \\ B_{n+l} \end{bmatrix}, \tag{A.16}$$

where,

$$\widetilde{\mathbf{M}} = \prod_{i=0}^{n} \mathbf{M}_{nn+1} = \begin{bmatrix} \widetilde{M}_{00} & \widetilde{M}_{01} \\ \widetilde{M}_{10} & \widetilde{M}_{11} \end{bmatrix}, \tag{A.17}$$

is the 2×2 transfer matrix for the whole structure.

Therefore, the mode condition is simply,

$$\widetilde{M}_{11} = 0. \tag{A.18}$$

With the mode condition given, the propagating constant can be found. Subsequently, the modal field in the structure can be expressed with the amplitude constant.

The amplitude constants can be normalized to the field that corresponds to a power flow, P_0, of 1 W (per unit width in the x direction) along the z axis in the mode. The normalization condition is thus given by,

$$\mathrm{P}_0 = \int_{-\infty}^{+\infty} \mathrm{S}_z \mathrm{d}y = \frac{1}{2} \int_{-\infty}^{+\infty} \mathrm{Re}(\mathrm{E} \times \mathrm{H}^*)_z \mathrm{d}y = 1. \tag{A.19}$$

for TE modes,

$$\mathrm{S}_z(y) = -\frac{1}{2} \mathrm{E}_x \mathrm{H}_y{}^* = \frac{\beta}{2\omega\mu_0} |\mathrm{E}_x(y)|^2, \tag{A.20}$$

and TM modes,

$$\mathrm{S}_z(y) = \frac{1}{2} \mathrm{H}_x{}^* \mathrm{E}_y = \frac{\beta}{2\omega \mathrm{n_i(y)}^2 \varepsilon_0} |\mathrm{H}_x(y)|^2. \tag{A.21}$$

Appendix B Data availability

Source code underlying this document along with the supporting data is uploaded to https://doi.org/10.5258/SOTON/D2771.

List of References

1. L. Jason-Moller, M. Murphy, and J. Bruno, "Overview of Biacore Systems and Their Applications," Current Protocols in Protein Science **45**(1), 19.13.1-19.13.14 (2006).

2. Y. Xu, P. Bai, X. Zhou, Y. Akimov, C. E. Png, L.-K. Ang, W. Knoll, and L. Wu, "Optical Refractive Index Sensors with Plasmonic and Photonic Structures: Promising and Inconvenient Truth," Advanced Optical Materials **7**(9), 1801433 (2019).

3. E. Benito-Peña, M. G. Valdés, B. Glahn-Martínez, and M. C. Moreno-Bondi, "Fluorescence based fiber optic and planar waveguide biosensors. A review," Analytica Chimica Acta **943**, 17–40 (2016).

4. G. S. Bumbrah and R. M. Sharma, "Raman spectroscopy – Basic principle, instrumentation and selected applications for the characterization of drugs of abuse," Egyptian Journal of Forensic Sciences **6**(3), 209–215 (2016).

5. G. Turrell and J. Corset, *Raman Microscopy: Developments and Applications* (Elsevier Science, 1996).

6. T. Itoh, V. Biju, M. Ishikawa, Y. Kikkawa, K. Hashimoto, A. Ikehata, and Y. Ozaki, "Surface-enhanced resonance Raman scattering and background light emission coupled with plasmon of single Ag nanoaggregates," Journal of Chemical Physics **124**(13), (2006).

7. R. F. Begley, A. B. Harvey, and R. L. Byer, "Coherent anti-Stokes Raman spectroscopy," Appl. Phys. Lett. **25**(7), 387–390 (1974).

8. M. Fan, G. F. S. Andrade, and A. G. Brolo, "A review on the fabrication of substrates for surface enhanced Raman spectroscopy and their applications in analytical chemistry," Anal. Chim. Acta **693**(1–2), 7–25 (2011).

9. K. Kneipp, M. Moskovits, and H. Kneipp, *Surface-Enhanced Raman Scattering: Physics and Applications* (Springer, 2006).

10. E. C. L. Ru and P. G. Etchegoin, eds., *Principles of Surface-Enhanced Raman Spectroscopy* (Elsevier, 2009).

11. S. Schlücker, "Surface-Enhanced Raman Spectroscopy: Concepts and Chemical Applications," Angew. Chem., Int. Ed. **53**(19), 4756–4795 (2014).

12. A. Otto, "The “chemical” (electronic) contribution to surface-enhanced Raman spectroscopy," J. Raman Spectrosc. **36**, 497–509 (2005).

13. A. Campion and P. Kambhampati, "Surface-enhanced Raman scattering," Chem. Soc. Rev. **27**, 241–250 (1998).

14. M. Moskovits, "Surface-enhanced Raman spectroscopy: a brief retrospective," J. Raman Spectrosc. **36**, 485–496 (2005).

15. A. Otto, I. Mrozek, H. Grabhorn, and W. Akemann, "Surface-enhanced Raman scattering," J. Phys.: Condens. Matter **4**, 1143–1212 (1992).

16. Z.-Q. Tian, B. Ren, and D.-Y. Wu, "Surface-enhanced Raman scattering: from noble to transition metals and from rough surfaces to ordered structures," J. Phys. Chem. B **106**, 9463–9483 (2002).

17. M. Moskovits and J. S. Suh, "Surface selection rules for surface-enhanced Raman spectroscopy: calculations and application to the surface-enhanced Raman spectrum of phthalazine on silver," J. Phys. Chem. **88**, 5526–5530 (1984).

18. J. S. Suh and M. Moskovits, "Surface-enhanced Raman spectroscopy of amino acids and nucleotide bases adsorbed on silver," J. Am. Chem. Soc. **108**(16), 4711–4718 (1986).

19. E. Garcia-Rico, R. A. Alvarez-Puebla, and L. Guerrini, "Direct surface-enhanced Raman scattering (SERS) spectroscopy of nucleic acids: from fundamental studies to real-life applications," Chem. Soc. Rev. **47**(13), 4909–4923 (2018).

20. S. Laing, K. Gracie, and K. Faulds, "Multiplex in vitro detection using SERS," Chem. Soc. Rev. **45**(7), 1901–1918 (2016).

21. G. Sabatte, R. Keir, M. Lawlor, M. Black, D. Graham, and W. E. Smith, "Comparison of surface-enhanced resonance Raman scattering and fluorescence for detection of a labeled antibody," Anal. Chem. **80**(7), 2351–2356 (2008).

22. K. Faulds, R. P. Barbagallo, J. T. Keer, W. E. Smith, and D. Graham, "SERRS as a more sensitive technique for the detection of labeled oligonucleotides compared to fluorescence," Analyst **129**, 567–568 (2004).

23. M. Jermyn, J. Desroches, K. Aubertin, K. St-Arnaud, W.-J. Madore, E. De Montigny, M.-C. Guiot, D. Trudel, B. C. Wilson, K. Petrecca, and F. Leblond, "A review of Raman spectroscopy advances with an emphasis on clinical translation challenges in oncology," Phys. Med. Biol. **61**(23), R370–R400 (2016).

24. G. E. Walrafen and J. Stone, "Intensification of spontaneous Raman spectra by use of liquid core optical fibers," Applied Spectroscopy **26**(6), 585–589 (1972).

25. Y. Levy, C. Imbert, J. Cipriani, S. Racine, and R. Dupeyrat, "Raman scattering of thin films as a waveguide," Optics communications **11**(1), 66–69 (1974).

26. J. Rabolt, R. Santo, and J. Swalen, "Raman spectroscopy of thin polymer films using integrated optical techniques," Applied Spectroscopy **33**(6), 549–551 (1979).

27. S. Clemmen, A. Raza, A. Dhakal, F. Peyskens, A. Z. Subramanian, P. Van Dorpe, P. Wuytens, H. Zhao, E. Ryckeboer, S. Severi, N. Le Thomas, and R. Baets, "Spectroscopic sensing with silicon nitride photonic integrated circuits," in *Integrated Optics: Devices, Materials, and Technologies XXI* (SPIE, 2017), **10106**.

28. A. Dhakal, P. C. Wuytens, F. Peyskens, K. Jans, N. L. Thomas, and R. Baets, "Nanophotonic Waveguide Enhanced Raman Spectroscopy of Biological Submonolayers," ACS Photonics **3**(11), 2141–2149 (2016).

29. S. A. Holmstrom, T. H. Stievater, D. A. Kozak, M. W. Pruessner, N. Tyndall, W. S. Rabinovich, R. Andrew McGill, and J. B. Khurgin, "Trace gas Raman spectroscopy using functionalized waveguides," Optica **3**(8), 891 (2016).

30. A. Dhakal, P. Wuytens, F. Peyskens, A. Subramanian, A. Skirtach, N. Le Thomas, and R. Baets, "Nanophotonic lab-on-a-chip Raman sensors: A sensitivity comparison with confocal Raman microscope," in *2015 International Conference on BioPhotonics (BioPhotonics)* (IEEE, 2015).

31. J. Michon, D. Kita, and J. Hu, "Sensitivity comparison of free-space and waveguide Raman for bulk sensing," J. Opt. Soc. Am. B **37**(7), 2012–2020 (2020).

32. S. Alberti, A. Datta, and J. Jagerska, "Integrated Nanophotonic Waveguide-Based Devices for IR and Raman Gas Spectroscopy," Sensors (Basel) **21**(21), (2021).

33. A. Raza, S. Clemmen, P. Wuytens, M. de Goede, A. S. K. Tong, N. Le Thomas, C. Liu, J. Suntivich, A. G. Skirtach, S. M. Garcia-Blanco, D. J. Blumenthal, J. S. Wilkinson, and R. Baets, "High index contrast photonic platforms for on-chip Raman spectroscopy," Optics Express **27**(16), 23067 (2019).

34. Z. Liu, Q. Zhao, P. Shi, B. Mitchell, H. Zhao, N. Le Thomas, D. J. Blumenthal, and R. Baets, "Tantalum Pentoxide Slot Waveguides for Waveguide Enhanced Raman Spectroscopy," in *2021 Conference on Lasers and Electro-Optics Europe & European Quantum Electronics Conference (CLEO/Europe-EQEC)* (IEEE, 2021), pp. 1–1.

35. Z. Wang, S. J. Pearce, Y.-C. Lin, M. N. Zervas, P. N. Bartlett, and J. S. Wilkinson, "Power Budget Analysis for Waveguide-Enhanced Raman Spectroscopy," Applied Spectroscopy **70**(8), 1384–1391 (2016).

36. Z. Wang, M. N. Zervas, P. N. Bartlett, and J. S. Wilkinson, "Surface and waveguide collection of Raman emission in waveguide-enhanced Raman spectroscopy," Optics Letters **41**(17), 4146 (2016).

37. A. Dhakal, A. Z. Subramanian, P. Wuytens, F. Peyskens, N. L. Thomas, and R. Baets, "Evanescent excitation and collection of spontaneous Raman spectra using silicon nitride nanophotonic waveguides," Optics Letters **39**(13), 4025 (2014).

38. C. A. Barrios, K. B. Gylfason, B. Sánchez, A. Griol, H. Sohlström, M. Holgado, and R. Casquel, "Slot-waveguide biochemical sensor," Optics Letters **32**(21), 3080 (2007).

39. J. Flueckiger, S. Schmidt, V. Donzella, A. Sherwali, D. M. Ratner, L. Chrostowski, and K. C. Cheung, "Sub-wavelength grating for enhanced ring resonator biosensor," Optics Express **24**(14), 15672 (2016).

40. J. S. Kanger, C. Otto, M. Slotboom, and J. Greve, "Waveguide Raman Spectroscopy of Thin Polymer Layers and Monolayers of Biomolecules Using High Refractive Index Waveguides," J. Phys. Chem. **100**(8), 3288–3292 (1996).

41. D. Kita, J. Michon, J. Hu, and S. G. Johnson, "Are slot and sub-wavelength grating waveguides better than strip waveguides for sensing?," Optica **5**(9), 1046–1054 (2018).

42. M. A. Ettabib, Z. Liu, M. N. Zervas, and J. S. Wilkinson, "Optimized design for grating-coupled waveguide-enhanced Raman spectroscopy," Optics Express **28**(25), 37226 (2020).

43. D. M. Kita, J. Michon, and J. Hu, "A packaged, fiber-coupled waveguide-enhanced Raman spectroscopic sensor," Optics Express **28**(10), 14963 (2020).

44. N. F. Tyndall, T. H. Stievater, D. A. Kozak, M. W. Pruessner, B. J. Roxworthy, W. S. Rabinovich, C. A. Roberts, R. A. McGill, B. L. Miller, E. Luta, and M. Z. Yates, "Figure-of-Merit Characterization of Hydrogen-Bond Acidic Sorbents for Waveguide-Enhanced Raman Spectroscopy," ACS Sens. **5**(3), 831–836 (2020).

45. N. Le Thomas, A. Dhakal, A. Raza, F. Peyskens, and R. Baets, "Impact of fundamental thermodynamic fluctuations on light propagating in photonic waveguides made of amorphous materials," Optica **5**(4), 328 (2018).

46. G. Roelkens, D. Vermeulen, D. Van Thourhout, R. Baets, S. Brision, P. Lyan, P. Gautier, and J.-M. Fédéli, "High efficiency diffractive grating couplers for interfacing a single mode optical fiber with a nanophotonic silicon-on-insulator waveguide circuit," Appl. Phys. Lett. **92**(13), 131101 (2008).

47. R. Halir, P. Cheben, S. Janz, D.-X. Xu, Í. Molina-Fernández, and J. G. Wangüemert-Pérez, "Waveguide grating coupler with subwavelength microstructures," Opt. Lett. **34**(9), 1408–1410 (2009).

48. A. Dhakal, P. Wuytens, F. Peyskens, A. Z. Subramanian, N. Le Thomas, and R. Baets, "Silicon-nitride waveguides for on-chip Raman spectroscopy," in F. Berghmans, A. G. Mignani, and P. De Moor, eds. (2014), p. 91411C.

49. M. Kawachi, "Silica waveguides on silicon and their application to integrated-optic components," Optical and Quantum Electronics **22**, 391–416 (1990).

50. K.-S. Lee and others, "Raman Microspectroscopy for Microbiology," Nature Reviews Methods Primers **1**(1), 80 (2021).

51. X. X. Han, R. S. Rodriguez, C. L. Haynes, Y. Ozaki, and B. Zhao, "Surface-enhanced Raman spectroscopy," Nature Reviews Methods Primers **1**(1), 87 (2021).

52. D. A. Long, *The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules* (Wiley Chichester, 2002).

53. J. H. Schmid, A. Delâge, B. Lamontagne, J. Lapointe, S. Janz, P. Cheben, A. Densmore, P. Waldron, D.-X. Xu, and K. P. Yap, "Interference effect in scattering loss of high-index-contrast planar waveguides caused by boundary reflections," Optics Letters **33**(13), 1479–1481 (2008).

54. L. Novotny, "Allowed and forbidden light in near-field optics I A single dipolar light source," Journal of the Optical Society of America A **14**(1), 91 (1997).

55. H. Zhao, B. Baumgartner, A. Raza, A. Skirtach, B. Lendl, and R. Baets, "Multiplex volatile organic compound Raman sensing with nanophotonic slot waveguides functionalized with a mesoporous enrichment layer," Optics Letters **45**(2), 447–450 (2020).

56. A. Dhakal, A. Z. Subramanian, P. Wuytens, F. Peyskens, N. Le Thomas, and R. Baets, "Evanescent Excitation and Collection of Spontaneous Raman Spectra Using Silicon Nitride Nanophotonic Waveguides," Optics Letters **39**(13), 4025–4028 (2014).

57. P. Lasch, "Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging," Chemometrics and Intelligent Laboratory Systems **117**, 100–114 (2012).

58. S. Li and L. Dai, "An Improved Algorithm to Remove Cosmic Spikes in Raman Spectra for Online Monitoring," Appl Spectrosc **65**(11), 1300–1306 (2011).

59. B. Hennelly and S. Barton, "Signal to noise ratio of Raman spectra of biological samples," in *Biophotonics: Photonic Solutions for Better Health Care VI*, J. Popp, V. V. Tuchin, and F. S. Pavone, eds. (SPIE, 2018), p. 160.

60. F. Ehrentreich and L. Summchen, "Spike removal and denoising of Raman spectra by wavelet transform methods," Analytical Chemistry **73**(17), 4364–4373 (2001).

61. Y. Katsumoto and Y. Ozaki, "Practical Algorithm for Reducing Convex Spike Noises on a Spectrum," Appl Spectrosc **57**(3), 317–322 (2003).

62. H. Takeuchi, S. Hashimoto, and I. Harada, "Simple and Efficient Method to Eliminate Spike Noise from Spectra Recorded on Charge-Coupled Device Detectors," Appl Spectrosc **47**(1), 129–131 (1993).

63. R. Gautam, S. Vanga, F. Ariese, and S. Umapathy, "Review of multidimensional data processing approaches for Raman and infrared spectroscopy," EPJ Techn Instrum **2**(1), 8 (2015).

64. X. Y. Zhao, G. Y. Liu, Y. T. Sui, M. Xu, and L. Tong, "Denoising method for Raman spectra with low signal-to-noise ratio based on feature extraction," Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy **250**, 119374 (2021).

65. A. Kumar and S. S. Sodhi, "Comparative analysis of gaussian filter, median filter and denoise autoenocoder," in *2020 7th International Conference on Computing for Sustainable Global Development (INDIACom)* (IEEE, 2020), pp. 45–51.

66. P. M. and K. V. Martin Clupek, "Noise reduction in Raman spectra: Finite impulse response filtration versus Savitzky–Golay smoothing," Journal of Raman Spectroscopy **38**(April), 1538–1553 (2007).

67. H. Chen, W. Xu, N. Broderick, and J. Han, "An adaptive denoising method for Raman spectroscopy based on lifting wavelet transform," Journal of Raman Spectroscopy **49**(9), 1529–1539 (2018).

68. A. R. Ferreira da Silva, "Wavelet denoising with evolutionary algorithms," Digital Signal Processing **15**(4), 382–399 (2005).

69. J. Xu, X. Yi, G. Jin, D. Peng, G. Fan, X. Xu, X. Chen, H. Yin, J. M. Cooper, and W. E. Huang, "High-Speed Diagnosis of Bacterial Pathogens at the Single Cell Level by Raman Microspectroscopy with Machine Learning Filters and Denoising Autoencoders," ACS Chem. Biol. **17**(2), 376–385 (2022).

70. A. P. Shreve, N. J. Cherepy, and R. A. Mathies, "Effective Rejection of Fluorescence Interference in Raman Spectroscopy Using a Shifted Excitation Difference Technique," Applied Spectroscopy **46**(4), 707–711 (1992).

71. J. Zhao, M. M. Carrabba, and F. S. Allen, "Automated fluorescence rejection using shifted excitation Raman difference spectroscopy," Applied Spectroscopy **56**(7), 834–845 (2002).

72. S. Guo, T. Bocklitz, and J. Popp, "Optimization of Raman-spectrum baseline correction in biological application," Analyst **141**(8), 2396–2404 (2016).

73. J. Peng, S. Peng, A. Jiang, J. Wei, C. Li, and J. Tan, "Asymmetric least squares for multiple spectra baseline correction," Analytica Chimica Acta **683**(1), 63–68 (2010).

74. Y. Cai, C. Yang, D. Xu, and W. Gui, "Baseline correction for Raman spectra using penalized spline smoothing based on vector transformation," Anal. Methods **10**(28), 3525–3533 (2018).

75. J. Zhao, H. Lui, D. I. Mclean, and H. Zeng, "Automated autofluorescence background subtraction algorithm for biomedical raman spectroscopy," Applied Spectroscopy **61**(11), 1225–1232 (2007).

76. A. K. Atakan, W. E. Blass, and D. E. Jennings, "Elimination of Baseline Variations from a Recorded Spectrum by Ultra-Low Frequency Filtering," Appl Spectrosc **34**(3), 369–372 (1980).

77. P. A. Mosier-Boss, S. H. Lieberman, and R. Newbery, "Fluorescence Rejection in Raman Spectroscopy by Shifted-Spectra, Edge Detection, and FFT Filtering Techniques," Applied Spectroscopy **49**(5), 630–638 (1995).

78. Y. Hu, T. Jiang, A. Shen, W. Li, X. Wang, and J. Hu, "A background elimination method based on wavelet transform for Raman spectra," Chemometrics and Intelligent Laboratory Systems **85**(1), 94–101 (2007).

79. Y. Xi, Y. Li, Z. Duan, and Y. Lu, "A novel pre-processing algorithm based on the wavelet transform for Raman spectrum," Applied spectroscopy **72**(12), 1752–1763 (2018).

80. G. Schulze, A. Jirasek, M. M. L. Yu, A. Lim, F. B. Turner, and M. W. Blades, "Investigation of Selected Baseline Removal Techniques as Candidates for Automated Implementation," Applied spectroscopy **59**(5), 545–574 (2005).

81. A. Jirasek, G. Schulze, M. M. L. Yu, M. W. Blades, and R. F. B. Turner, "Accuracy and Precision of Manual Baseline Determination," Appl Spectrosc **58**(12), 1488–1499 (2004).

82. V. A. Lórenz-Fonfría and E. Padrós, "Maximum Entropy Deconvolution of Infrared Spectra: Use of a Novel Entropy Expression without Sign Restriction," Appl Spectrosc **59**(4), 474–486 (2005).

83. T. F. Chan and C. K. Wong, "Total variation blind deconvolution," IEEE Transactions on Image Processing **7**(3), 370–375 (1998).

84. B. L. K. Davey, R. G. Lane, and R. H. T. Bates, "Blind deconvolution of noisy complex-valued image," Optics Communications **69**(5–6), 353–356 (1989).

85. T. Liu, H. Liu, Z. Zhang, and S. Liu, "Nonlocal low-rank-based blind deconvolution of Raman spectroscopy for automatic target recognition," Applied Optics **57**(22), 6461 (2018).

86. H. Liu, Z. Zhang, J. Sun, and S. Liu, "Blind spectral deconvolution algorithm for Raman spectrum with Poisson noise," Photon. Res. **2**(6), 168 (2014).

87. H. Liu, L. Yan, Y. Chang, H. Fang, and T. Zhang, "Spectral Deconvolution and Feature Extraction With Robust Adaptive Tikhonov Regularization," IEEE Trans. Instrum. Meas. **62**(2), 315–327 (2013).

88. S. C. Warren-Smith, S. Afshar, and T. M. Monro, "Fluorescence-based sensing with optical nanowires: a generalized model and experimental validation," Optics Express **18**(9), 9474 (2010).

89. L. Tong, R. R. Gattass, J. B. Ashcom, S. He, J. Lou, M. Shen, I. Maxwell, and E. Mazur, "Subwavelength-diameter silica wires for low-loss optical wave guiding," Nature **426**(6968), 816–819 (2003).

90. G. Zhai and L. Tong, "Roughness-induced radiation losses in optical micro or nanofibers," Optics Express **15**(21), 13805 (2007).

91. P. K. Tien, "Light waves in thin films and integrated optics," Applied Optics **10**(11), 2395–2413 (1971).

92. D. Marcuse and R. M. Derosier, "Mode Conversion Caused by Diameter Changes of a Round Dielectric Waveguide," Bell System Technical Journal **48**(10), 3217–3232 (1969).

93. D. Marcuse, "Mode Conversion Caused by Surface Imperfections of a Dielectric Slab Waveguide," Bell System Technical Journal **48**(10), 3187–3215 (1969).

94. A. Snyder, "Radiation on losses due to variations of radius dielectric or optical fibers," IEEE Transactions on Microwave Theory and Techniques **18**(9), 608–615 (1970).

95. S. G. Johnson, M. L. Povinelli, M. Soljačić, A. Karalis, S. Jacobs, and J. D. Joannopoulos, "Roughness losses and volume-current methods in photonic-crystal waveguides," Applied Physics B: Lasers and Optics **81**(2–3), 283–293 (2005).

96. F. P. Payne and J. P. R. Lacey, "A theoretical analysis of scattering loss from planar optical waveguides," Optical and Quantum Electronics **26**, 977–986 (1994).

97. J. P. R. R. Lacey and F. P. Payne, "Radiation loss from planar waveguides with random wall imperfections," IEE proceedings. Part J, Optoelectronics **137**(4), 282–288 (1990).

98. A. Dhakal, A. Subramanian, R. Baets, and N. Le Thomas, "The role of index contrast in the efficiency of absorption and emission of a luminescent particle near a slab waveguide," in *16th European Conference on Integrated Optics (ECIO-2012)* (2012).

99. R. Baets, A. Z. Subramanian, A. Dhakal, S. K. Selvaraja, K. Komorowska, F. Peyskens, E. Ryckeboer, N. Yebo, G. Roelkens, and N. Le Thomas, "Spectroscopy-on-chip applications of silicon photonics," Integrated Optics: Devices, Materials, and Technologies XVII **8627**(March 2013), 86270I (2013).

100. Y. C. Jun, R. M. Briggs, H. A. Atwater, and M. L. Brongersma, "Broadband enhancement of light emission in silicon slot waveguides," Optics Express **17**(9), 7479–7490 (2009).

101. A. Dhakal, A. Raza, F. Peyskens, A. Z. Subramanian, S. Clemmen, N. Le Thomas, and R. Baets, "Efficiency of evanescent excitation and collection of spontaneous Raman scattering near high index contrast channel waveguides," Optics Express **23**(21), 27391 (2015).

102. A. W. Snyder and J. Love, *Optical Waveguide Theory* (Springer Science & Business Media, 2012).

103. S. J. Byrnes, "Multilayer optical calculations," arXiv:1603.02720 1–20 (2016).

104. P. Yeh, *Optical Waves in Layered Media*, Wiley Series in Pure and Applied Optics (Wiley, 2005).

105. M. A. Ettabib, A. Marti, Z. Liu, B. M. Bowden, M. N. Zervas, P. N. Bartlett, and J. S. Wilkinson, "Waveguide Enhanced Raman Spectroscopy for Biosensing: A Review," ACS Sensors **6**(6), 2025–2045 (2021).

106. J.-C. Zhou, D. Luo, Y. Li, and L. Zheng, "Effect of sputtering pressure and rapid thermal annealing on optical properties of Ta_2O_5 thin films," Transactions of Nonferrous Metals Society of China **19**(2), 359–363 (2009).

107. C. Mannequin, T. Tsuruoka, T. Hasegawa, and M. Aono, "Identification and roles of nonstoichiometric oxygen in amorphous Ta_2O_5 thin films deposited by electron beam and sputtering processes," Applied Surface Science **385**, 426–435 (2016).

108. T. W. Huang, H. Y. Lee, Y. W. Hsieh, and C. H. Lee, "X-ray study of the surface morphology of crystalline and amorphous tantalum peroxide thin films prepared by RF magnetron sputtering," Journal of Crystal Growth **237–239**(1 4 I), 492–495 (2002).

109. K. Lehovec, "Lattice structure of β-Ta_2O_5," Journal of the Less Common Metals **7**(6), 397–410 (1964).

110. F. Ladouceur, J. D. Love, and T. J. Senden, "Effect of side wall roughness in buried channel waveguides," IEE Proceedings - Optoelectronics **141**(4), 242–248 (1994).

111. J. Staub, V. Scheuer, and T. Tschudi, "Investigation Of Dielectric Laser Mirrors Using Transmission Electron Microscopy," in *Thin Films in Optics* (SPIE, 1990), **1125**, pp. 61–67.

112. A. Duparré, E. Welsch, H. G. Walther, H. J. Kühn, and G. Schirmer, "Optical losses of sputtered Ta2O5 films," J. Phys. France **48**(7), 1155–1159 (1987).

113. M. Vaidulych, P. Pleskunov, J. Kratochvíl, H. Mašková, P. Kočová, D. Nikitin, J. Hanuš, O. Kylián, J. Štěrba, H. Biederman, and A. Choukourov, "Convex vs concave surface nano-curvature of Ta2O5 thin films for tailoring the osteoblast adhesion," Surface and Coatings Technology **393**, 125805 (2020).

114. M. A. Duguay, Y. Kokubun, T. L. Koch, and L. Pfeiffer, "Antiresonant reflecting optical waveguides in SiO_2-Si multilayer structures," Appl. Phys. Lett. **49**(1), 13–15 (1986).

115. S. Février, B. Beaudou, and P. Viale, "Understanding origin of loss in large pitch hollow-core photonic crystal fibers and their design simplification," Opt. Express **18**(5), 5142 (2010).

116. N. M. Litchinitser, A. K. Abeeluck, C. Headley, and B. J. Eggleton, "Antiresonant reflecting photonic crystal optical waveguides," Opt. Lett. **27**(18), 1592 (2002).

117. K. Reynkens, S. Clemmen, A. Raza, H. Zhao, J. S.-D. Peñaranda, C. Detavernier, and R. Baets, "Mitigation of photon background in nanoplasmonic all-on-chip Raman sensors," Optics Express **28**(22), 33564 (2020).

118. Y. Li, H. Zhao, R. Baets, and J.-J. He, "A novel Raman spectrum detection system based on an integrated optical biosensor," in *Advanced Sensor Systems and Applications X* (SPIE, 2020), **11554**, pp. 30–36.

119. M. Ettabib, B. Bowden, Z. Liu, A. Marti, G. Churchill, J. Gates, M. Zervas, P. N. Bartlett, and J. S. Wilkinson, "Grating-Incoupled Waveguide-Enhanced Raman Sensor," PLOS ONE (in press).

120. Lord Rayleigh, "On the Dynamical Theory of Gratings," Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **79**(532), 399–416 (1907).

121. B. Wohlfeil, L. Zimmermann, and K. Petermann, "Optimization of fiber grating couplers on SOI using advanced search algorithms," Optics Letters **39**(11), 3201 (2014).

122. L. Cheng, S. Mao, Z. Li, Y. Han, and H. Fu, "Grating Couplers on Silicon Photonics: Design Principles, Emerging Trends and Practical Issues," Micromachines **11**(7), 666 (2020).

123. V. A. Sychugov, A. V. Tishchenko, N. M. Lyndin, and O. Parriaux, "Waveguide coupling gratings for high-sensitivity biochemical sensors," Sensors and Actuators, B: Chemical **39**(1–3), 360–364 (1997).

124. T. Tamir, *Integrated Optics (Topics in Applied Physics; v. 7)* (Springer-Verlag Berlin Heidelberg, 1975).

125. D. G. Dalgoutte and C. D. W. Wilkinson, "Thin grating couplers for integrated optics: an experimental and theoretical study," Applied Optics **14**(12), 2983 (1975).

126. N. M. Lyndin, O. Parriaux, and V. A. Sychugov, "Waveguide excitation by a Gaussian beam and a finite size grating," Sensors and Actuators, B: Chemical **41**(1–3), 23–29 (1997).

127. K. Edee, J. P. Plumey, and J. Chandezon, "On the Rayleigh-Fourier method and the Chandezon method: Comparative study," Optics Communications **286**(1), 34–41 (2013).

128. A. A. Zlenko, V. A. Kiselev, A. M. Prokhorov, A. A. Spikhal'skii, and V. A. Sychugov, "Emission of Surface Light Waves From a Corrugated Part of a Thin-Film Waveguide.," Sov J Quantum Electron **4**(7), 839–842 (1975).

129. V. A. Kiselev, "Diffraction coupling of radiation into a thin-film waveguide," Sov J Quantum Electron **4**(8), 872–875 (1975).

130. V. A. Sychugov and A. V. Tishchenko, "Light Emission From a Corrugated Dielectric Waveguide.," Soviet journal of quantum electronics **10**(2), 186–189 (1980).

131. A. M. Prokhorov, A. A. Spikhal'skii, and V. A. Sychugov, "Emission of E waves from a corrugated section of a waveguide," Sov J Quantum Electron **6**(10), 1211–1213 (1976).

132. T. Tamir and S. T. Peng, "Analysis and Design of Grating Couplers," Applied Physics **21**(4), 410 (1980).

133. D. Gloge, "Offset and Tilt Loss in Optical Fiber Splices," Bell System Technical Journal **55**(7), 905–916 (1976).

134. K. Horiguchi, S. Kobayashi, and O. Sugihara, "Calculation Model for Multimode Fiber Connection Using Measured Near- and Far-Field Patterns," IEEE Photon. Technol. Lett. **33**(6), 285–288 (2021).

135. W. van Etten, W. Lambo, and P. Simons, "Loss in multimode fiber connections with a gape," Appl. Opt. **24**(7), 970 (1985).

136. R. P. Brent, *Algorithms for Minimization without Derivatives* (Courier Corporation, 2013).

137. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," Nature Methods **17**, 261–272 (2020).

138. R. N. Bracewell and R. N. Bracewell, *The Fourier Transform and Its Applications* (McGraw-Hill New York, 1986), **31999**.

139. J. E. Sipe, "New Green-function formalism for surface optics," Journal of the Optical Society of America B **4**(4), 481 (1987).

140. T. W. Ang, G. T. Reed, A. Vonsovici, A. G. Evans, P. R. Routley, and M. R. Josey, "Highly efficient unibond silicon-on-insulator blazed grating couplers," Applied Physics Letters **77**(26), 4214–4216 (2000).

141. M. Fan, M. A. Popović, and F. X. Kärtner, "High directivity, vertical fiber-to-chip coupler with anisotropically radiating grating teeth," in *Conference on Lasers and Electro-Optics (CLEO)* (Optical Society of America, 2007), p. CTuDD3.

142. Y. Chen, R. Halir, Í. Molina-Fernández, P. Cheben, and J.-J. He, "High-efficiency apodized-imaging chip-fiber grating coupler for silicon nitride waveguides," Opt. Lett. **41**(21), 5059 (2016).

143. M. Kwon, "A Numerically Stable Analysis Method for Complex Multilayer Waveguides Based on Modified Transfer-Matrix Equations," **27**(20), 4407–4414 (2009).

144. Bin Wang, Jianhua Jiang, and G. P. Nordin, "Embedded slanted grating for vertical coupling between fibers and silicon-on-insulator planar waveguides," IEEE Photon. Technol. Lett. **17**(9), 1884–1886 (2005).

145. J. Schrauwen, F. Van Laere, D. Van Thourhout, and R. Baets, "Focused-Ion-Beam Fabrication of Slanted Grating Couplers in Silicon-on-Insulator Waveguides," IEEE Photon. Technol. Lett. **19**(11), 816–818 (2007).

146. A. Buades, B. Coll, and J.-M. Morel, "A Non-Local Algorithm for Image Denoising," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (IEEE, 2005), **2**, pp. 60–65.

147. J. Froment, "Parameter-Free Fast Pixelwise Non-Local Means Denoising," Image Processing On Line **4**, 300–326 (2014).

148. P. Jain and V. Tyagi, "A survey of edge-preserving image denoising methods," Inf Syst Front **18**(1), 159–170 (2016).

149. Y. Lv, Q. Zhu, and R. Yuan, "Fault Diagnosis of Rolling Bearing Based on Fast Nonlocal Means and Envelop Spectrum," Sensors **15**(1), 1182–1198 (2015).

150. J. Laumer and S. K. O'Leary, "A root-mean-square-error analysis of two-peak Gaussian and Lorentzian fittings of thin-film carbon Raman spectral data," Journal of Applied Physics **126**(4), 045706 (2019).

151. A. Buades, B. Coll, and J.-M. Morel, "Non-Local Means Denoising," in *Image Processing On Line* (2011), **1**, pp. 208–212.

152. P. H. C. Eilers, "Parametric Time Warping," Analytical Chemistry **76**(2), 404–411 (2004).

153. M. J. L. Orr, "Introduction to radial basis function networks," University of Edinburgh (1996).

154. S. A. Billings, H.-L. Wei, and M. A. Balikhin, "Generalized multiscale radial basis function networks," Neural Networks **20**(10), 1081–1094 (2007).

155. R. Vio, J. Bardsley, M. Donatelli, and W. Wamsteker, "Dealing with edge effects in least-squares image deconvolution problems," A&A **442**(1), 397–403 (2005).

156. K. Chandrawansa, F. H. Ruymgaart, and A. Van Rooij, "Controlling the Gibbs phenomenon in noisy deconvolution of irregular multivariable input signals," Journal of Applied Mathematics and Stochastic Analysis **13**(1), 1–14 (2000).

157. P. H. C. Eilers, "A Perfect Smoother," Anal. Chem. **75**(14), 3631–3636 (2003).