RAMAN SPECTROSCOPY FOR EXTRACELLULAR VESICLE STUDY

WOOJE LEE

RAMAN SPECTROSCOPY FOR EXTRACELLULAR VESICLE STUDY

Wooje Lee

Graduation Committee:

Chairman / secretary:

Prof. dr. K.J. Boller

University of Twente

Supervisor:

Prof. dr. H.L. Offerhaus

University of Twente

Committee Members:

Prof. dr. A.G. Ryder

National University of Ireland - Galway

Prof. dr. A.G.J.M. van Leeuwen

University of Amsterdam

Prof. dr. L.W.M.M. Terstappen

University of Twente

Prof. dr. P.W.H. Pinkse

University of Twente

Research described in this thesis was carried out at the Optical Sciences (OS) group within the Faculty of Science and Technology, and the MESA+ Institute for Nanotechnology, University of Twente, Enschede, The Netherlands.

This work was financially supported by the Netherlands Organization for Scientific Research (NWO) domain Applied and Engineering Sciences (TTW) under the Cancer-ID program (project number 14197).

UNIVERSITY
OF TWENTE.

MESA+
INSTITUTE

Cover design: Pixelated image of Waveguide Raman setup.

Photo and design by Wooje Lee

Printed by: Gildeprint, Enschede, the Netherlands

ISBN: 978-90-365-4951-6

DOI: <https://doi.org/10.3990/1.9789036549516>

© 2020 Wooje Lee, The Netherlands. All rights reserved.

No parts of this thesis may be reproduced, stored in a retrieval system or transmitted in any form or by any means without permission of the author. Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd, in enige vorm of op enige wijze, zonder voorafgaande schriftelijke toestemming van de auteur.

RAMAN SPECTROSCOPY FOR EXTRACELLULAR VESICLE STUDY

DISSERTATION

to obtain
the degree of doctor at the Universiteit Twente,
on the authority of the rector magnificus,
Prof. dr. T.T.M. Palstra,
on account of the decision of the graduation committee
to be publicly defended
on Thursday the 15th of October 2020 at 10.45

by

Wooje Lee

born on the 28th of August 1988
in Bucheon-si, Republic of Korea

This dissertation has been approved by:

Supervisor

Prof. dr. H.L. Offerhaus (Promotor)

This book is dedicated
to my wife, Jinhwa Jang,
and my two kids, Philip and Hannah
with love.

Table of Contents

CHAPTER 1 Introduction.....	1
1.1 Motivation and Scientific Challenge	1
1.2 Thesis outline	3
 CHAPTER 2 Theoretical Background	 9
2.1 Extracellular Vesicles.....	10
2.1.1 Biogenesis of Vesicle.....	11
2.1.2 Clinical Relevance of EVs	12
2.2 Raman effect and Raman Spectroscopy	13
2.2.1 Raman Theory	14
2.2.2 Molecular Vibration and Selection Rule.....	17
2.2.3 Advantages and Application of Raman Spectroscopy	20
2.3 Optical Trapping.....	21
 CHAPTER 3 Characterization and Classification of Extracellular Vesicles using Raman Spectroscopy and Principal Component Analysis.....	 29
3.1 Introduction	30
3.2 Experiments.....	31
3.2.1 Preparation of EVs	31

3.2.2	Sample validation	34
3.2.3	Raman setup	37
3.2.4	Raman spectral data acquisition	37
3.2.5	Data processing and principal component analysis	39
3.3	Result and discussion	40
3.4	Conclusion.....	43

CHAPTER 4 Classifying Raman Spectra of Extracellular Vesicles based on Convolutional Neural Networks for Prostate Cancer Detection..... 47

4.1	introduction.....	48
4.2	Introduction to Neural Network.....	50
4.3	Convolutional Neural Network.....	51
4.3.1	Convolution Layer	53
4.3.2	Feedforward Neural Network.....	54
4.4	Raman spectral signature collection	58
4.5	Result and Discussion	60
4.6	Conclusion.....	63

CHAPTER 5 Waveguide Raman Spectroscopy: Different Waveguide Materials and their Intrinsic Background..... 69

5.1	Introduction	70
5.2	Experimental setup	72
5.2.1	Wavenumber Calibration.....	75
5.3	Waveguides Design Consideration for on-chip Raman Spectroscopy ..	78
5.4	Summary of Waveguide Fabrication	81

5.5	Results and discussion.....	84
5.6	Conclusion and Future Work	93

CHAPTER 6 Micro Ring Resonator Enhanced Waveguide Raman Spectroscopy..... 101

6.1	Introduction	102
6.2	Ring Resonator	102
6.2.1	Spectral Characteristics of a Micro-Ring Resonator.....	106
6.3	Design consideration.....	108
6.4	Mask Design of Micro Ring Resonators.....	111
6.5	Fabrication of Micro Ring Resonator	112
6.6	Result and discussion	115
6.7	Conclusion and Future works	121

CHAPTER 7 Summary..... 125

HOOFDSTUK 8 Samenvatting / Dutch Summary 129

APPENDIX A 133

A.1	The Python script coded for realization of the CNN.....	133
-----	---	-----

APPENDIX B..... 143

B.1	The Matlab script coded for wavenumber calibration.....	143
B.1.1	Main script: "Pixel2Wvn.m"	143
B.1.2	Toluene peak finding function: "ToluenePeakFinder.m"	147

B.1.3	ArHg peak finding function: “ArHgPeakFinder.m”	149
APPENDIX C		153
C.1	Mask Design	153
C.2	Device fabrication process	154
Scientific Output		155
Acknowledgements		157

Introduction

1.1 Motivation and Scientific Challenge

Almost every cell releases tiny particles into their extracellular environment[1-3]: the particles are known as extracellular vesicles (EVs). The particles have a spherical shape, and their size ranges from 30 nm to 1 μm [1,4]. The size of EVs in comparison to other micro-organism is shown in Figure 1-1. It has been demonstrated that cells use EVs for intercellular communication, waste control, and disease metastasis[1,5]. Although the first cell-derived vesicles were discovered in 1940, research on vesicles was very limited due to the lack of detection techniques for nanoparticles. By leveraging advanced detection techniques, the significance of EVs has gained attention since the early 2000s.

EVs are presented at concentration exceeding 10^{10} particles/ml in body fluids such as blood, saliva, and urine[1,4,6]. The particles transport biomolecules, such as protein, RNA, and DNA. Since the EVs originate from cells, the contents of EVs are dependent on their cellular origin. Therefore certain EVs include information related to diseases such as cancer, allergies, cardiovascular and autoimmune diseases, and investigating EVs' cellular origin/cargo is useful as diagnosis and for monitoring the prognosis of therapy[7,8]. However, current state-of-the-art techniques for EV characterization are still insufficient in terms of sensitivity. This is a significant bottleneck of EVs research and the application of EVs as clinical biomarkers.

The aim of this research project is the characterization of EVs using vibrational spectroscopy to study the contents and cellular origin of different EVs subtypes. Of the various vibrational spectroscopic techniques, Raman spectroscopy will be used for this study[9,10], which is a nondestructive and non-labeling technique. As the term 'vibrational spectroscopy' implies, Raman spectroscopy provides molecular vibration information.

Analyzing molecular vibrations not only reveals the chemical composition of the specimen but also allows for a quantitative study, simple comparison between samples, and detection of specific molecules in samples. Raman spectroscopy has proven to be a useful tool for many different applications: material science, biomedical science, and real-life applications such as forensics. Although Raman spectroscopy is a powerful and straight forward technique, the ability of a conventional Raman microscope is limited by the diffraction limit. In a free-space optical system, the diffraction limit sets a lower limit on the total probed volume and, therefore, a limit on the surface-to-volume ratio when studying nanoparticles.

Figure 1-1 Size comparison among various bio-products. The term ‘EVs’ comprises exosomes and microvesicles. The size of EVs ranges from 300 nm up to 1 μm. The small size of EVs makes difficult to study with conventional optical microscopic techniques.

Each cell contains different biomolecules depending on the presence of disease, the function of the cell, and the location in the body[11-13]. However, many biomolecules are derived from similar building blocks, such as amino acids or nucleic acids, and related to the general functioning of the cell so that that difference can be subtle. These small differences can be a significant challenge in studying cells using spectroscopic techniques.

EVs are fairly small. Small variations of the cell are transferred to small differences in EVs. Due to these hallmarks of EVs, differentiating EVs based on Raman spectrum requires not

only high signal to noise ratio (SNR) Raman spectra but also a huge effort to analyze Raman spectrum, which can lead low-throughput of the analysis.

This thesis will provide several ways to improve the current Raman technique for EV research; algorithmic analysis and use of the evanescent field for Raman spectroscopy. Firstly, we will discuss two types of Raman spectra classifiers based on Principal Component Analysis (PCA) and Neural Network, especially Convolutional Neural Network (CNN). The automated algorithmic analysis methods can be a solution to enhance the throughput of Raman analysis and create an objective classification. Alongside this approach, we will discuss integrated optics for on-chip Raman spectroscopy, called waveguide Raman spectroscopy. The evanescent wave propagates outside the waveguide and decays exponentially from the interface. Waveguide Raman uses the evanescent field for analyte excitation and collection of scattered photons. Although the resolution of the system cannot be better than the diffraction limit of the system, the use of the evanescent field allows one to probe a shallow layer above the waveguide surface. This increases the surface to volume ration, and therefore, it is expected to achieve a high SNR Raman signal of the specimen from the waveguide Raman chip.

1.2 Thesis outline

Chapter 2 briefly introduces the theoretical backgrounds of the extracellular vesicles and Raman scattering, which are essential topics for developing the discussion.

Chapter 3 describes a Raman spectrum classifier based on PCA. For this experiment, we isolated EVs from two blood cells (red blood cells and platelets) and two cancer cell lines (PC3 and LNCaP). The isolation methods of each EVs subtype are described. Four EVs subtypes (i.e., red blood cell-, platelet-, PC3-, LNCaP-derived EVs) were captured and measured by Raman optical tweezers: Raman optical tweezers is a combined technique of Raman with optical tweezers. Firstly, single or multiple particles were captured in the laser focus. Raman spectrum of the particle(s) was then recorded. The Raman spectra of the EVs subtypes were classified by cellular origin based on the PCA classifier.

Chapter 4 describes an automated classification algorithm based on a Convolutional Neural Network (CNN). Though the classifier based on PCA shows high classification accuracy, the model requires background correction prior to the analysis: this background presence and rejection can bias the result of PCA. In addition, the SNR of the Raman spectra of EVs is not so good. Thus, the Raman spectra include backgrounds of

surroundings (i.e., sample suspension and sample container). Since PCA projects the original data onto a new axis based on input values, backgrounds can bias the PCA results. In this chapter, Raman spectra of EVs were collected in the same way as described in Chapter 3. A classifier based on a supervised machine learning algorithm, namely Convolutional Neural Network, will be described. CNN was modified for 1-dimensional spectral data and trained on the Raman spectra of EVs to create a classification model. The performance of PCA and CNN based classifiers will be compared.

Chapter 5 demonstrates waveguide Raman (WGR) spectroscopy. Raman optical tweezers allow recording of Raman spectra of EVs. However, the contribution of the surrounding is often more prominent than the contribution of EVs, and the size of EVs makes it challenging to achieve a high SNR Raman signal. WGR can achieve higher SNR Raman spectra of EVs. In a WGR system, the evanescent field of an optical waveguide is used to probe an analyte on top of the waveguide. Since the evanescent field decays exponentially from the waveguide surface, it probes only a shallow area at the waveguide interface; the probing depth can be influenced by the design and the fabrication of the waveguide. Samples can be enriched on the waveguide by means of functionalizing the waveguide surface. WGR has huge potential to be used for nanostructure investigation. However, there is a major bottleneck of this technique, which is the intrinsic background of waveguide material. Therefore, in this chapter, the inherent background of aluminum oxide, silicon nitride, and titanium oxide (Al_2O_3 , Si_3N_4 , and TiO_2 , respectively) will be discussed. An analyte, toluene, will be measured to prove the WGR concept and compare the performance of the different waveguide materials.

In **chapter 6**, I will discuss a way to improve the SNR of a WGR system. We suggest a microring resonator (MRR) to enhance the WGR signal. However, the SNR of the WGR signal is limited because of the intrinsic Raman or fluorescent background of the materials. In this chapter, an optical microring resonator (MRR) will be discussed as a solution. The MRR is a special type of waveguide having a closed-loop waveguide (ring) adjacent to a straight waveguide (bus waveguide). When the incident light meets certain conditions, the light travels in the ring continuously. This phenomenon is called resonance. By using resonance, the SNR of the WGR signal can be enhanced. Thus, this chapter will discuss some theoretical background of the MRR, design, and fabrication and will demonstrate the MRR enhanced waveguide Raman spectroscopy (RE-WGR) using Al_2O_3 MRRs.

Chapter 7 will summarize and conclude the thesis. Following **Appendices** will provide more details on the software and PDMS microfluidic device fabrication process for those who want to reproduce part of this research

Bibliography

- [1] E. van der Pol, A. N. Böing, P. Harrison, A. Sturk, R. Nieuwland, "Classification, functions, and clinical relevance of extracellular vesicles", *Pharmacological Reviews*, **2012**; *64*, 676.
- [2] M. Verma, T. K. Lam, E. Hebert, R. L. Divi, "Extracellular vesicles: potential applications in cancer diagnosis, prognosis, and epidemiology", *BMC Clinical Pathology*, **2015**; *15*, 6.
- [3] L. Margolis, Y. Sadovsky, "The biology of extracellular vesicles: The known unknowns", *PLoS Biology*, **2019**; *17*, e3000363.
- [4] N. Arraud, R. Linares, S. Tan, C. Gounou, J. M. Pasquet, S. Mornet, A. R. Brisson, "Extracellular vesicles from blood plasma: determination of their morphology, size, phenotype and concentration", *Journal of Thrombosis and Haemostasis*, **2014**; *12*, 614.
- [5] H. Kimura, H. Kato, A. Faried, M. Sohda, M. Nakajima, Y. Fukai, T. Miyazaki, N. Masuda, M. Fukuchi, H. Kuwano, "Prognostic significance of EpCAM expression in human esophageal cancer", *International Journal of Oncology*, **2007**; *30*, 171.
- [6] E. van der Pol, Ph.D. Dissertation, "Detection of extracellular vesicles: size does matter", *Universiteit van Amsterdam*, **2015**.
- [7] J. Skog, T. Würdinger, S. Van Rijn, D. H. Meijer, L. Gainche, W. T. Curry Jr, B. S. Carter, A. M. Krichevsky, X. O. Breakefield, "Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers", *Nature Cell Biology*, **2008**; *10*, 1470.
- [8] K. Al-Nedawi, B. Meehan, J. Rak, "Microvesicles: messengers and mediators of tumor progression", *Cell cycle*, **2009**; *8*, 2014.
- [9] C. V. Raman, "A new radiation", *Indian Journal of Physics*, **1928**; *2*, 387.
- [10] C. V. Raman, K. S. Krishnan, "A new type of secondary radiation", *Nature*, **1928**; *121*, 501.
- [11] W. A. Osta, Y. Chen, K. Mikhitarian, M. Mitas, M. Salem, Y. A. Hannun, D. J. Cole, W. E. Gillanders, "EpCAM is overexpressed in breast cancer and is a potential target for breast cancer gene therapy", *Cancer Research*, **2004**; *64*, 5818.

- [12] M. Munz, P. A. Baeuerle, O. Gires, "The emerging role of EpCAM in cancer and stem cell signaling", *Cancer Research*, **2009**; 69, 5627.

- [13] P. Baeuerle, O. Gires, "EpCAM (CD326) finding its role in cancer", *British Journal of Cancer*, **2007**; 96, 417.

Theoretical Backgrounds

The main topic of this thesis is ‘characterization of extracellular vesicles using Raman spectroscopy.’ Prior to the main discussion, some understanding of EVs and Raman spectroscopy is required. Therefore, this chapter will briefly explain the biology of extracellular vesicles, the basics of Raman spectroscopy and optical trapping.

This chapter consists of three subchapters. First, general knowledge about EVs and the formation of EVs will be covered, and the clinical relevance of EVs will also be discussed with some examples. In the second section, the theory of the Raman effect will be discussed to understand what Raman scattering is and why it happens. Lastly, optical trapping will be explained.

2.1 Extracellular Vesicles

EVs are ultra-small particles produced and released by almost all cells[1-4]. The first cell-derived vesicles and their function were reported by Chargaff and West in the 1940s[5]. Initially, vesicles were understood to play an important role in the coagulation of blood. Although the existence and function of EVs have been recognized since the 1940s, many of the details of EVs have been under a veil of mystery owing to a lack of techniques to explore the sub-micron scale. Recently, the field of EVs is rapidly growing on the strength of the advanced nanotechnologies for EVs isolation and detection.

Cells release EVs into their microenvironment. The released particle travels around the body through all kinds of body fluids, for example, blood, urine, and saliva. These body fluids are known to contain $10^6 \sim 10^{12}$ particles/ml[2,6,7]. An EV is a spherically shaped particle with phospholipid bilayer shell as if cells have. The diameter of EVs typically ranges from 30 nm to 1 μ m. The smallest EVs are about 100-fold smaller than the smallest cell[6]. Figure 2-1 describes the structure of EVs and shows a TEM image of a prostate cancer cell line-derived (PC3) EV.

According to recent studies, EVs have specialized functions and play an important role in trash disposal, transfer of functional biomolecules, intercellular signaling, disease metastasis, and molecular recycling. EVs carry cargos of nucleic acids, metabolites, proteins, and organelle such as mitochondria[8]. These bioactive compounds can vary depending on their cellular origin[2]. Therefore, EVs have considerable potential to be used as disease biomarkers, therapeutic aid, and monitoring[9].

Figure 2-1 Structure of extracellular vesicles (A) and transmission electron microscope image of PC3 derived EVs (B).

2.1.1 Biogenesis of Vesicle

There is a wide variety of terminologies that describe the vesicles: exosomes, exosome-like vesicles, microvesicles, dexosomes, oncosomes, endosomes, and texosomes[10]. The terminology has not been standardized yet, and the differences between the various types of vesicles are still ambiguous. For example, a particle called “exosomes” by some is called “microvesicles” by others[9,11]; nevertheless, vesicles are isolated using the same method. The term microvesicles also arouse a belief that the microvesicles are much bigger than exosomes, although some of them are in the same size range; exosomes range from 30 to 150 nm, and microvesicles range from 50 to 1000 nm[2,12-14]. A terminology “extracellular vesicles” has been suggested by the International Society for Extracellular Vesicles (ISEV) as a standard and collective term for the entire population of cell derived-vesicles[15]. The term EVs aims to prevent unnecessary confusion.

It is important to recognize the fact that there are many different types of vesicles to use EVs as a biomarker[16]. In this thesis, cell derived-vesicles are classified based on their mode of release (e.g., biogenic pathway). Cells produce EVs through the outward budding of plasma membrane or endosomal pathway: exosomes, in general, are generated through the endosomal pathway, and microvesicles are generated by the direct outward budding of the plasma membrane. Figure 2-2 shows two main mechanisms of EV generation.

Figure 2-2 Biogenesis pathway of cell-derived vesicles. (A) describes the formation of microvesicles through the outward budding and (B) shows the biogenesis of exosomes based on endosomal pathway.

The biogenetic mechanism of microvesicles is believed to be simpler than the biogenesis of exosomes. The microvesicles are produced and released into the extracellular space via the outward budding and fission of the cellular membrane (Figure 2-2A)[17]. Exosomes are created through the endosomal pathway[18]. Firstly, endosomes are formed by the invagination of the cell membrane. The inward budding of the endosomal membrane forms intraluminal vesicles (ILVs) in the endosomes. Through this process, endosomes accumulate ILVs and become multivesicular bodies (MVBs) containing multiple small vesicles in the endosomes. The outer membrane of MVBs is fused with the plasma membrane of the cell. This exposes the inside of the MVBs to the extracellular environment resulting in the release of exosomes (Figure 2-2B)[8,19,20].

2.1.2 Clinical Relevance of EVs

Recent studies have revealed that cancer cells or diseased cells contain specific bioactive components associated with the disease; these components are not found in healthy cells[21-26]. Both diseased cells and healthy cells produce and release EVs into their microenvironment. Therefore, diseased cell-derived-EVs have distinctive biomolecules because EVs contain biomolecules that are originated from the mother cell.

The Epithelial Cell Adhesion Molecule (EpCAM, also known as CD326) has been discovered to be overexpressed in many types of cancer (i.e., esophageal, breast, pancreatic and ovarian cancer)[27-31]. The increased population of the EpCAM positive EVs have also been found in ovarian cancer, breast cancer, and prostate cancer[27,32,33]. Recently, it has been demonstrated that the tumor derived-EVs (tdEVs) are mediators between cancer cells in the local or distant microenvironment[34]. Thus, EVs are believed to be involved in metastasis and cancer growth.

In addition to the cancer biomarker, there are more findings showing the potential of EV utilization for disease biomarkers. The house dust mite is known for causing respiratory ailments. The small creature also produces EVs. These EVs induce an immune response and the production of cytokines associated with respiratory diseases[35]. It has also been discovered that Alzheimer's disease releases EVs containing amyloid- β , which is critically involved in the development of Alzheimer's disease[36-38].

These findings in recent studies make the clinical relevance of EVs clearer and suggest EVs as a novel therapeutic target for certain diseases.

2.2 Raman effect and Raman Spectroscopy

When photons collide with small particles such as atoms and molecules, the particles can absorb the photons and re-emit the light in a different direction in such a way that the intensity of the re-emitted light is not the same as the intensity of the incoming light. This phenomenon is generally called scattering. In nature, almost all of the photons are scattered without energy shift; this scattering is called elastic scattering and is also known as Rayleigh scattering. In 1928 the Indian scientist C.V. Raman and K.S. Krishnan discovered that when monochromatic light is incident on a liquid, the scattered light includes not only the original color but also other colors. This physical phenomenon is called the Raman effect named after C.V. Raman[39,40].

A small fraction of photons (approximately 1 in 10 million photons) is scattered with the energy shift[41,42]. In Raman scattering, photons can be either red or blue shifted. The red shift is known as the Stokes shift, and the red shifted photon has less energy than the energy of the incident photon. The blue shift is called the anti-Stokes shift. The anti-Stokes scattered photon has higher energy than the energy of the incident photon. The energy shifts of the scattering processes can be described with Jablonski diagrams [43] in Figure 2-3. Since most of the electrons are naturally in the ground state, which can only absorb additional energy, Stokes Raman scattering is the predominant phenomena.

Figure 2-3 Jablonski diagram describes Rayleigh scattering, Stokes Raman scattering and anti-Stokes Raman scattering.

2.2.1 Raman Theory

The Raman effect arises when vibrating molecules interact with light; the interaction is known as scattering. The scattering can be explained in both classical and quantum theory. In this chapter, the scattering mechanism will be discussed as a classical theory[41,42,44-46].

2

Light is a form of electromagnetic radiation. Electromagnetic radiation includes gamma rays, X-rays, ultraviolet, visible light, infrared, microwave, and radio waves. Light generally refers to the visible light ranging from 400 to 700 nm in wavelength. Light as electromagnetic radiation has an oscillating electric field, E [41,42,44,45]. The oscillating electric field can be expressed in the form of sinusoidal function with the frequency of the light, ν , and the amplitude of the electric field, E_0 (Only the scalar notation will be used to simplify the equations):

$$E = E_0 \cos(2\pi\nu t) . \quad 2-1$$

The electric field can induce a dipole moment, $p_{induced}$, to the polarizable molecule in the field. The induced dipole moment is proportional to the external electric field, E .

$$p_{induced} = \alpha E + \frac{1}{2}\beta E^2 + \frac{1}{6}\gamma E^3 + \dots , \quad 2-2$$

where α is the polarizability of molecules, β is the hyperpolarizability, and γ is the second hyperpolarizability. Because the non-linear effects at commonly used laser power (up to 1 kW) are negligible[41], only the first term of Equation 2-2 will be taken into account for the explanation.

The polarizability represents the ability of an electron cloud to interact with an electric field to form dipoles. The induced dipole will radiate light at the frequency of the incident light, ν , which gives rise to Rayleigh scattering. The Raman effect arises from the change in polarizability of the molecule caused by the molecular vibration.

Figure 2-4 Diatom molecule as a harmonic oscillator. K is bonding strength of the molecule, m_A and m_B are mass of atom A and B, and x_A and x_B are displacement.

A molecular vibration can be modeled using a diatomic molecule as described in Figure 2-4[41,42,45,47]; the diatomic model can be regarded as a spring/mass system which is a simple harmonic oscillator. The molecule consists of two atoms (mass m_A and m_B) connected with a spring or bonding strength, K . The vibration force is proportional to the bond strength and the displacement. Thus, the vibration can be expressed using Hooke's law.

$$\mu \left(\frac{d^2 q}{dt^2} \right) = -Kq, \quad 2-3$$

where μ is reduced mass, $[m_A m_B / (m_A + m_B)]$, and q is a displacement of the atoms ($x_A + x_B$). By solving the differential equation (Equation 2-3), we get a solution for the displacement. This solution can be written as

$$q = q_0 \cos(2\pi \nu_{vib} t), \quad \nu_{vib} = \frac{1}{2\pi} \sqrt{\frac{K}{\mu}}. \quad 2-4$$

Equation 2-4 shows that the molecule is oscillating at the particular frequency, ν_{vib} where ν_{vib} is the normal mode vibration frequency of the molecule: this frequency is defined by the bond strength and the weight of the atoms of the molecule.

The molecular vibration, however, can change the polarizability of the molecules. If we assume that the polarizability is a function of the displacement, the polarizability α can be denoted as Equation 2-5, using a Maclaurin series.

$$\alpha = \alpha_0 + q \left(\frac{\partial \alpha}{\partial q} \right)_{q=0} + \dots, \quad 2-5$$

where we take into account only the first two terms for the small-amplitude approximation.

Equation 2-2 can be re-written as a combination of Equation 2-1, 2-4, and 2-5. It will result in,

$$p_{induced} = [\alpha_0 E_0 \cos(2\pi\nu t)] + \frac{q_0 E_0}{2} \left(\frac{\partial \alpha}{\partial q} \right)_{q=0} [\cos(2\pi(\nu - \nu_{vib})t) + \cos(2\pi(\nu + \nu_{vib})t)]. \quad 2-6$$

The first term, $[\alpha_0 E_0 \cos(2\pi\nu t)]$, represents Rayleigh scattering. As the formula shows, the photon is scattered without any energy shift. There is an energy shift in the second term of Equation 2-6 when $\partial\alpha/\partial q \neq 0$. This energy shift describes Raman scattering; the decrease in frequency ($\nu - \nu_{vib}$) is the Stokes shift, and the increase ($\nu + \nu_{vib}$) is the anti-Stokes shift. The equation also shows that the vibration frequency of a molecular bond can be directly measured by measuring the frequency shift of the scattered photon.

Figure 2-5 shows an example of a Stokes Raman spectrum of toluene. Monochrome light is absorbed by toluene, and the molecules scatter photon. A Raman spectrum can be obtained by dispersing the scattered photons using a prism or a grating. The Raman shift is usually expressed in the wavenumber scale that intuitively shows the energy shift. The wavenumber is defined as

$$\bar{\nu} = \left(\frac{1}{\lambda_{excitation}} - \frac{1}{\lambda_{scattered}} \right) \times 10^7, \quad 2-7$$

where the wavelength of the excitation beam and the scattered photon are in nanometers (nm). The unit of wavenumber is reciprocal centimeter (cm^{-1}).

Figure 2-5 Toluene, C_7H_8 , molecule (left) and its Stokes Raman spectrum (right). Monochrome light at the frequency of ν interacts with the toluene molecule, which results in Rayleigh, Stokes Raman and anti-Stokes Raman scattering.

2.2.2 Molecular Vibration and Selection Rule

In general, a molecule consisting of N atoms has several degrees of freedom including rotation, translation, and vibration[42,48]. The number of vibrational modes can be described as $3N-6$ for all molecules except for linear molecules. The linear molecules have $3N-5$ modes. However, not all vibrational modes are Raman active. As can be seen in Equation 2-6, to be Raman active, a change in the polarizability of the molecule must be induced while the molecule is vibrating. This can be mathematically described like

$$\left. \frac{\partial \alpha}{\partial q} \right|_{q=0} \neq 0. \quad 2-8$$

A molecule not only scatters an incoming photon but also absorbs. It is called IR absorption when the absorbed frequency matches the vibrational frequency of the molecule. It is only required that there is a dipole induced to be IR active. Vibrational modes of a simple molecule, for example, the carbon dioxide molecule (CO_2), give us more tangible insight into the selection rules. A CO_2 is a linear triatomic molecule and has four normal modes of vibration including symmetric stretching, asymmetric stretching, and two bending (scissoring) motions[42,46,48]. Although a CO_2 molecule has two bending modes, they are essentially the same vibrational mode occurring along different axes. Therefore, only one bending mode will be considered in this discussion.

The asymmetric stretching and bending motion change the dipole moment of the molecule, but these vibrational motions do not change the polarizability of the molecule. Since there

is no change in the polarizability, these two vibrational modes are not (or only weakly) Raman active modes. Since the dipole is induced, these modes are IR active

In contrast to the asymmetric stretching and bending modes, the symmetric stretching mode does induce a change in the polarizability of the molecule during the vibration without change the dipole moment. Thus, the symmetric stretching mode is a strong Raman active mode with a low IR absorption. Figure 2-6 illustrates the three vibrational modes of a CO₂ molecule.

Figure 2-6 A carbon dioxide (CO_2) has three vibrational modes; symmetric stretching, asymmetric stretching, and a bending mode. The symmetric stretching mode changes the polarizability and, thus, the mode is Raman active. However, the other two modes do not change the polarizability. These two modes are not Raman active. They are IR active modes.

2.2.3 Advantages and Application of Raman Spectroscopy

Spectroscopy is a method to study the interaction between electromagnetic radiation and matter. There are many diverse spectroscopic techniques such as optical spectroscopy, mass spectroscopy, nuclear magnetic resonance spectroscopy, and so on[41].

Raman spectroscopy uses the Raman effect to investigate the vibrational modes in a specimen. After an interaction between monochrome light and molecules, the scattered light can be dispersed by an optical component (e.g., grating or prism). This spectrum will include several peaks. Because the location of the peaks corresponds to the energy difference in molecular vibration modes, the Raman spectrum gives insight into the chemical composition of the sample. Raman spectroscopy has several advantages[41].

- It is a non-destructive technique.
- It can be combined with imaging techniques to produce a hyperspectral image.
- The linearity of the Raman scattering facilitates quantitative analysis compared to coherent anti-Stokes Raman spectroscopy (CARS).
- Since the interaction with water is very weak in the visible range, Raman is a suitable technique to study the water in aqueous solutions or biological samples.
- It requires little sample preparation.
- It can probe solid, liquid, and gas-phase material.

Although Raman spectroscopy has some disadvantages (for example, it is ~1000 times weaker than Rayleigh scattering, and it tends to excite a fluorescent background), the advantages are such that Raman spectroscopy has a wide variety of applications[42] such as material studies[49-51], biological and medical studies[22,23,26,52-59], pharmaceutical applications[60-62], forensic applications[63-65], art and archeology [66-68]. Raman spectroscopy is used progressively in real-life applications. For example, in the clinic, Raman spectroscopy is used as a tool for surgical support [54,69].

2.3 Optical Trapping

Optical tweezers or optical trapping uses highly focused light to manipulate microscopic objects. An optical tweezer can attract a particle into the focus or can repel a particle away depending on the refractive index contrast between the suspension and the particle. Since the first optical trap was reported by Ashkin *et al.* in 1986[70], it has been widely used in the field of chemistry and biology[71-74]. The trap has not only been used for particle transport or manipulation[75-77] but also for studying mechanical properties of the single biomolecule[78-80].

When individual optical rays enter and exit the particle, the propagation direction of the exit rays will be changed due to the refraction. Since light has momentum[81], altering the propagation direction indicates that the momentum of the photon is also changed. According to Newton's third law, a change in momentum causes the same level of momentum in the opposite direction to conserve the total momentum of the system.

A strong electric field gradient is induced at the center of a focused beam. Thus, the center of the beam has a higher field density than the surroundings. If a particle is displaced from the center, the particle will feel an imbalanced momentum because of the different field densities. Thus, a net force toward the center of the beam is induced, and the particle will be pulled to the center of the beam, Figure 2-7. Once the particle located at the center of the trap, the particle finds a momentum equilibrium and will stay in the trap.

Figure 2-7 Due to the gradient of the light, there are more photons carried at the center of the beam along optical axis. A particle displaced from the center will experience a net force toward center.

Bibliography

- [1] J. O. Park, D. Y. Choi, D. S. Choi, H. J. Kim, J. W. Kang, J. H. Jung, J. H. Lee, J. Kim, M. R. Freeman, K. Y. Lee, "Identification and characterization of proteins isolated from microvesicles derived from human lung cancer pleural effusions", *Proteomics*, **2013**; *13*, 2125.
- [2] E. van der Pol, A. N. Böing, P. Harrison, A. Sturk, R. Nieuwland, "Classification, functions, and clinical relevance of extracellular vesicles", *Pharmacological Reviews*, **2012**; *64*, 676.
- [3] M. Verma, T. K. Lam, E. Hebert, R. L. Divi, "Extracellular vesicles: potential applications in cancer diagnosis, prognosis, and epidemiology", *BMC Clinical Pathology*, **2015**; *15*, 6.
- [4] L. Margolis, Y. Sadovsky, "The biology of extracellular vesicles: The known unknowns", *PLoS Biology*, **2019**; *17*, e3000363.
- [5] E. Chargaff, R. West, "The biological significance of the thromboplastic protein of blood", *Journal of Biological Chemistry*, **1946**; *166*, 189.
- [6] E. van der Pol, Ph.D. Dissertation, "Detection of extracellular vesicles: size does matter", *Universiteit van Amsterdam*, **2015**.
- [7] N. Arraud, R. Linares, S. Tan, C. Gounou, J. M. Pasquet, S. Mornet, A. R. Brisson, "Extracellular vesicles from blood plasma: determination of their morphology, size, phenotype and concentration", *Journal of Thrombosis and Haemostasis*, **2014**; *12*, 614.
- [8] E. R. Abels, X. O. Breakefield, "Introduction to extracellular vesicles: biogenesis, RNA cargo selection, content, release, and uptake", *Cellular and Molecular Neurobiology*, **2016**; *36*, 301.
- [9] J. Skog, T. Würdinger, S. Van Rijn, D. H. Meijer, L. Gainche, W. T. Curry Jr, B. S. Carter, A. M. Krichevsky, X. O. Breakefield, "Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers", *Nature Cell Biology*, **2008**; *10*, 1470.
- [10] K. Al-Nedawi, B. Meehan, J. Rak, "Microvesicles: messengers and mediators of tumor progression", *Cell cycle*, **2009**; *8*, 2014.

-
- [11] H. Valadi, K. Ekström, A. Bossios, M. Sjöstrand, J. J. Lee, J. O. Lötval, "Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells", *Nature Cell Biology*, **2007**; *9*, 654.
- [12] B. György, T. G. Szabó, M. Pásztói, Z. Pál, P. Misják, B. Aradi, V. László, E. Pállinger, E. Pap, A. Kittel, "Membrane vesicles, current state-of-the-art: emerging role of extracellular vesicles", *Cellular and Molecular Life Sciences*, **2011**; *68*, 2667.
- [13] M. F. Baietti, Z. Zhang, E. Mortier, A. Melchior, G. Degeest, A. Geeraerts, Y. Ivarsson, F. Depoortere, C. Coomans, E. Vermeiren, "Syndecan–syntenin–ALIX regulates the biogenesis of exosomes", *Nature Cell Biology*, **2012**; *14*, 677.
- [14] A. N. Böing, E. van der Pol, A. E. Grootemaat, F. A. W. Coumans, A. Sturk, R. Nieuwland, "Single-step isolation of extracellular vesicles by size-exclusion chromatography", *Journal of Extracellular Vesicles*, **2014**; *3*, 23430.
- [15] F. A. Coumans, A. R. Brisson, E. I. Buzas, F. Dignat-George, E. E. Drees, S. El-Andaloussi, C. Emanuelli, A. Gasecka, A. Hendrix, A. F. Hill, "Methodological guidelines to study extracellular vesicles", *Circulation Research*, **2017**; *120*, 1632.
- [16] J. C. Akers, D. Gonda, R. Kim, B. S. Carter, C. C. Chen, "Biogenesis of extracellular vesicles (EV): exosomes, microvesicles, retrovirus-like vesicles, and apoptotic bodies", *Journal of Neuro-Oncology*, **2013**; *113*, 1.
- [17] V. R. Minciocchi, M. R. Freeman, D. Di Vizio In *Seminars in cell & developmental biology*; Elsevier: **2015**; Vol. 40, p 41.
- [18] C. Théry, M. Ostrowski, E. Segura, "Membrane vesicles as conveyors of immune responses", *Nature reviews immunology*, **2009**; *9*, 581.
- [19] B. D. Grant, J. G. Donaldson, "Pathways and mechanisms of endocytic recycling", *Nature reviews Molecular cell biology*, **2009**; *10*, 597.
- [20] M. S. Pols, J. Klumperman, "Trafficking and function of the tetraspanin CD63", *Experimental Cell Research*, **2009**; *315*, 1584.
- [21] L. A. Austin, S. Osseiran, C. L. Evans, "Raman technologies in cancer diagnostics", *The Analyst*, **2016**; *141*, 476.
- [22] A. S. Haka, K. E. Shafer-Peltier, M. Fitzmaurice, J. Crowe, R. R. Dasari, M. S. Feld, "Diagnosing breast cancer by using Raman spectroscopy", *Proceedings of the National Academy of Sciences of the United States of America*, **2005**; *102*, 12371.

- [23] R. E. Kast, S. C. Tucker, K. Killian, M. Trexler, K. V. Honn, G. W. Auner, "Emerging technology: applications of Raman spectroscopy for prostate cancer", *Cancer and Metastasis Reviews*, **2014**; *33*, 673.
- [24] Y. H. Ong, M. Lim, Q. Liu, "Comparison of principal component analysis and biochemical component analysis in Raman spectroscopy for the discrimination of apoptosis and necrosis in K562 leukemia cells", *Optics Express*, **2012**; *20*, 22158.
- [25] X. Song, R. D. Airan, D. R. Arifin, A. Bar-Shir, D. K. Kadayakkara, G. Liu, A. A. Gilad, P. C. van Zijl, M. T. McMahon, J. W. Bulte, "Label-free in vivo molecular imaging of underglycosylated mucin-1 expression in tumour cells", *Nature communications*, **2015**; *6*, 6719.
- [26] Z. Huang, A. McWilliams, H. Lui, D. I. McLean, S. Lam, H. Zeng, "Near-infrared Raman spectroscopy for optical diagnosis of lung cancer", *International Journal of Cancer*, **2003**; *107*, 1047.
- [27] W. A. Osta, Y. Chen, K. Mikhitarian, M. Mitas, M. Salem, Y. A. Hannun, D. J. Cole, W. E. Gillanders, "EpCAM is overexpressed in breast cancer and is a potential target for breast cancer gene therapy", *Cancer Research*, **2004**; *64*, 5818.
- [28] M. Munz, P. A. Baeuerle, O. Gires, "The emerging role of EpCAM in cancer and stem cell signaling", *Cancer Research*, **2009**; *69*, 5627.
- [29] P. Baeuerle, O. Gires, "EpCAM (CD326) finding its role in cancer", *British Journal of Cancer*, **2007**; *96*, 417.
- [30] H. Kimura, H. Kato, A. Faried, M. Sohda, M. Nakajima, Y. Fukai, T. Miyazaki, N. Masuda, M. Fukuchi, H. Kuwano, "Prognostic significance of EpCAM expression in human esophageal cancer", *International Journal of Oncology*, **2007**; *30*, 171.
- [31] J. Ni, P. J. Cozzi, W. Duan, S. Shigdar, P. H. Graham, K. H. John, Y. Li, "Role of the EpCAM (CD326) in prostate cancer metastasis and progression", *Cancer and Metastasis Reviews*, **2012**; *31*, 779.
- [32] D. D. Taylor, C. Gercel-Taylor, "MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer", *Gynecologic Oncology*, **2008**; *110*, 13.
- [33] T. Liu, D. E. Mendes, C. E. Berkman, "Functional prostate-specific membrane antigen is enriched in exosomes from prostate cancer cells", *International Journal of Oncology*, **2014**; *44*, 918.
- [34] A. Becker, B. K. Thakur, J. M. Weiss, H. S. Kim, H. Peinado, D. Lyden, "Extracellular vesicles in cancer: cell-to-cell mediators of metastasis", *Cancer Cell*, **2016**; *30*, 836.

-
- [35] J. P. Choi, S. G. Jeon, Y. K. Kim, Y. S. Cho, "Role of house dust mite-derived extracellular vesicles in a murine model of airway inflammation", *Clinical and Experimental Allergy*, **2019**; 49, 227.
- [36] L. Rajendran, M. Honsho, T. R. Zahn, P. Keller, K. D. Geiger, P. Verkade, K. Simons, "Alzheimer's disease β -amyloid peptides are released in association with exosomes", *Proceedings of the National Academy of Sciences*, **2006**; 103, 11172.
- [37] S. Sölvander, E. Nikitidou, R. Brolin, L. Söderberg, D. Sehlin, L. Lannfelt, A. Erlandsson, "Accumulation of amyloid- β by astrocytes result in enlarged endosomes and microvesicle-induced apoptosis of neurons", *Molecular Neurodegeneration*, **2016**; 11, 38.
- [38] C. Quek, A. F. Hill, "The role of extracellular vesicles in neurodegenerative diseases", *Biochemical and Biophysical Research Communications*, **2017**; 483, 1178.
- [39] C. V. Raman, K. S. Krishnan, "A new type of secondary radiation", *Nature*, **1928**; 121, 501.
- [40] C. V. Raman, "A new radiation", *Indian Journal of Physics*, **1928**; 2, 387.
- [41] G. Gauglitz, T. Vo-Dinh *Handbook of spectroscopy*; Wiley Online Library, **2003**; Vol. 1.
- [42] E. Smith, G. Dent *Modern Raman spectroscopy: a practical approach*; Wiley, **2005**.
- [43] A. Jablonski, "Efficiency of anti-Stokes fluorescence in dyes", *Nature*, **1933**; 131, 839.
- [44] J. R. Ferraro *Introductory raman spectroscopy*; Elsevier, **2003**.
- [45] J.-X. Cheng, X. S. Xie *Coherent Raman scattering microscopy*; CRC press, **2016**.
- [46] D. A. Long *The raman effect*; Wiley, **2002**.
- [47] D. C. Harris, M. D. Bertolucci *Symmetry and spectroscopy: an introduction to vibrational and electronic spectroscopy*; Courier Corporation, **1989**.
- [48] A. Beiser *Concepts of modern physics*; Tata McGraw-Hill Education, **2003**.
- [49] M. J. Pelletier *Analytical applications of Raman spectroscopy*; Blackwell science Oxford, **1999**; Vol. 427.

- [50] S. Reich, C. Thomsen, "Raman spectroscopy of graphite", *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, **2004**; 362, 2271.
- [51] N. Schlotter, J. Rabolt, "Raman spectroscopy in polymeric thin film optical waveguides. 1. Polarized measurements and orientational effects in two-dimensional films", *The Journal of Physical Chemistry*, **1984**; 88, 2062.
- [52] S. Sigurdsson, P. A. Philipsen, L. K. Hansen, J. Larsen, M. Gniadecka, H.-C. Wulf, "Detection of skin cancer by classification of Raman spectra", *IEEE Transactions on Biomedical Engineering*, **2004**; 51, 1784.
- [53] J. Desroches, M. Jermyn, K. Mok, C. Lemieux-Leduc, J. Mercier, K. St-Arnaud, K. Urmeý, M.-C. Guiot, E. Marple, K. Petrecca, "Characterization of a Raman spectroscopy probe system for intraoperative brain tissue classification", *Biomedical optics express*, **2015**; 6, 2380.
- [54] M. Jermyn, J. Desroches, J. Mercier, M.-A. Tremblay, K. St-Arnaud, M.-C. Guiot, K. Petrecca, F. Leblond, "Neural networks improve brain cancer detection with Raman spectroscopy in the presence of operating room light artifacts", *Journal of biomedical optics*, **2016**; 21, 094002.
- [55] L. Hartsuiker, W. Petersen, R. G. Rayavarapu, A. Lenferink, A. A. Poot, L. W. Terstappen, T. G. Van Leeuwen, S. Manohar, C. Otto, "Raman and fluorescence spectral imaging of live breast cancer cells incubated with PEGylated gold nanorods", *Applied Spectroscopy*, **2012**; 66, 66.
- [56] H. Abramczyk, J. Surmacki, M. Kopeć, A. Olejnik, K. Lubecka-Pietruszewska, K. Fabianowska-Majewska, "The role of lipid droplets and adipocytes in cancer. Raman imaging of cell cultures: MCF10A, MCF7, and MDA-MB-231 compared to adipocytes in cancerous human breast tissue", *Analyst*, **2015**; 140, 2224.
- [57] C. W. Freudiger, W. Min, B. G. Saar, S. Lu, G. R. Holtom, C. He, J. C. Tsai, J. X. Kang, X. S. Xie, "Label-free biomedical imaging with high sensitivity by stimulated Raman scattering microscopy", *Science*, **2008**; 322, 1857.
- [58] W. Lee, A. Nanou, L. Rikkert, F. A. W. Coumans, C. Otto, L. W. M. M. Terstappen, H. L. Offerhaus, "Label-Free Prostate Cancer Detection by Characterization of Extracellular Vesicles Using Raman Spectroscopy", *Analytical Chemistry*, **2018**; 90, 11290.
- [59] X. Nan, J.-X. Cheng, X. S. Xie, "Vibrational imaging of lipid droplets in live fibroblast cells with coherent anti-Stokes Raman scattering microscopy", *Journal of Lipid Research*, **2003**; 44, 2202.

- [60] T. Vankeirsbilck, A. Vercauteren, W. Baeyens, G. Van der Weken, F. Verpoort, G. Vergote, J. P. Remon, "Applications of Raman spectroscopy in pharmaceutical analysis", *TRAC Trends in Analytical Chemistry*, **2002**; *21*, 869.
- [61] C. Shende, W. Smith, C. Brouillette, S. Farquharson, "Drug stability analysis by Raman spectroscopy", *Pharmaceutics*, **2014**; *6*, 651.
- [62] S. Kazarian, G. Martirosyan, "Spectroscopy of polymer/drug formulations processed with supercritical fluids: in situ ATR-IR and Raman study of impregnation of ibuprofen into PVP", *International Journal of Pharmaceutics*, **2002**; *232*, 81.
- [63] J. M. Chalmers, H. G. Edwards, M. D. Hargreaves *Infrared and Raman spectroscopy in forensic science*; John Wiley & Sons, **2012**.
- [64] A. Braz, M. López-López, C. García-Ruiz, "Raman spectroscopy for forensic analysis of inks in questioned documents", *Forensic Science International*, **2013**; *232*, 206.
- [65] J. De Gelder, P. Vandenabeele, F. Govaert, L. Moens, "Forensic analysis of automotive paints by Raman spectroscopy", *Journal of Raman Spectroscopy: An International Journal for Original Work in all Aspects of Raman Spectroscopy, Including Higher Order Processes, and also Brillouin and Rayleigh Scattering*, **2005**; *36*, 1059.
- [66] D. Bersani, J. M. Madariaga, "Applications of Raman spectroscopy in art and archaeology", *Journal of Raman Spectroscopy*, **2012**; *43*, 1523.
- [67] M. Bouchard, D. Smith, "Catalogue of 45 reference Raman spectra of minerals concerning research in art history or archaeology, especially on corroded metals and coloured glass", *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **2003**; *59*, 2247.
- [68] P. Vandenabeele, H. G. Edwards, L. Moens, "A decade of Raman spectroscopy in art and archaeology", *Chemical Reviews*, **2007**; *107*, 675.
- [69] A. S. Haka, Z. Volynskaya, J. A. Gardecki, J. Nazemi, J. Lyons, D. Hicks, M. Fitzmaurice, R. R. Dasari, J. P. Crowe, M. S. Feld, "In vivo margin assessment during partial mastectomy breast surgery using Raman spectroscopy", *Cancer Research*, **2006**; *66*, 3317.
- [70] A. Ashkin, J. Dziedzic, J. Bjorkholm, S. Chu, "Observation of a single-beam gradient force optical trap for dielectric particles", *Optics Letters*, **1986**; *11*, 288.
- [71] M. D. Wang, H. Yin, R. Landick, J. Gelles, S. M. Block, "Stretching DNA with optical tweezers", *Biophysical Journal*, **1997**; *72*, 1335.

- [72] H. Zhang, K.-K. Liu, "Optical tweezers for single cells", *Journal of the Royal Society interface*, **2008**; *5*, 671.
- [73] D. P. Cherney, J. C. Conboy, J. M. Harris, "Optical-trapping Raman microscopy detection of single unilamellar lipid vesicles", *Analytical Chemistry*, **2003**; *75*, 6621.
- [74] K. Ajito, K. Torimitsu, "Single nanoparticle trapping using a Raman tweezers microscope", *Applied Spectroscopy*, **2002**; *56*, 541.
- [75] A. Ashkin, J. M. Dziedzic, "Optical trapping and manipulation of viruses and bacteria", *Science*, **1987**; *235*, 1517.
- [76] G. Weber, K. O. Greulich, In *International review of cytology*; Elsevier: **1992**; Vol. 133, p 1.
- [77] D. G. Grier, "A revolution in optical manipulation", *Nature*, **2003**; *424*, 810.
- [78] L. Osinkina, S. Carretero-Palacios, J. Stehr, A. A. Lutich, F. Jäckel, J. Feldmann, "Tuning DNA binding kinetics in an optical trap by plasmonic nanoparticle heating", *Nano Letters*, **2013**; *13*, 3140.
- [79] K. Pant, R. L. Karpel, M. C. Williams, "Kinetic regulation of single DNA molecule denaturation by T4 gene 32 protein structural domains", *Journal of Molecular Biology*, **2003**; *327*, 571.
- [80] C. Kleimann, A. Sischka, A. Spiering, K. Tönsing, N. Sewald, U. Diederichsen, D. Anselmetti, "Binding kinetics of bisintercalator Triostin a with optical tweezers force mechanics", *Biophysical Journal*, **2009**; *97*, 2780.
- [81] D. L. Andrews, M. Babiker *The angular momentum of light*, Cambridge University Press, **2012**.

Characterization and Classification of Extracellular Vesicles using Raman Spectroscopy and Principal Component Analysis

All mammalian cells release extracellular vesicles (EVs) into their micro-environment. The vesicles travel through the body along the stream of bodily fluids. EVs contain a wide range of biomolecules. The transported cargo varies depending on the EV origin. Knowledge of the cellular origin and chemical composition of EVs can potentially be used as a biomarker to detect, stage, and monitor diseases. In this paper, we demonstrate the potential of EVs as a prostate cancer biomarker. A Raman optical tweezer was employed to obtain Raman signatures from four types of EV samples, which were red blood cell- and platelet-derived EVs of healthy donors and the prostate cancer cell lines- (PC3 and LNCaP) derived EVs. EVs' Raman spectra could be separated/classified into distinct groups using principal component analysis (PCA) which permits the discrimination of the investigated EV subtypes. These findings may provide a new methodology to detect and monitor early-stage cancer.

Part of this chapter has been published in *Analytical Chemistry*, DOI: 10.1021/acs.analchem.8b01831.

Label-free Prostate Cancer Detection by Characterization of Extracellular Vesicles using Raman Spectroscopy
Wooje Lee, Afroditi Nanou, Linda Rikkert, Frank A.W. Coumans, Cees Otto, Leon W.M.M. Terstappen, and Herman L. Offerhaus

3.1 Introduction

Extracellular vesicles (EVs)[1-3] are small spherical particles (diameter between 30 nm and 1 μm) enclosed by a phospholipid bilayer, shed by living cells into their extracellular environment[2]. Both healthy and unhealthy cells secrete EVs so that EVs are found in all body fluids, such as blood plasma[4], urine[5], and breast milk[6]. These small particles play a significant role in both intercellular communication and waste control[2,7].

3

EVs are formed through several biogenesis pathways, for example, the endolysosomal pathway or budding from the plasma membrane[3]. The vesicle formation process allows the parent cells to package biomolecules with the generated EVs, such as membrane lipids, proteins, receptors, and genetic information[3]. These biomolecules are transported by the EVs from the parent cell to a recipient cell[2,4,8,9]. The molecular composition of the transported cargo has been shown to change depending on the origin of the EVs. Therefore, EVs released from healthy and diseased cells are likely to contain different combinations of biological molecules. The different types of cargo imply that EVs can be utilized as a disease biomarker[2,3], and the clinical relevance of EVs[2,3] has been explored in various studies[10].

Recently it was shown that EVs secreted by tumor cells contain tumor antigens[11,12]. Various biochemical compositions of cancer-derived EVs suggest a potential of EVs as a biomarker not only for cancer diagnosis but also for cancer prognosis and the monitoring of patients after or during treatment[13]. Furthermore, the alterations in EV molecular content are reflected in a different spectral response. The spectroscopy can be used for the analysis. Raman spectroscopy is an analytical tool long used to determine molecular composition without external labels. Therefore, this vibrational spectroscopic technique presents a potentially useful opportunity for such an analysis[14-17].

Spontaneous Raman spectroscopy is a type of vibrational spectroscopy based on inelastic scattering by molecules. When incident photons are scattered by molecules, some are scattered with particular energy shifts, a phenomenon called Raman scattering[18]. Raman microscopy is used exclusively to investigate structural and compositional information of a specimen[18,19]. Since the optical technique yields the fingerprint of chemicals, it has been widely used in biological and pharmaceutical fields[10,20-22]. It has been applied to identify differences in tissues and cells. Convincing spectral differences have been demonstrated between cancer cells and healthy cells based on lipid droplet content, carotenoids, and ratios between different proteins[14,21-25].

Therefore, Raman spectroscopy is a promising tool to reveal the structural differences among EVs of various origins. However, the vibrational differences across the EV subtypes are subtle. Such subtle differences require sensitive and reliable analysis, such as principal component analysis (PCA). This statistical technique is used to interpret high dimensional data with several inter-correlated variables[26]. PCA is widely utilized in pattern recognition, image processing, and spectroscopy. PCA differs from supervised learning in the sense that all variation is evaluated unsupervised so that dependence on peculiarities of the assignments in the training set are avoided as all spectra are used without assignment

In this study, spontaneous Raman[18] was utilized to obtain spectral fingerprints of four different EVs subsets that had been derived from two prostate cancer cell lines (LNCaP and PC3) and platelet and red blood cells from healthy donors. We obtained the spectral fingerprints of each EV subtype and used PCA to identify the four vesicle subtypes based on 300 spectra. The discrimination that we aim for is not between EVs from healthy prostate cells and EVs from cancer prostate cells since this is not a discrimination that would be useful in diagnosis (a healthy person lacks EVs from cancer cells). Rather we seek to discriminate EVs from prostate cancer cells from EVs derived from (healthy) platelets and red blood cells.

3.2 Experiments

For this study, we prepared EVs from four cell types; red blood, platelet, PC3, and LNCaP. In this part, sample preparation, confirmation, and Raman experiment will be described.

3.2.1 Preparation of EVs

3.2.1.1 Preparation of blood cells-derived EVs

Red blood cell concentrate (150 mL) obtained from Sanquin (Amsterdam, The Netherlands) was diluted 1:1 with filtered phosphate-buffered saline (PBS; 154 mM NaCl, 1.24 mM $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$, 0.2 mM $\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$, pH 7.4; supplemented with 0.32% trisodium citrate; 0.22 mm filter (Merck chemicals BV, Darmstadt, Germany)) and centrifuged three times for 20 minutes at 1,560xg, 20°C using a Rotina 46RS centrifuge (Hettich, Tuttlingen, Germany). The EV-containing supernatant was pooled, and aliquots of 50 μL were frozen in liquid nitrogen and stored at -80°C.

Platelet concentrate (100 mL) obtained from Sanquin (Amsterdam, The Netherlands) was diluted 1:1 with filtered PBS. Next, 40 mL acid citrate dextrose (ACD; 0.85 M trisodium citrate, 0.11 M D-glucose, and 0.071 M citric acid) was added, and the suspension was centrifuged for 20 minutes at 800xg, 20°C. Thereafter, the supernatant was centrifuged (20 minutes at 1,560xg, 20°C). This centrifugation procedure was repeated twice to ensure the removal of platelets. The vesicle-containing supernatant was pooled, and aliquots of 50 µL were frozen in liquid nitrogen and stored at -80°C. Samples were thawed on melting ice for 30 min before use.

3.2.1.2 Preparation of prostate cancer-derived EVs

Two prostate cancer cell lines (PC3 and LNCaP) were used as a model to produce prostate cancer-derived EVs. Cell lines were cultured at 37°C and 5% CO₂ in Dulbecco's modified Eagle medium, RPMI 1640 with L-glutamine (Thermo Fischer Scientific, 11875) supplemented with 10% v/v fetal bovine serum, 10 units/mL penicillin and 10 µg/ml streptomycin. The medium was refreshed every second day. When cells reached 80-90% confluence, they were washed three times with PBS and FBS-free RPMI medium supplemented with 1 unit/mL penicillin and 1 µg/ml streptomycin was added to the cells. After 48 h of cell culture, the cell supernatant was collected and centrifuged at 1000xg for 30 minutes. The invisible pellet containing dead or apoptotic cells and the biggest in size population of EVs was discarded. The supernatant was pooled, and aliquots of 50 µl were frozen in liquid nitrogen and stored at -80°C. Size distribution and presence of the harvested EVs were assessed with Nanoparticle Tracking Analysis (NTA) and Transmission Electron Microscopy (TEM) images were taken to provide some examples of EVs.

Figure 3-1 EVs sample preparation flow. (A) RBC-EVs, (B) platelet-EVs and (C) PC3- and LNCaP-EVs preparation flow.

3.2.2 Sample validation

The term “EVs” is somewhat ambiguous and there was no firm definition for the term. Thus, International Society of Extracellular Vesicles (ISEV) recommended a guideline to fast characterize EVs prior to further study[10]. The ISEV criteria recognize TEM and NTA as suitable techniques to validate isolated samples. Thus, we carried out the validation of the EVs samples with TEM and NTA.

Figure 3-2 Concentration and size distribution of EV samples measured using NTA. Panel (A), (B), (C) and (D) represent NTA result of red blood cell-derived EVs, platelet derived-EVs, PC3-derived EVs and LNCaP derived-EVs, respectively. Mean size of red blood cell derived EVs is 148 ± 3.7 nm and its concentration is $0.85 \times 10^8 \pm 0.03 \times 10^8$ particles/ml. Platelet-derived EVs is 89 ± 4.6 nm and $0.42 \times 10^8 \pm 0.02 \times 10^8$ particles/ml. PC3-derived EVs is 172 ± 3.7 nm and $1.00 \times 10^8 \pm 0.03 \times 10^8$ particles/ml. LNCaP-derived EVs is 167 ± 4.4 nm and $1.06 \times 10^8 \pm 0.05 \times 10^8$ particles/ml.

3.2.2.1 Size distribution measurement using nanoparticle tracking analysis (NTA)

The concentration and size distribution of particles in the EV-containing samples were measured by NTA (NS500; Nanosight, Amesbury, UK), equipped with an EMCCD camera and a 405 nm diode laser. Silica beads (105 nm diameter; Microspheres-Nanospheres, Cold Spring, NY) were used to configure and calibrate the instrument. Fractions were diluted 10 to 2,000-fold in filtered PBS to reduce the number of particles in the field of view below 200/image. Of each sample, 10 videos, each of 30-seconds duration, were captured with the camera shutter set at 33.31 ms and the camera gain set at 400. All samples were analyzed using the same threshold, which was calculated by custom-made software (MATLAB v.7.9.0.529). The analysis was performed by the instrument software (NTA 2.3.0.15). The size distribution and concentration of EV samples are shown in Figure 3-2.

3.2.2.2 Visualizing prepared sample using transmission electron microscopy

Size exclusion chromatography was used to isolate EVs from the platelet and red blood cell EV-containing samples[27]. Sepharose CL-2B (30 mL, GE Healthcare; Uppsala, Sweden) was washed with PBS containing 0.32% trisodium citrate (pH 7.4, 0.22 mm filtered). Subsequently, a frit was placed at the bottom of a 10 mL plastic syringe (Becton Dickinson (BD), San Jose, CA)), and the syringe was stacked with 10 mL washed sepharose CL-2B to create a column with 1.6 cm in diameter and 6.2 cm in height. Platelet or red blood cell EV-containing samples (125 μ L) were loaded on the respective column, followed by elution with PBS/0.32% citrate (pH 7.4, 0.22 mm filtered). The first 1 mL was discarded, and the following 500 μ L was collected.

All EV samples were fixed 1:1 in a 0.1% final concentration (v/v) paraformaldehyde (Electron Microscopy Science, Hatfield, PA) for 30 min. Then, a 300-mesh carbon-coated Formvar film nickel grid (Electron Microscopy Science) was placed on 10 μ L of fixed sample for 7 minutes. Thereafter, the grid was transferred onto drops of 1.75% uranyl acetate (w/v) for negative staining, blotted after 7 minutes and air-dried. Each grid was studied through a transmission electron microscope (Fei, Tecnai-12; Eindhoven, the Netherlands) operated at 100 kV using a Veleta 2,048 x 2,048 side-mounted CCD camera and Imaging Solutions software (Olympus, Shinjuku, Tokyo, Japan). All steps were performed at room temperature, and all used liquids were filtered through 0.22 μ m filters. TEM images of the various groups of EVs are shown in Figure 3-3.

Figure 3-3 Transmission electron microscope images of EV subtypes. Arrows point EVs in the figure. (A) red blood cell-derived EVs, (B) platelet-derived EVs, (C) PC3-derived EVs and (D) LNCaP-derived EVs. Scale bar in each panel is 500 nm.

Figure 3-4 Schematic diagram of confocal Raman microscope. Kr-ion laser emits 647 nm light. The pump light is cleaned up by laser line filter and focuses on the sample. Scattered light is collected by same objective.

3.2.3 Raman setup

A custom-built Raman microscope has been used to obtain the Raman signature of EVs. The Raman microscope is briefly described in Figure 3-4. The microscope has a Kr⁺ laser (Innova 90-K, Coherent Inc., Santa Clara, CA) which emits a wavelength of 647 nm for excitation. The laser beam is cleaned up by a laser clean-up filter (LL01-647-12.5; Semrock Inc., Rochester, NY) and is focused onto the sample by a microscope objective (40X/0.95NA UPLSAPO, Olympus corp., Tokyo, Japan). The scattered photons are collected by the same objective lens. The light passes through a dichroic beamsplitter (Di02-R35-25x36; Semrock Inc.) and a long-pass edge filter (LP02-647RU-25; Semrock Inc.). Then, the light is focused on a 15 μm pinhole at the entrance of a custom-made spectrograph. The prism-based spectrograph disperses collected light in the range of 646-849 nm[28]. The pinhole allows us to achieve confocal configuration with a lateral laser spot size of about 350 nm and an axial resolution of about 1.5 μm . The dispersed light is recorded by a -70 °C cooled EMCCD camera which is cooled by an embedded Peltier based cooler (Newton DU-970N-BV, Andor Technology Ltd., Belfast, Northern Ireland)[28].

3.2.4 Raman spectral data acquisition

For the Raman measurements, 25 aliquots were prepared for each EV subtype (in total 100 aliquots). Each aliquot contains 50 μl of EVs sample. An aliquot was placed in a microscope slide glass. The microscope slide is made from borosilicate glass and has about 50 μl hollow cavity in the middle. After the sample was added in the cavity, the cavity was covered by a thin glass disk (0.25 μm , borosilicate glass) to prevent evaporation and contamination.

It is quite difficult to measure EVs with a conventional Raman microscope since they are very small and float in suspension. Therefore, we used Raman optical tweezers. Optical trapping captures the vesicles in the waist of the highly focused beam[10,29]. To minimize fluorescence and a Raman background generated by the 647 nm pump beam, we focused the excitation beam 50 μm below the bottom of the disk coverslip. The focal plane can be controlled by a piezo with a 1 μm precision. The power of the excitation beam was 50 mW under the objective. The exposure time per spectrum was 10 seconds and 16 spectra were obtained at a fixed position (160 seconds in total). After each data acquisition, we closed the laser shutter and moved the sample stage to capture new vesicles. Uniform experimental conditions were applied during all the experiments. The aliquot was replaced after every third measurement; 25 aliquots were measured three times to measure each subtype 75 times, and each measurement data includes 16 spectra of trapped EVs.

Figure 3-5 Raman spectra of each vesicle EV subtypes. (A1-3) show Raman spectra of red blood cell- derived EVs. (B1-3) are Raman spectra of platelet- derived EVs. (C1-3) are PC3- derived EVs. (D1-3) are LNCaP- derived EVs. (E) and (F) shows Raman spectra of negative controls which are PBS and RPMI-1640, respectively. Insets of panel (E) and (F) show fingerprint of the suspension. The first column shows Raman spectra with suspension signal. The second column shows Raman spectra of EVs without suspension, and the last column shows background corrected Raman spectra.

3.2.5 Data processing and principal component analysis

All programs were implemented in MATLAB R2016b (version 9.1.0, The MathWorks, Natick, MA). Cosmic rays and the readout noise of the system were corrected by custom made software. The software was realized by LabVIEW (LabVIEW 2013; National Instruments, Austin, TX). The raw Raman signal was recorded as a function of the pixel number. Pixel numbers were converted into the wavenumber scale using toluene peaks and ArHg lines (520, 785, 1003, 1030, 1210, 1604, 2919, 3056, 1097, 1303, 1705, 1910, 2126, 2145, 2357, 2508, 2873, 2964, 2977, 3114, 3131, 3354, 3560 and 3582 cm^{-1}) for calibration.

Figure 3-5 shows the Raman spectra of each EV subtypes and two suspensions that are PBS and RPMI-1640. RBC- and platelet-derived EVs are suspended in PBS, and PC3- and LNCaP-derived EVs are suspended in RPMI-1640. Although a Raman a band between 1500 - 1700 cm^{-1} was observed from the Raman spectra of PBS and RPMI-1640, a peak around 1450 cm^{-1} and a band 2800 - 3050 cm^{-1} , that are commonly appeared in the Raman spectra of EVs (Figure 3-5 A1, B1, C1 and D1), are not observed in the Raman spectra of two suspensions. Therefore, it is believed that we measured the Raman spectra of EVs in the suspension.

Since the volume of EVs is about 100 folds smaller than the confocal volume, the contribution of the vesicles to the total signal was much weaker than the background from the suspension (PBS or RPMI-1640 cell culture medium). The Raman spectrum of the suspensions was subtracted from EVs measurements to retrieve the contribution of the EVs. Two tdEVs subtypes are suspended in RPMI, and the other two subtypes are suspended in PBS. Thus, the PBS contributions were removed from the Raman spectra of RBC- and platelet-EVs by subtraction. The Raman contributions of the cell culture medium were removed from two tdEVs subtypes (Figure 3-5 A2, B2, C2 and D2).

16 Raman spectra of each measurement were averaged to reduce the shot noise of the data. As a result, we produced 75 Raman spectra for each subtype. Nevertheless, the processed data still contained several sources of noise, such as the offset and auto-fluorescence contributions. We applied baseline correction using the 'msbackadj' with a default value, which is a function of the Bioinformatics Toolbox of MATLAB (See Figure 3-5 A3, B3, C3 and D3).

Figure 3-6 Raman spectra of EV subtypes. The curves are normalized using the feature scaling method to enable comparison of the spectra in same scale. Each EV fingerprint shows spectral differences across the fingerprint area. The shaded area shows the main contribution to the separation by PCA. The spectra are vertically segregated for clarity purpose. The high frequency region (right) also shows small differences between EV subtypes.

For the multivariate analysis, we selected the spectral fingerprint region of each EV subtype, which is the range of 400 – 1800 cm^{-1} . The pre-processed spectra were normalized using unity-based normalization (feature scaling) for PCA. It brings all values into the range of 0 to 1, which prevents the emergence of artifacts as a result of variations in the intensity. PCA was performed with the function in MATLAB.

3.3 Result and discussion

Figure 3-6 shows the averaged Raman spectra of EV subtypes in 400 – 1800 (left; fingerprint region) and 2700-3050 cm^{-1} (right; high-frequency region) after baseline correction and normalization. Data pre-processing was required due to the weak Raman contribution of EVs. Spectral features were observed in both the fingerprint and the high-frequency region. Figure 4 shows the spectral differences across the EV subtypes (670-770, 998, 1146-1380, 1504-1590 and 1710-1780 cm^{-1} in fingerprint, 2834-2897 and 1985-3025 cm^{-1} in high frequency region). Each EV subtype showed distinctive spectral features; for example, lipid contents at 2847 & 2876 cm^{-1} , protein contribution at 2932 cm^{-1} , CH_2 deformation in lipids at 1296 cm^{-1} , CH_2 & CH_3 deformation in proteins & lipids at 1440

cm^{-1} , phenylalanine at 1603 cm^{-1} , amide II at 1544 cm^{-1} and C=C stretching in lipids 1650 cm^{-1} [30-32].

Multivariate analysis using PCA was conducted on the Raman spectra of four different EV subtypes (red blood cell-, platelet-, PC3- and LNCaP-derived EVs). The Raman spectra in the range of 400 to 1800 cm^{-1} with a 2 cm^{-1} interval ($n=300$; 654 variables) were selected for PCA. We also performed PCA in the high-frequency region (2700 to 3050 cm^{-1} ; 113 variables) and full spectra (400 to 3050 cm^{-1} ; 1152 variables), but these resulted in a weaker separation (see Figure 3-7 D ~ I). PCA in the fingerprint region performed better. The PCA score plot of PComp1 (87.47% of data variance) vs. PComp2 (5.27% of data variance), PComp3 (1.36% of data variance) vs. PComp1 and PComp3 vs. PComp2 are shown in Figure 3-7 (A), (B) and (C), respectively. Hematopoietic cell-derived EVs are marked with circles and cancer-derived EVs with triangles. In Figure 3-7 (B) and (C), PCA score plots clearly separate the prostate cancer EV group from the healthy EV group with 94.67% and 98%, respectively, of the data being accurately classified. This result indicates the clear discrimination of these two groups based on their spectral fingerprints. PCA loading plots show that those classifications might be contributed by a peak at 750 cm^{-1} (lactic acid) and a spectral band between 1500 and 1700 cm^{-1} , which contains the contribution of phenylalanine at 1603 cm^{-1} and amide II at 1544 cm^{-1} (see Figure 3-8). However, some EVs were not well separated. This could be caused by the heterogeneous nature of cancer EVs and/or the low signal-to-noise ratio of EV spectra. Despite these limitations, the result of multivariate analysis suggests the need for further study on EVs detection and recognition for disease monitoring.

Figure 3-7 PCA score plots for the Raman spectra obtained from four EV subtypes (red blood cell-EVs ●, platelet-EVs ●, PC3-EVs ▲ and LNCaP-EVs ▲). Circles represent blood cell-derived EVs and triangles show cancer-derived EVs. (A)~(C) were performed on the fingerprint region ($400 - 1800 \text{ cm}^{-1}$), (D)~(F) were performed on the high frequency region ($2700 - 3050 \text{ cm}^{-1}$) and (G)~(I) were performed on the full spectrum ($400 - 3050 \text{ cm}^{-1}$). Panels (B) and (C) show a good separation among EVs with various cellular origins. Principal component 1 (PComp1), PComp2 and PComp3 account for 87.47%, 5.27% and 1.36% of the total variance, respectively. (A) Score plot for PComp1 and Pcomp2, (B) score plot for PCom1 and PComp3 and (C) score plot for PComp2 and PComp3. In panel B and C, 94.67% and 98% of the data is classified into two categories, respectively; one containing the healthy cell-derived EVs and the other one the prostate cancer-derived EVs.

Figure 3-8 PCA loading plots corresponding to (A) PComp1, (B) PComp2 and (C) PComp3.

3.4 Conclusion

In conclusion, we have explored spectral differences between cancer-derived EVs and healthy control- derived EVs to examine the potential of cancer-derived EVs as a cancer biomarker. To clarify the role of EVs as a disease biomarker, Raman spectroscopy was employed to obtain the spectral fingerprint of EV subtypes. We obtained 300 data sets from four EV subtypes, and Raman spectra of EVs were analyzed with PCA to classify vesicle subtypes. Based on the principal component of the spectral information, the result of the multivariate analysis shows the spectral differences between healthy cells derived EVs (red blood cell and platelet) and prostate cancer cell-derived EVs (PC3 and LNCaP). The PC score plot shows that more than 90% of EVs were classified into two categories. This result suggests the potential of EVs as a cancer biomarker and makes them worthy of investigating further.

Bibliography

- [1] J. O. Park, D. Y. Choi, D. S. Choi, H. J. Kim, J. W. Kang, J. H. Jung, J. H. Lee, J. Kim, M. R. Freeman, K. Y. Lee, "Identification and characterization of proteins isolated from microvesicles derived from human lung cancer pleural effusions", *Proteomics*, **2013**; *13*, 2125.
- [2] E. van der Pol, A. N. Böing, P. Harrison, A. Sturk, R. Nieuwland, "Classification, functions, and clinical relevance of extracellular vesicles", *Pharmacological Reviews*, **2012**; *64*, 676.
- [3] M. Verma, T. K. Lam, E. Hebert, R. L. Divi, "Extracellular vesicles: potential applications in cancer diagnosis, prognosis, and epidemiology", *BMC Clinical Pathology*, **2015**; *15*, 6.
- [4] M.-P. Caby, D. Lankar, C. Vincendeau-Scherrer, G. Raposo, C. Bonnerot, "Exosomal-like vesicles are present in human blood plasma", *International Immunology*, **2005**; *17*, 879.
- [5] T. Pisitkun, R.-F. Shen, M. A. Knepper, "Identification and proteomic profiling of exosomes in human urine", *Proceedings of the National Academy of Sciences of the United States of America*, **2004**; *101*, 13368.
- [6] C. Admyre, S. M. Johansson, K. R. Qazi, J.-J. Filén, R. Lahesmaa, M. Norman, E. P. Neve, A. Scheynius, S. Gabrielsson, "Exosomes with immune modulatory features are present in human breast milk", *The Journal of immunology*, **2007**; *179*, 1969.
- [7] K. Denzer, M. J. Kleijmeer, H. Heijnen, W. Stoorvogel, H. J. Geuze, "Exosome: from internal vesicle of the multivesicular body to intercellular signaling device", *Journal of Cell Science*, **2000**; *113*, 3365.
- [8] A. Beach, H.-G. Zhang, M. Z. Ratajczak, S. S. Kakar, "Exosomes: an overview of biogenesis, composition and role in ovarian cancer", *Journal of ovarian research*, **2014**; *7*, 14.
- [9] C. Kahlert, S. A. Melo, A. Protopopov, J. Tang, S. Seth, M. Koch, J. Zhang, J. Weitz, L. Chin, A. Futreal, "Identification of double-stranded genomic DNA spanning all chromosomes with mutated KRAS and p53 DNA in the serum exosomes of patients with pancreatic cancer", *Journal of Biological Chemistry*, **2014**; *289*, 3869.
- [10] I. Tatischeff, E. Larquet, J. M. Falcón-Pérez, P.-Y. Turpin, S. G. Kruglik, "Fast characterisation of cell-derived extracellular vesicles by nanoparticles tracking analysis,

- cryo-electron microscopy, and Raman tweezers microspectroscopy", *Journal of extracellular vesicles*, **2012**; *1*, 19179.
- [11] F. Andre, N. E. Scharztz, M. Movassagh, C. Flament, P. Pautier, P. Morice, C. Pomel, C. Lhomme, B. Escudier, T. Le Chevalier, "Malignant effusions and immunogenic tumour-derived exosomes", *The Lancet*, **2002**; *360*, 295.
- [12] G. Rabinowits, C. Gerçel-Taylor, J. M. Day, D. D. Taylor, G. H. Kloecker, "Exosomal microRNA: a diagnostic marker for lung cancer", *Clinical Lung Cancer*, **2009**; *10*, 42.
- [13] F. Coumans, C. J. M. Doggen, G. Attard, J. De Bono, L. W. M. M. Terstappen, "All circulating EpCAM+ CK+ CD45– objects predict overall survival in castration-resistant prostate cancer", *Annals of Oncology*, **2010**; *21*, 1851.
- [14] X. Song, R. D. Airan, D. R. Arifin, A. Bar-Shir, D. K. Kadayakkara, G. Liu, A. A. Gilad, P. C. van Zijl, M. T. McMahon, J. W. Bulte, "Label-free in vivo molecular imaging of underglycosylated mucin-1 expression in tumour cells", *Nature communications*, **2015**; *6*, 6719.
- [15] C. W. Freudiger, W. Min, B. G. Saar, S. Lu, G. R. Holtom, C. He, J. C. Tsai, J. X. Kang, X. S. Xie, "Label-free biomedical imaging with high sensitivity by stimulated Raman scattering microscopy", *Science*, **2008**; *322*, 1857.
- [16] M. Miljković, T. Chernenko, M. J. Romeo, B. Bird, C. Matthäus, M. Diem, "Label-free imaging of human cells: algorithms for image reconstruction of Raman hyperspectral datasets", *Analyst*, **2010**; *135*, 2002.
- [17] R. P. Carney, S. Hazari, M. Colquhoun, D. Tran, B. Hwang, M. S. Mulligan, J. D. Bryers, E. Girda, G. S. Leiserowitz, Z. J. Smith, K. S. Lam, "Multispectral Optical Tweezers for Biochemical Fingerprinting of CD9-Positive Exosome Subpopulations", *Analytical Chemistry*, **2017**; *89*, 5357.
- [18] C. V. Raman, "A new radiation", *Indian Journal of Physics*, **1928**; *2*, 387.
- [19] K. Kneipp, H. Kneipp, I. Itzkan, R. R. Dasari, M. S. Feld, "Ultrasensitive chemical analysis by Raman spectroscopy", *Chemical Reviews*, **1999**; *99*, 2957.
- [20] T. Vankeirsbilck, A. Vercauteren, W. Baeyens, G. Van der Weken, F. Verpoort, G. Vergote, J. P. Remon, "Applications of Raman spectroscopy in pharmaceutical analysis", *TRAC Trends in Analytical Chemistry*, **2002**; *21*, 869.
- [21] L. A. Austin, S. Osseiran, C. L. Evans, "Raman technologies in cancer diagnostics", *The Analyst*, **2016**; *141*, 476.

- [22] R. E. Kast, S. C. Tucker, K. Killian, M. Trexler, K. V. Honn, G. W. Auner, "Emerging technology: applications of Raman spectroscopy for prostate cancer", *Cancer and Metastasis Reviews*, **2014**; 33, 673.
- [23] A. S. Haka, K. E. Shafer-Peltier, M. Fitzmaurice, J. Crowe, R. R. Dasari, M. S. Feld, "Diagnosing breast cancer by using Raman spectroscopy", *Proceedings of the National Academy of Sciences of the United States of America*, **2005**; 102, 12371.
- [24] Y. H. Ong, M. Lim, Q. Liu, "Comparison of principal component analysis and biochemical component analysis in Raman spectroscopy for the discrimination of apoptosis and necrosis in K562 leukemia cells", *Optics Express*, **2012**; 20, 22158.
- [25] Z. Huang, A. McWilliams, H. Lui, D. I. McLean, S. Lam, H. Zeng, "Near-infrared Raman spectroscopy for optical diagnosis of lung cancer", *International Journal of Cancer*, **2003**; 107, 1047.
- [26] H. Abdi, L. J. Williams, "Principal component analysis", *Wiley interdisciplinary reviews: computational statistics*, **2010**; 2, 433.
- [27] A. N. Böing, E. van der Pol, A. E. Grootemaat, F. A. W. Coumans, A. Sturk, R. Nieuwland, "Single-step isolation of extracellular vesicles by size-exclusion chromatography", *Journal of Extracellular Vesicles*, **2014**; 3, 23430.
- [28] L. Hartsuiker, PhD, "Microspectroscopic characterisation of gold nanorods for cancer cell detection", *University of Twente*, **2011**.
- [29] K. Ajito, K. Torimitsu, "Single nanoparticle trapping using a Raman tweezers microscope", *Applied Spectroscopy*, **2002**; 56, 541.
- [30] Z. J. Smith, C. Lee, T. Rojalin, R. P. Carney, S. Hazari, A. Knudson, K. Lam, H. Saari, E. L. Ibañez, T. Viitala, "Single exosome study reveals subpopulations distributed among cell lines with variability related to membrane content", *Journal of extracellular vesicles*, **2015**; 4, 28533.
- [31] H. Abramczyk, J. Surmacki, M. Kopeć, A. Olejnik, K. Lubecka-Pietruszewska, K. Fabianowska-Majewska, "The role of lipid droplets and adipocytes in cancer. Raman imaging of cell cultures: MCF10A, MCF7, and MDA-MB-231 compared to adipocytes in cancerous human breast tissue", *Analyst*, **2015**; 140, 2224.
- [32] Z. Movasaghi, S. Rehman, I. U. Rehman, "Raman spectroscopy of biological tissues", *Applied Spectroscopy Reviews*, **2007**; 42, 493.

Classifying Raman Spectra of Extracellular Vesicles based on Convolutional Neural Networks for Prostate Cancer Detection

Since the early 2000s, machine learning algorithms have been widely used in many research and industrial fields; most prominently in computer vision. Lately, many fields of study have tried to use these automated methods and there are several reports from the field of spectroscopy. In this study, we demonstrate a classification model based on machine learning to classify Raman spectra. We obtained Raman spectra from extracellular vesicles (EVs) to find tumor-derived EVs. The convolutional neural network (CNN) was trained on preprocessed Raman data and raw Raman data. We compare the result from CNN with results from principal component analysis (PCA), which is widely used among spectroscopists. The new model classifies EVs with an accuracy of $> 90\%$. Moreover, the new model based on CNN is also suitable for classifying the raw Raman data directly without pre-processing with an accuracy exceeding 93% .

Part of this chapter has been published in Journal of Raman Spectroscopy, 2019, DOI: 10.1002/jrs.5770.

Classifying Raman Spectra of Extracellular Vesicles based on Convolutional Neural Networks for Prostate Cancer Detection

Wooje Lee, Aufried T.M. Lenferink, Cees Otto and Herman L. Offerhaus

4.1 introduction

Raman spectroscopy allows extracting chemical information from a sample without labeling[1,2]. When we obtain the vibrational spectrum from a pure sample like toluene, ethanol, or silicon, we can readily identify the chemical contents. In real life, samples are unlikely to include only one pure chemical component[3-5]. Especially in clinical or biological applications, samples include very many different types of molecules indicative of their function or cellular origin[3,6-10]. Thus, we obtain very complex Raman spectra, and analyzing spectral data requires an extended effort [4,5,11-13].

4

To analyze these deeply convoluted data, principal component analysis (PCA) has commonly been employed. PCA is mostly used to reduce the dimension of the data and to make a prediction model. PCA calculates principal components of the data and projects the given data onto a newly generated coordinate system[14,15]. PCA shows optimal performance if spectral data are linearly correlated to their chemical content. Spontaneous Raman data is linear in the first approximation, but practical Raman spectra are unlikely to be only linear since they contain background and other features that disturb the scaling. Therefore signal processing is generally a prerequisite and this can bias the result of PCA[12,13].

The main sources of background signal in a biological sample are (1) autofluorescence [4,5,12,13], and (2) the suspension solution and sample containers, such as PBS or the cell culture medium. These solutions can contribute to the Raman spectrum such as peaks or bands [12]. Background signals strongly affect the result of the analysis so that we need to perform a rigorous background removal to avoid biasing the analysis result. However, removing background and noise, partially associated with the background, is challenging since the background subtraction itself can induce peak shifting, leaning, rejection of small peaks and distortion. It is difficult to distinguish background noise from a useful Raman spectrum in a complex sample. Since the background removal is both essential and a source of errors, using a type of data analysis that can automatically handle raw data directly is useful. We can use computing power where it is most useful and we can eliminate a human source of artifacts.

In Chapter 2, we classified Raman spectra obtained from extracellular vesicles (EVs) to detect prostate cancer without labeling[16]. EVs are small spherical particles (about 30 nm ~ 1 μ m in diameter) secreted by mammalian cells[17,18]. Cells, in general, will contain many different molecules depending on their function or disease state. Thus, EVs originating from diseased cells are likely to have different cellular content compared to EVs

derived from healthy cells[17,19-21]. Thus, finding the cellular origin of EV's gives us an insight and potential route to disease diagnosis.

We obtained Raman spectra from four types of extracellular vesicles (EVs); two out of four are blood product derived EVs and the other two are prostate cancer cell line derived EVs (also known as tumor-derived EVs, or tdEVs). In this research, we are aiming to discriminate tdEVs from blood as a disease diagnostic tool. Healthy people should not have tdEVs or only a very low presence of tdEVs in their body fluids[22]. Hematopoietic cell-derived EVs are always present in healthy people and patients. Hence, we aim to distinguish between hematopoietic cell-derived EVs and tdEVs using an algorithmic analysis technique.

In the previous chapter, we demonstrated that PCA (after pre-processing) can classify EVs depending on their cellular origin. With PCA, we were able to classify the Raman spectra with 95% accuracy using the spectral fingerprint region ($400\text{-}1800\text{ cm}^{-1}$) [16]. Although the method showed good results, it classified data that was preprocessed and that model is not suitable for raw data. Here, we propose a prediction model based on a machine learning (ML) algorithm. ML is widely used in computer vision[23,24], voice recognition and voice synthesis[25,26], and there were some attempts to use ML in the field of spectroscopy[27-31]. We demonstrate an ML-based prediction model to classify the Raman spectra of EVs without data preprocessing. Specifically, we use a Convolutional Neural Network (CNN) to build our prediction model.

The CNN [32-34] is inspired by the mammalian brain. Layers in the brain extract features from the input before this information enters the deeper areas of the brain for further processing. It was revealed by Hubel and Wiesel in the 1950s that feature extraction is used in pattern recognition tasks[35]. In 1998 LeCun et al. applied a feature extractor in their pioneer convolutional networks[36]. In that neural network algorithm, the “feature extracting layer” is known as the “convolution layer,” and it precedes the feed-forward neural network. The convolution layer in CNN allows the model to extract small details and to be trained on the extracted details of the input data, which improves its prediction accuracy[36,37]. Since then, CNNs have been widely used for image recognition or image classification. There have also been several demonstrations of the use of artificial intelligence (AI) algorithm to study Raman spectral data[27-31] as well as different types of spectroscopic data[38-41]. In this chapter, we will discuss neural networks and we will suggest a platform for Raman signature classification of EVs based on CNN. The platform

approach will provide an automated and robust classification tool for potential prostate cancer biomarker detection.

Figure 4-1 Structure of biological neuron (top) and perceptron (bottom). Neuron in the artificial neural network (ANN) also has similar structure.

4.2 Introduction to Neural Network

In 1957, the perceptron algorithm was invented by Rosenblatt *et al*, which was one of the very early versions of a machine learning algorithm[42]. The perceptron is a simplified model of neurons in the mammalian brain. It is a simple binary classifier that can decide yes or no. The structure of the perceptron and the biological neuron is described and compared in Figure 4-1. The perceptron has an input vector x and a vector w that contains real-valued weights. Its output value $f(x)$ is defined as the sum of $w \cdot x$. Then, the final output, y , of the algorithm is decided by a threshold or an activation function. The working principle of the perceptron algorithm can be described as

$$f(x) = \sum_{i=1}^n W_i \cdot x_i + b, \quad 4-1$$

$$y = \begin{cases} 1 & \text{if } f(x) > \text{threshold,} \\ 0 & \text{otherwise} \end{cases}, \quad 4-2$$

where x is an input vector, W is a vector of connection weights and b is a constant called bias. The output value y is compared to a label. If the algorithm gives the wrong answer, the weights are adjusted manually to get the correct answer.

The early algorithms performed only for linearly separable data. In 1969, Marvin Minsky suggested a multilayer perceptron (MLP) model for solving more complex problems in his book 'Perceptron'[43]. The proposed MLP consists of multiple layers with multiple perceptrons in each layer. These days, the concept of MLP has been widely used for many different types of machine learning algorithms.

4.3 Convolutional Neural Network

Machine learning algorithms can be separated into two categories: supervised learning and unsupervised learning. Unsupervised learning algorithms use given data to find underlying patterns/structures in the data without example. Principal Component Analysis (PCA) and cluster analysis belong to unsupervised learning. The unlikely unsupervised learning algorithm, supervised learning algorithms find a particular solution with pre-labeled data. In this type of algorithm, the output value of the model is compared to a pre-labeled output (target) value to calculate an error (also known as cost) of the model. Then, the algorithm uses the calculated error to minimize the distance between its output value and the correct answer. In this Chapter, we use a supervised learning algorithm, Convolutional Neural Networks (CNN), to identify the Raman spectra of EVs.

In section 4.3.1, the operating principle of CNN will be discussed in detail with the structure used for this study

Figure 4-2 Simplified schematic diagram of the 1-dimensional CNN used in this work. The model has three convolution-maxpooling layers alternatingly and a fully connected network with four hidden layers. The convolution-max pooling layer extracts features from the input spectrum and the fully connected layer is trained on the extracted features.

4.3.1 Convolution Layer

The CNN architecture that is used in this study is illustrated in Figure 4-2. The network has three convolution layers with a max-pooling layer for each convolution layer for feature extraction. The feature extractor is followed by a fully connected network for learning on the extracted features. The output from the fully connected layers is normalized by softmax into a probability distribution which is the set of probabilities of K possible outcomes. Thus, the normalized output must be in the interval $(0, 1)$. The networks were realized in Python (Python Software Foundation. Python Language Reference, version 3.6.6. Available at <http://www.python.org>) using Tensorflow (TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org).

Figure 4-3 A schematic diagram of convolution and max pooling layer. In the convolution layer a scalar product is performed of two vectors with input values and filter properties. Next the result of the scalar product is activated by Leaky ReLU and down sampled to reduce the size of the extracted feature.

Figure 4-3 shows a diagram of a convolution layer for 1-dimensional input data. In our CNN architecture, the input spectral data in the moving window is convoluted with $n \times 1$ filter(s) where the filter(s) determine the size of the moving window. Next, the convoluted input is activated by Leaky ReLU (rectified linear unit). Leaky ReLU is given by

$$f(x) = \begin{cases} x, & \text{if } x \geq 0 \\ ax, & \text{otherwise} \end{cases}, \quad 4-3$$

where x is the input to the neuron and the parameter a is typically less than or equal to one and greater than zero for ReLU. After convolution, the convoluted data is down-sampled by an operation known as max pooling. The max-pooling reduces the spatial dimension of the convoluted feature by selecting the maximum value in the moving window and allows the creation of a translation-invariant feature.

4.3.2 Feedforward Neural Network

The extracted feature will enter the fully connected (FC) layers. In this study, the FC network is a feed-forward neural network[44-46] with four hidden layers. The simple structure of a feedforward neural network is described in Figure 4-4. In the feed-forward neural network, I inputs are propagated to the adjacent hidden layer. This process is continued in every hidden layer until the end of the FC network. If the network has I inputs connected to the next hidden layer in which the layer has J neurons, the forward propagation can be expressed as

$$a_j = \tanh\left(\sum_{i=1}^I W_{ji} \cdot x_i + b_n\right) \quad j = 1, 2, 3, \dots, J, \quad 4-4$$

where W_{ji} is the weight for the i^{th} input to the j^{th} neuron, x_i is i^{th} input, b_n is the bias of n^{th} hidden layer and a_j is the output of j^{th} neuron in the hidden layer. The output of the hidden layer will be the input to the next hidden layer or the output of the FC network with K classes; we have four classes in this study. The output of the FC is given by

$$y_k = \sum_{j=1}^J W_{kj} \cdot a_j + b_{out} \quad k = 1, 2, 3, \dots, K, \quad 4-5$$

where W_{ji} are the weights connected to the output of the FC network, a_j is the output of the previous hidden layer, b_{out} is the bias of the output layer and y_k is the k -dimensional non-activated output of the FC. The output will be activated by the softmax function [24,47], Equation 4-6.

$$S(y_k) = \frac{e^{y_k}}{\sum_{m=1}^K e^{y_m}} \quad k = 1, 2, 3, \dots, K. \quad 4-6$$

The softmax calculates the probability distribution of the event over ‘K’ different events that sum to one. To train a model on a given input, the model calculates the distance between its prediction and the given label. The distance is called cost and the cost is calculated by the cross entropy function[47,48] written as:

$$D(S, L) = -\sum_{k=1}^K L_k \log(S(y_k)), \quad 4-7$$

where S is the probability of each class and L is the given label. Here, we used one-hot encoded label which means the character label is expressed as a vector, for example, ‘RBC-EVs’ is expressed as $[0 \ 0 \ 0 \ 1]$, ‘platelet-EVs’ is $[0 \ 0 \ 1 \ 0]$ and so on. The Adam optimizer updates weights based on the cost to minimize the distance between the prediction result and the target[49]. The outcome of the cross-entropy function will be closer to zero if the model is trained well. Then, the model propagates new data forward and the new cost will be backpropagated iteratively for the training.

There are two notable issues in the training of the neural network, namely vanishing gradients and overfitting. The vanishing gradient can be found in neural networks that use a gradient-based learning algorithm and backpropagation. In this type of learning algorithm, weights get updated in every iteration based on the partial derivative of the cost function with respect to current weights, for example

$$W_{ji}(n+1) := W_{ji}(n) - \eta \frac{\partial E}{\partial W_{ji}}, \quad 4-8$$

where $W_{ji}(n+1)$ is a new weight, $W_{ji}(n)$ is a current weight, η is a learning rate and E is the cost function. The vanishing gradient occurs when the gradient is too small, and the algorithm does not really update its weights. The learning process can slow down or stop completely. The other problem is overfitting. As the goal of the training, the network model changes its weights to reduce the cost. However, overfitting occurs when the network

focuses too much on irrelevant details of the given dataset. In this case, the trained network gives correct answers for the training dataset, but not when a new dataset is given as an input.

We applied the dropout method to avoid overfitting and vanishing gradients[50]. The dropout algorithm randomly selects 50% of the weights in each layer for every iteration, thereby varying the weights that determine the gradient. As part of the weights are used for training, we can minimize the overfitting and vanishing gradients problem. An early stop algorithm is also used to avoid the overfitting problem. For the early stop algorithm, we use one more dataset called the validation dataset. The validation dataset is not directly used for training, but it is used only for monitoring the training progress. When the network is overfitted to the training dataset, the cost of the training set keeps reducing. But, the validation cost increase at some point, Figure 4-5. If the model starts over-training, a gap between training cost and validation cost increases with each iteration. By monitoring the cost change, the algorithm can decide when to stop learning to prevent overfitting. The Python script used for the CNN training can be found in Appendix A.

Figure 4-4 Structure of the fully connected layer, the input layer and the 1st hidden layer (A). The last hidden layer and the output layer connected to the softmax function (B). The output of the hidden layer will be the input of the next hidden layer. If the network has n hidden layers, the structure described in panel A will be repeated $n-1$ times and last n^{th} layer will be connected to the output layer.

Figure 4-5 Training and validation cost curves. The red shaded area represents under fitting which means the model has to be trained more. The blue shaded area shows over fitting. The unshaded area or beginning of the blue shaded area can be regarded as the position where the model is well trained.

4.4 Raman spectral signature collection

The experimental method described in Chapter 2 was used to obtain the Raman spectra of EVs. As described in Chapter 2, four EVs subtypes (RBC, platelet, PC3, and LNCaP) were measured on a house-built Raman microscope. We obtained 300 spectra from the four EV subtypes (75 spectra per each subtype). Figure 4-6 shows averaged Raman spectra of each subtype. In the figure, the blue line shows the Raman spectrum after the background removal and the red curve represents averaged raw data. The raw data is shifted for clarity.

Figure 4-6 The averaged Raman spectra of EVs from (A) red blood cells-, (B) platelet- (C) PC3-, and (D) LNCaP-derived EVs. In each panel, the blue line represents preprocessed data and the red line shows raw data. The shaded area shows the standard deviation of the measurement.

Figure 4-7 Example of data augmentation. Randomly generated white Gaussian noise is added to the Raman spectrum of RBC-derived EVs. In the figure, the blue curve represents the original Raman signal. The red line shows the noise added signal and the noise is plotted separately in yellow.

4.5 Result and Discussion

We performed PCA and CNN both on baseline-corrected data and raw data. To make a prediction model based on PCA, the EVs' Raman data is divided into two subsets: the training and the testing set. The PCA training set consists of 240 spectra and the testing set has 60 spectra. To create the training set and testing set, the spectra are selected on a random basis from the four EV subtypes: 15 spectra from each EV subtypes were selected to make the testing dataset. PCA is done on the training dataset and validated on the training set. The PCA based prediction model was realized in MATLAB R2016b (version 9.1.0, The MathWorks, Natick, MA). The PCA and CNN model was trained on three different spectral regions to find the most relevant spectral area for the classification; 400-3050 cm^{-1} (full spectrum), 400-1800 cm^{-1} (fingerprint) and 2700-3050 cm^{-1} (high frequency, also known as C-H stretch region).

The 300 Raman spectra used in Chapter 2 were also used for the CNN based prediction model. Artificial neural network models usually require a large volume of data to achieve high accuracy and avoid overfitting to the given data. Moreover, the dimension of the Raman spectrum is 1152×1 . It means that the dimension of the training data is far bigger than the number of data points which can readily cause overfitting.

To solve this problem, we conducted data augmentation which is a commonly used method to increase the training data. For the data augmentation, we generated white Gaussian noise with SNR of 15, 25 and 30 and added it to the original signal, which was done using the '*awgn*' function provided by MATLAB. Figure 4-7 shows an example of data augmentation done for this research. After the augmentation, we had 1200 spectra; 300 original and 900 random noise added spectra. This dataset was divided into three subsets; a training, validation and testing dataset. We randomly selected 90 spectra from each subtype. The testing set was prepared from 50% of 90 spectra and the other 50% became a validation set. In the end, we had a training set of 840, a testing set of 180 and a validation set of 180 Raman spectra. The structure of the model for this particular analysis has three sets of convolution layers followed by a max-pooling layer. The feature extractor is followed by four hidden layers, which have 1000, 500, 200 and 200 neurons. The output layer has four neurons and is connected to softmax to convert output scores to a probability distribution. Prior to the network training, all the weights of the network were initialized on a random basis. We assigned the weights with the Xavier initializer, which assigns weights from a Gaussian distribution[46]. The initialization method keeps the variance of the weights the same in each hidden layer.

Table 4-1 Prediction accuracy of PCA on the preprocessed dataset

Spectral region (cm ⁻¹)	LDA	QDA
400-3050	0.6500	0.7833
400-1800	0.8333	0.9500
2700-3050	0.7333	0.8667

Table 4-2 Prediction accuracy of PCA on raw data

Spectral region (cm ⁻¹)	LDA	QDA
400-3050	0.6167	0.6833
400-1800	0.6167	0.6167
2700-3050	0.5833	0.6000

Table 4-3 Classification accuracy of the CNN based prediction model.

Spectral region (cm ⁻¹)	Preprocessed data	Raw data
400-3050	0.9089 ± 0.0101	0.9522 ± 0.0101
400-1800	0.9022 ± 0.0050	0.9656 ± 0.0091
2700-3050	0.9122 ± 0.0120	0.9311 ± 0.0050

The CNN training time was about 10 to 70 minutes depending on the input dimension and all the training was done by a simulation computer which is powered by an Intel Core i7-8700, NVIDIA GTX1080Ti, 500 GB SSD and 32GB RAM. For the parallelized calculation, a graphic processor unit (GPU) was mainly used for the model training.

Table 4-1 and Table 4-2 show the classification result of PCA-LDA and PCA-QDA on preprocessed data (Table 4-1) and raw data (Table 4-2). In Table 4-1, PCA-QDA shows a fairly good classification, especially in the fingerprint region. However, the results show that the PCA based model has a high prediction/classification accuracy under certain conditions; it performed well on the fingerprint region ($400\text{-}1800\text{cm}^{-1}$) of the background-corrected Raman data. In general, however, the PCA model classified the Raman spectra of the EVs poorly for the full spectral width and in the high-frequency region. Table 4-2 shows the result of PCA-LDA and PCA-QDA trained on untreated data. As can be seen in the table, the classification accuracy of PCA on raw data is very low. The result shows that PCA is not suitable for handling raw Raman data and requires a decent background/noise removal process as discussed in the introduction.

We trained the CNN model on preprocessed and raw Raman data and Table 4-3 shows the prediction accuracy on both datasets. In both cases, the prediction accuracy is higher than 90%. Originally, we assumed that the CNN trained on the cleaner signal would show a better classification accuracy since, after removing the background contribution, the remaining data should have a cleaner EV contribution. While the model trained on preprocessed data classified the spectra with an accuracy of 90.89% over the full spectral area, 90.22% for the fingerprint and 91.22% for the high frequency, the model trained on raw Raman data shows better prediction accuracies of 95.22% for the full spectral region, 96.56% for the fingerprint region and 93.11% for the high-frequency region. We attribute this improvement to small signals buried in the untreated spectral data.

The mean size of the EVs used in this study is about 150 nm[16]. The single-particle is therefore about 100-fold smaller than the focal volume of the Raman microscope. Thus, the solution in which the EVs are suspended contributes to the Raman signal more than the trapped particles do, which leads to poor SNR of the Raman spectra of EVs; about 7dB. At such an SNR, small spectral features can be concealed by the background contribution and small peaks might be eliminated by background correction. The raw data retains such

small spectral information that is not clearly visible in the raw Raman spectrum. This subtle information allows the CNN model to learn more details about the input signal.

We have tried identifying the most relevant spectral regions that contain most of the meaningful information for the classification. The result in Table 4-3 shows that every spectral segment used in this study shows high accuracy of 95.22%, 96.56% and 93.11% in fingerprint ($400\text{-}1800\text{ cm}^{-1}$), high frequency ($2700\text{-}3050\text{ cm}^{-1}$) and full spectrum ($400\text{-}305\text{ cm}^{-1}$), respectively. The fact that the model trained on fingerprint performed better suggests that the fingerprint region ($400\text{-}1800\text{ cm}^{-1}$) has more relevant information for the classification than the high-frequency region ($2700\text{-}3050\text{ cm}^{-1}$). However, it does not imply that spectral information in the high-frequency region is less important than information in the spectral fingerprint region. While protein and lipid contribution are more prominent in the high-frequency region, many other biomolecules contribute to the fingerprint region. The PCA model shows a similar result, namely a classification accuracy on the fingerprint and high-frequency region of 95.00% and 86.67%, respectively.

4.6 Conclusion

In this chapter, we have demonstrated that a CNN based prediction model can be used as a classifier of Raman spectra of EVs and that the model is suitable for raw data handling. The study shows that a PCA based prediction model can classify spectral data by EVs' cellular origin, but its classification ability is limited by background noise. On the other hand, the CNN model suggested in this paper shows a better classification accuracy ($> 90\%$) on both preprocessed data and raw data. Interestingly, the model trained on raw data classifies the Raman spectra of EVs better than the model trained on preprocessed data. It suggests that the use of raw data is useful for the classification and that the raw data preserves features so that the result is improved and background subtraction can be omitted.

Bibliography

- [1] C. V. Raman, "A new radiation", *Indian Journal of Physics*, **1928**; 2, 387.
- [2] K. Kneipp, H. Kneipp, I. Itzkan, R. R. Dasari, M. S. Feld, "Ultrasensitive chemical analysis by Raman spectroscopy", *Chemical Reviews*, **1999**; 99, 2957.
- [3] Y. H. Ong, M. Lim, Q. Liu, "Comparison of principal component analysis and biochemical component analysis in Raman spectroscopy for the discrimination of apoptosis and necrosis in K562 leukemia cells", *Optics Express*, **2012**; 20, 22158.
- [4] R. Petry, M. Schmitt, J. Popp, "Raman spectroscopy—a prospective tool in the life sciences", *Chemphyschem*, **2003**; 4, 14.
- [5] S. Nie, S. R. Emory, "Probing single molecules and single nanoparticles by surface-enhanced Raman scattering", *Science*, **1997**; 275, 1102.
- [6] X. Song, R. D. Airan, D. R. Arifin, A. Bar-Shir, D. K. Kadayakkara, G. Liu, A. A. Gilad, P. C. van Zijl, M. T. McMahon, J. W. Bulte, "Label-free in vivo molecular imaging of underglycosylated mucin-1 expression in tumour cells", *Nature communications*, **2015**; 6, 6719.
- [7] L. A. Austin, S. Osseiran, C. L. Evans, "Raman technologies in cancer diagnostics", *The Analyst*, **2016**; 141, 476.
- [8] R. E. Kast, S. C. Tucker, K. Killian, M. Trexler, K. V. Honn, G. W. Auner, "Emerging technology: applications of Raman spectroscopy for prostate cancer", *Cancer and Metastasis Reviews*, **2014**; 33, 673.
- [9] A. S. Haka, K. E. Shafer-Peltier, M. Fitzmaurice, J. Crowe, R. R. Dasari, M. S. Feld, "Diagnosing breast cancer by using Raman spectroscopy", *Proceedings of the National Academy of Sciences of the United States of America*, **2005**; 102, 12371.
- [10] Z. Huang, A. McWilliams, H. Lui, D. I. McLean, S. Lam, H. Zeng, "Near-infrared Raman spectroscopy for optical diagnosis of lung cancer", *International Journal of Cancer*, **2003**; 107, 1047.
- [11] R. Gautam, S. Vanga, F. Ariese, S. Umapathy, "Review of multidimensional data processing approaches for Raman and infrared spectroscopy", *EPJ Techniques and Instrumentation*, **2015**; 2, 8.

-
- [12] F. Bonnier, S. M. Ali, P. Knief, H. Lambkin, K. Flynn, V. McDonagh, C. Healy, T. Lee, F. M. Lyng, H. J. Byrne, "Analysis of human skin tissue by Raman microspectroscopy: dealing with the background", *Vibrational Spectroscopy*, **2012**; *61*, 124.
- [13] J. Zhao, H. Lui, D. I. McLean, H. Zeng, "Automated autofluorescence background subtraction algorithm for biomedical Raman spectroscopy", *Applied Spectroscopy*, **2007**; *61*, 1225.
- [14] H. Abdi, L. J. Williams, "Principal component analysis", *Wiley interdisciplinary reviews: computational statistics*, **2010**; *2*, 433.
- [15] S. Wold, K. Esbensen, P. Geladi, "Principal component analysis", *Chemometrics and Intelligent Laboratory Systems*, **1987**; *2*, 37.
- [16] W. Lee, A. Nanou, L. Rikkert, F. A. W. Coumans, C. Otto, L. W. M. M. Terstappen, H. L. Offerhaus, "Label-Free Prostate Cancer Detection by Characterization of Extracellular Vesicles Using Raman Spectroscopy", *Analytical Chemistry*, **2018**; *90*, 11290.
- [17] E. van der Pol, A. N. Böing, P. Harrison, A. Sturk, R. Nieuwland, "Classification, functions, and clinical relevance of extracellular vesicles", *Pharmacological Reviews*, **2012**; *64*, 676.
- [18] J. Park, M. Hwang, B. Choi, H. Jeong, J.-h. Jung, H. K. Kim, S. Hong, J.-h. Park, Y. Choi, "Exosome classification by pattern analysis of surface-enhanced Raman spectroscopy data for lung cancer diagnosis", *Analytical Chemistry*, **2017**; *89*, 6695.
- [19] M. Verma, T. K. Lam, E. Hebert, R. L. Divi, "Extracellular vesicles: potential applications in cancer diagnosis, prognosis, and epidemiology", *BMC Clinical Pathology*, **2015**; *15*, 6.
- [20] F. Andre, N. E. Scharzt, M. Movassagh, C. Flament, P. Pautier, P. Morice, C. Pomel, C. Lhomme, B. Escudier, T. Le Chevalier, "Malignant effusions and immunogenic tumour-derived exosomes", *The Lancet*, **2002**; *360*, 295.
- [21] G. Rabinowits, C. Gerçel-Taylor, J. M. Day, D. D. Taylor, G. H. Kloecker, "Exosomal microRNA: a diagnostic marker for lung cancer", *Clinical Lung Cancer*, **2009**; *10*, 42.
- [22] A. Nanou, F. A. Coumans, G. van Dalum, L. L. Zeune, D. Dolling, W. Onstenk, M. Crespo, M. S. Fontes, P. Rescigno, G. Fowler, "Circulating tumor cells, tumor-derived extracellular vesicles and plasma cytokeratins in castration-resistant prostate cancer patients", *Oncotarget*, **2018**; *9*, 19283.

- [23] E. Rosten, T. Drummond In *European conference on computer vision*; Springer: **2006**, p 430.
- [24] C. M. Bishop *Pattern recognition and machine learning*; springer, **2006**.
- [25] S. Kang, X. Qian, H. Meng In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*; IEEE: **2013**, p 8012.
- [26] L. Muda, M. Begam, I. Elamvazuthi, "Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques", *arXiv preprint arXiv:1003.4083*, **2010**.
- [27] S. Sigurdsson, P. A. Philipsen, L. K. Hansen, J. Larsen, M. Gniadecka, H.-C. Wulf, "Detection of skin cancer by classification of Raman spectra", *IEEE Transactions on Biomedical Engineering*, **2004**; *51*, 1784.
- [28] M. G. Madden, A. G. Ryder In *Opto-Ireland 2002: Optics and Photonics Technologies and Applications*; International Society for Optics and Photonics: **2003**; Vol. 4876, p 1130.
- [29] T. A. Dolenko, S. A. Burikov, A. V. Sugonjaev In *Opto-Ireland 2005: Optical Sensing and Spectroscopy*; International Society for Optics and Photonics: **2005**; Vol. 5826, p 298.
- [30] J. Liu, M. Osadchy, L. Ashton, M. Foster, C. J. Solomon, S. J. Gibson, "Deep convolutional neural networks for Raman spectrum recognition: a unified solution", *Analyst*, **2017**; *142*, 4067.
- [31] M. Jermyn, J. Desroches, J. Mercier, M.-A. Tremblay, K. St-Arnaud, M.-C. Guiot, K. Petrecca, F. Leblond, "Neural networks improve brain cancer detection with Raman spectroscopy in the presence of operating room light artifacts", *Journal of biomedical optics*, **2016**; *21*, 094002.
- [32] N. Aloysius, M. Geetha In *2017 International Conference on Communication and Signal Processing (ICCSP)*; IEEE: **2017**, p 0588.
- [33] R. Yamashita, M. Nishio, R. K. G. Do, K. Togashi, "Convolutional neural networks: an overview and application in radiology", *Insights into imaging*, **2018**; *9*, 611.
- [34] W. Rawat, Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review", *Neural Computation*, **2017**; *29*, 2352.
- [35] D. H. Hubel, T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex", *The Journal of physiology*, **1962**; *160*, 106.

-
- [36] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition", *Proceedings of the IEEE*, **1998**; *86*, 2278.
- [37] A. Krizhevsky, I. Sutskever, G. E. Hinton In *Advances in neural information processing systems* **2012**, p 1097.
- [38] J. Acquarelli, T. van Laarhoven, J. Gerretzen, T. N. Tran, L. M. Buydens, E. Marchiori, "Convolutional neural networks for vibrational spectroscopic data analysis", *Analytica Chimica Acta*, **2017**; *954*, 22.
- [39] S. Malek, F. Melgani, Y. Bazi, "One-dimensional convolutional neural networks for spectroscopic signal regression", *Journal of Chemometrics*, **2018**; *32*, e2977.
- [40] C. Yuanyuan, W. Zhibin, "Quantitative analysis modeling of infrared spectroscopy based on ensemble convolutional neural networks", *Chemometrics and Intelligent Laboratory Systems*, **2018**; *181*, 1.
- [41] K. Ghosh, A. Stuke, M. Todorović, P. B. Jørgensen, M. N. Schmidt, A. Vehtari, P. Rinke, "Deep learning spectroscopy: Neural networks for molecular excitation spectra", *Advanced Science*, **2019**; *6*, 1801367.
- [42] F. Rosenblatt *The perceptron, a perceiving and recognizing automaton Project Para*; Cornell Aeronautical Laboratory, **1957**.
- [43] M. Minsky, S. Papert, "Perceptron: an introduction to computational geometry", *The MIT Press, Cambridge, expanded edition*, **1969**; *19*, 2.
- [44] M. Minsky, S. A. Papert *Perceptrons: An introduction to computational geometry*; MIT press, **2017**.
- [45] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks", *Neural Networks*, **2004**; *2*, 985.
- [46] X. Glorot, Y. Bengio In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* **2010**, p 249.
- [47] I. Goodfellow, Y. Bengio, A. Courville *Deep learning*; MIT press, **2016**.
- [48] R. Y. Rubinstein, D. P. Kroese *The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation and machine learning*; Springer Science & Business Media, **2013**.
- [49] D. P. Kingma, J. Ba, "Adam: A method for stochastic optimization", *arXiv preprint arXiv:1412.6980*, **2014**.

- [50] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting", *The Journal of Machine Learning Research*, **2014**; 15, 1929.

Waveguide Raman Spectroscopy: Different Waveguide Materials and their Intrinsic Background

This chapter discusses Waveguide Raman spectroscopy (WGS). WGS uses the evanescent field outside the waveguide to probe an analyte on the surface of the chip. This makes it possible to study thin films or nanostructures on the waveguide. In this chapter, the working principle of WGS will be discussed. Waveguides for WGS were designed and fabricated using Si_3N_4 , Al_2O_3 and TiO_2 as the core material and the intrinsic background of each of the three waveguide materials was characterized. An analyte, toluene, was measured on these three different waveguide platforms. Both the background and the analyte were measured for quasi- transverse electric (quasi-TE) and quasi- transverse magnetic (quasi-TM) modes. TM modes generate less background than TE modes due to the lower overlap of the mode with the waveguide core material. Furthermore, TM modes exhibit a higher coupling efficiency of the generated Raman signal back into the waveguide. A lower background in combination with a higher coupling efficiency results in a higher signal-to-noise ratio (SNR) for the Raman signal of the analyte. The combination of Si_3N_4 and quasi-TM polarization led to the highest SNR in this study.

Part of this chapter has been submitted to OSA continuum and accepted by the journal.

Study on Multiple Waveguide Platforms for Waveguide Integrated Raman Spectroscopy

Wooje Lee, Pablo Muñoz Galindo, Ivo Hegeman, Yean-Sheng Yong, Meindert Dijkstra, Sonia M. García-Blanco and Herman L. Offerhaus

5.1 Introduction

Optical sensors and fiber optic sensors have some inherent advantages compared to their electrical counterparts: they can provide high selectivity, immunity to electromagnetic interference (EMI), a wide dynamic range, easy multiplexing and the possibility to be deployed in hostile environments thanks to their electrical and chemical inertness[1,2]. By measuring intensity, phase and/or refractive index changes, optical sensors can sense a wide variety of parameters, including temperature, pressure, vibration, and specific chemical species[1,3]. Optical sensors also have critical disadvantages, such as the fragility of the optical interface and (often) a high development cost. In the last few decades, optical sensors have been integrated on-chip leveraging existing semiconductor fabrication technologies. Integrated optical sensors not only enable miniaturization of these systems and mass production but also lead to more robust sensors. Since their conception, integrated optical sensors have attracted a great deal of interest[4-11]. However, the detection of different analytes is restricted to the availability of suitable capture layers, complicating the fingerprinting of unknown substances.

Raman spectroscopy is a useful analytical tool to identify molecules, chemical bonds and polymer structures of an analyte in a label-free and nondestructive manner. It is compatible with a wide range of samples that can be either liquid, solid or in a gas phase. Raman spectroscopy has a wide range of applications in analytical chemistry, food science[12], pharmaceutical industry[13], art and archaeology[14], biomedical studies and in the clinic[15,16]. Several studies have demonstrated the detection of single or multiple nanoparticles using Raman spectroscopy combined with optical tweezers [17-20]. Raman spectroscopy has also been used to study thin surface layers[21]. Surface-enhanced Raman spectroscopy allows collecting the Raman spectrum from thin-film surfaces or nanoparticles[22,23].

A clean Raman spectrum with a high SNR is required to separate the signal of a nanostructure or monolayer from the surrounding bulk background signals. In a free-space optical system, the resolution of the Raman microscope cannot be better than the diffraction limit, which sets a lower limit on the total probed volume. When a large surface must be studied, the volume scales with the numerical aperture to the fourth power so that the surface-to-volume ratio becomes increasingly unfavorable. To overcome these limitations, waveguide Raman (WGR) spectroscopy has been proposed. WGR uses the evanescent field above a waveguide to probe an analyte in close proximity to the waveguide. Furthermore, the interaction length between waveguide and analyte can be longer by

increasing the length of the waveguide that is limited by the propagation losses. WGR reduces the collection time 10-100 fold to achieve the same Raman signature as a conventional Raman microscope[24].

Since WGR was first demonstrated by Schlotter and Rabolt in 1984[25], WGR spectroscopy has proven to be a useful technique to study nanostructures. However, waveguide fabrication technologies proved to be a major bottleneck. Leveraging the advances in microfabrication technologies for integrated photonics, Dhakal et al. successfully demonstrated WGR detection of isopropyl alcohol (IPA) in transmission mode using Si_3N_4 waveguides and 785 nm excitation[24]. Since then, WGR has been shown to be a promising technique to investigate thin films and different liquid- and gas-phase molecules[26-30]. Although there were attempts to measure aqueous sample, thin film, gas phase molecules and polystyrene particles[31], measuring EVs using waveguide Raman has not been done yet.

Si_3N_4 was mainly used as the waveguide core material[26,29,30,32]. Si_3N_4 has a high refractive index, which increases the coupling efficiency of the generated Raman signal. Although some Si_3N_4 waveguides showed a low fluorescence background[32,33], many Si_3N_4 waveguides have a relatively high fluorescent background in the near infrared[34,35]. This inherent background can limit the SNR of the Raman signals, which has motivated the search for other high refractive index materials. Evans et al. studied WGR in TiO_2 waveguides[36]. Raza et al. compared four high refractive index contrast materials typically used in integrated photonics, namely Al_2O_3 , Si_3N_4 , TiO_2 , and Ta_2O_5 . In their study, the relatively low refractive index contrast of Al_2O_3 led to poor coupling efficiency, while the elevated background and high propagation losses of TiO_2 led to a poor SNR. Their study contained only TE polarization.

In this study, we compare the performance of Al_2O_3 , Si_3N_4 , and TiO_2 for both TE and TM polarization to find the most suitable platform for on-chip Raman spectroscopy. The TM polarization shifts part of the field outside the core of the waveguide. Therefore, a reduction in the background is expected. At the same time, the increased intensity at the core-cladding boundary enhances the coupling of the emitted Raman photons back into the waveguide core.

The background of each waveguide platform is measured for both TE and TM polarizations. This approach is proposed to measure EVs bound to the waveguide surface. Catching EVs on a surface requires surface chemistry that enables one to attach specific

antibodies to functionalize the surface. Due to the time limit, however, surface functionalization was not able to be done in the given time. Therefore, toluene will be used in this chapter and the next chapter for the proof of the concept. Since toluene has a very distinctive Raman fingerprint, it is believed that the chemical can be used for this purpose.

5.2 Experimental setup

We built a spontaneous Raman microscope and the in-house built setup was used for all the waveguide experiments. In contrast with the conventional Raman microscope, the custom-built microscope is neither an upright nor an inverted microscope. In this system, the incident beam is parallel to the optical table. This structure makes the user easy to couple the light into the waveguide and to mount a sample on the waveguide chip. The scattered photon can be collected by two different ways; backscattering configuration (Figure 5-1A) and vertical collection configuration (Figure 5-2B). These two options can be selected by a flipping mirror located behind the dichroic mirror.

Figure 5-1 describes the in-house built Raman microscope. The setup uses the 785 nm pump beam from a laser diode (LD-785-SE400; Thorlabs Inc., Newton, NJ) mounted on a temperature-controlled mount (TCLDM9; Thorlabs Inc.). The laser is driven by a laser controller (LDC-3722; ILX Lightwave, Bozeman, MT) that controls the temperature and the driving current. To achieve a stable laser output, we operated the laser at a driving current of 260 mA and a temperature of 25°C. We controlled the laser power with a half-wave plate and polarizing beam splitter. The pump beam was passed through a laser clean-up filter (LL01-785; Semrock Inc., Rochester, NY). The pump beam was then p- or s-polarized by a half-wave plate to excite either the TE₀₀ or TM₀₀ mode of the optical waveguide, respectively. The polarized pump light was coupled into the waveguide with infinity corrected microscope objective (40x/0.60; Nached, France). To optimize coupling, the Rayleigh scattering from the waveguide was monitored by a camera (BFLY-U3-23S6M-C; Point Grey Research Inc., Richmond, BC, Canada) from the top of the waveguide.

In the back-scattering configuration, the same objective lens was used to collect the light. The collected light passed through a dichroic mirror (LPD02-785RU-25; Semrock Inc.) and long pass filter (LD02-785RU-25; Semrock Inc.).

In the vertical collection mode, scattered photons are collected by an objective lens on top of the waveguide. This light passed through a beamsplitter (BS023; Thorlabs Inc.) and was

focused on an optical fiber (ZQP200-2-VIS-FC-BX; Ocean optics) by a lens (AC254-100-B; Thorlabs Inc.). The light coming out of the optical fiber was collimated (CFC-5X-B; Thorlabs Inc.) and passed through a long pass filter (LD02-785RU-25; Semrock Inc.).

The filtered light was then dispersed by a spectrograph with a 300 grooves/mm grating (IsoPlane 160; Princeton Instruments, Trenton, NJ, USA). The dispersed light was recorded on a 1002x1004 electron multiplying charge-coupled device (EMCCD) camera (Andor iXon DV885JCS-VP; Andor Instruments, Belfast, UK) cooled down to -70°C to reduce the dark noise generated by the imaging sensor.

Although the setup has a front-illuminated EMCCD (F-EMCCD), the setup was originally built with a back-illuminated EMCCD (B-EMCCD) camera (Andor iXon DV887ECS-BV; Andor Instruments). The B-EMCCD is widely used since it has higher quantum efficiency than the F-EMCCD. For the B-EMCCD, the photons first encounter the photosensitive layer whereas the incoming photon on the F-EMCCD first meets the electrodes, see Figure 5-2C. This structure causes the F-EMCCD to reflect some of the incoming photons, which result in lower quantum efficiency. The B-EMCCD however, has a critical drawback; an intrinsic interference pattern can show up as shown in Figure 5-2A. Since the photosensitive layer of the B-EMCCD is exposed to the air or vacuum, the CCD has a relatively high refractive index contrast step at the CCD-to-air interface. Therefore, the layer can act as a Fabry-Perot cavity which causes the interference pattern on the CCD. The photon that enters the F-EMCCD does not feel a big refractive index mismatch so that this CCD does show the interference pattern, Figure 5-2B.

Figure S-1 Schematic of the in-house built Raman microscope for the waveguide Raman experiments. The pump beam is in the horizontal plane and the coupling into the chip is monitored from the top by a CCD camera. The scattered photons can be collected in (A) the backscattering configuration and (B) the top-collection mode. The two collection modes cannot be used at the same time.

Figure 5-2 Broad band illumination, recorded on (A) the B-EMCCD and (B) F-EMCCD. The interference pattern is clearly visible on the B-EMCCD. A cartoon in the panel (C) shows the structure of the B-EMCCD (left) and the F-EMCCD (right) with a cross section view.

5.2.1 Wavenumber Calibration

The dispersed light on the EMCCD is recorded as photon-counts on each pixel of the camera. The pixel number must be converted to a wavenumber scale to compare the measurement to known spectra. For the calibration toluene and an ArHg lamp (HG-1; Ocean Optics, Duiven, the Netherlands) were used as references. The conversion of wavelength to wavenumber can be carried out using Equation 5-1 and 5-2,

$$\text{Raman shift (cm}^{-1}\text{)} = \left(\frac{1}{\lambda_{\text{ex}}} - \frac{1}{\lambda} \right) \times 10^7, \quad 5-1$$

$$\lambda = \frac{1}{\frac{1}{\lambda_{\text{ex}}} - \frac{\text{Raman shift (cm}^{-1}\text{)}}{10^7}}, \quad \text{5-2}$$

where λ_{ex} is the wavelength of the excitation beam that is 785 nm, and λ is the wavelength of the detected peak. The spectral lines of the references used for the calibration are summarized in Table 5-1.

The calibration protocol can be described as:

5

- 1) Recording the Raman spectrum of toluene and the emission lines of ArHg lamp.
- 2) Finding well-defined peaks of two references.
- 3) Mapping the pixel location of the peaks to wavenumber positions. This produces two-dimensional coordinates of pixels and wavenumber.
- 4) Polynomial fitting of the coordinates to get the relationship between each pixel and corresponding wavenumber.

The calibration protocol was implemented in MATLAB. To retrieve the peaks of interest, the 'findpeaks' function in MATLAB was used. The peaks of toluene and the ArHg were combined and were sorted by pixel numbers. In the mapping process, interpolation was performed to achieve subpixel accuracy. The order of the polynomial has to be carefully selected to avoid over- and under-fitting[37]. We used 5th order polynomial fitting which showed good performance. The Matlab script can be found in Appendix B.

Table 5-1 Toluene and ArHg peak locations used for the pixel-to-wavenumber calibration. For the fingerprint region, peaks above the dotted lines are used. The peaks below the dotted line are used for the high-frequency region

Toluene peaks		ArHg peaks	
wavelength (nm)	wavenumber (cm ⁻¹)	wavelength (nm)	wavenumber (cm ⁻¹)
785.000	0.000	794.818	157.357
818.474	521.000	800.616	248.471
825.392	623.400	811.531	416.465
836.683	786.900	826.452	638.937
852.090	1003.000	842.465	868.924
854.055	1030.000	852.144	1003.748
867.464	1211.000	866.794	1202.087
880.293	1379.000	912.297	1777.511
898.242	1606.000	922.450	1898.158
1018.397	2919.500	935.422	2048.491
1032.817	3056.600

Figure 5-3 (A) Raman spectrum of toluene and ArHg emission lines. Triangle markers and diamond markers represent the selected peaks of toluene and ArHg, respectively. (B) The toluene spectrum after wavenumber calibration. The dotted vertical lines show the wavenumbers where the toluene peaks are supposed to be located after the calibration.

5.3 Waveguides Design Consideration for on-chip Raman Spectroscopy

Monomode channel waveguides were designed for three materials, Al_2O_3 , Si_3N_4 , and TiO_2 for operation in both the TE and TM polarization. Channel waveguides were selected because this geometry maximizes the contact area with the analyte compared to other structures such as slab waveguides or rib waveguides. Furthermore, channel waveguides support smaller bending radii which is advantageous for long spiral waveguides that increase the interaction length with the analyte.

Lumerical MODE solutions (v7.11.1541, Lumerical Inc., Vancouver, BC, Canada) was used for eigenmode calculations on the different waveguide structures. A Cauchy model was used to represent the dispersion relation for the different materials. The Cauchy's equation is given by:

$$n(\lambda) = B + \frac{C}{\lambda^2} + \frac{D}{\lambda^4} \quad (\lambda = 785 \text{ nm}) . \quad 5-3$$

Table 5-2 shows Cauchy's coefficients of the materials and shows the calculated refractive indices of Al_2O_3 , Si_3N_4 , and TiO_2 at $\lambda = 785 \text{ nm}$. Perfect dielectric materials with $n_{\text{cl}} = 1.45$ and $n_{\text{cl}} = 1.48$ were used to represent the SiO_2 cladding outside the measurement window and the toluene in contact with the waveguides during the experiments, respectively. A fine mesh ($dx = 5 \text{ nm}$, $dy = 2 \text{ nm}$) was used to improve the accuracy of the calculations in the core region. A rough mesh ($dx = 200 \text{ nm}$, $dy = 200 \text{ nm}$) was used for the calculations in the remaining region. We use metallic boundary conditions to reduce the calculation time. A distance of $12 \text{ }\mu\text{m}$ between the waveguide core and the boundary condition was set to eliminate artifacts. To detect non-physical modes, the gradient of each mode's intensity was calculated, and its phase was explored to detect the existence of maxima outside the core region. This calculation was done for the external simulation region separated by at least $1 \text{ }\mu\text{m}$ from the core. To reduce processing time, the calculations were parallelized in a 5x Intel® NUC5i7RYH (Intel Core i7-5557U, 256GB SSD, 16GB RAM) cluster running Ubuntu 16.04 LTS GNU/Linux and TORQUE PBS, Figure 5-4.

Figure 5-4 A clustered system to perform the simulation in parallel.

Figure 5-5 The mask design for waveguide fabrication. The mask has one test area and four different waveguide designs: WP_C6, WP_C7, WP_C8 and WP_C10.

Table 5-3 summarizes the dimensions of the waveguides for the three materials systems. The sizes were chosen considering the guiding properties for both the pump (785 nm) and the Raman signal (880 nm) wavelengths as well as a desire for a lithography-based production process. Two features were incorporated into the design to simplify and improve the coupling: inverted tapers from 1 μm to the width of the waveguide and a 90-degree S-turn with a large bending radius to minimize stray light from the Raman excitation/collection region. Spirals of about 14 mm long between the two end facets were designed to maximize the collected signal in the backscattering configuration.

Clewin5, in combination with MATLAB scripts, was used to generate the mask for the different designs, as shown in Figure 5-5. A chip size of 19x12 mm² was chosen. Three types of designs were produced: straight waveguides with a fixed length (~12 mm) and a varying width (0.8-1.6 μm), spirals of increasing length (14.2-163.9 mm) and fixed-width (1.0, 1.2 or 1.4 μm), and spirals of increasing width (0.8-1.6 μm) and fixed-length (98.2 mm). A width-sweep was used in all cases to compensate for possible variations in the refractive indices of the fabricated layers, which may affect their guiding properties. Waveguides with a 50 μm width (i.e., slab waveguides) are used as a marker to locate the groups of thinner waveguides and to perform a rough initial coupling

Table 5-2 Cauchy's coefficients for the calculation of the refractive index of each material

Coefficients	B	C	D	$n(\lambda) _{\lambda=785\text{nm}}$
Al_2O_3	1.6487	6.1e-3	1e-4	1.6589
Si_3N_4	1.9832	9.6e-3	6e-4	2.0004
TiO_2	2.2105	6.03e-2	-2e-3	2.3031

Table 5-3 Dimension of the three waveguides used for the experiments

Coefficients	Core height / nm	Core width / μm	Length / mm
Al_2O_3	346	1.2	14.208
Si_3N_4	127	1.2	14.208
TiO_2	181	1.2	14.208

5.4 Summary of Waveguide Fabrication

The fabrication starts with the deposition of the optical material of the core onto and oxidized (i.e., 8 μm thermal oxide) silicon wafer. Low-pressure chemical vapor deposition (LPCVD) was used to deposit the Si_3N_4 layer. Al_2O_3 and TiO_2 were deposited by reactive RF magnetron sputtering. A standard UV lithography (i-line, $\lambda = 365\text{ nm}$) process follows. First, the substrates were cleaned in a nitric acid (HNO_3) bath to eliminate organic residues for 10 minutes, rinsed with a quick dump rinse (QDR) and spin-dried. Then, a dehydration baking and deposition of an HMDS monolayer using a vacuum oven were performed. A photoresist (Olin OiR 907/17) was deposited by spin coating. A UV mask aligner (Electronic Vision Group, EVG620) was used to transfer the patterns into the resist layer. The development of the structures was done by direct immersion in an OPD4262 developer. Reactive ion etching (RIE) followed to transfer the patterns created by photolithography into the optical layer. After the etching of the optical layer, resist stripping was done in an oxygen plasma. An Argon/oxygen plasma was used to process the Si_3N_4 layers and a nitrogen/oxygen plasma was used for the TiO_2 and Al_2O_3 layers. A SiO_2 cladding (3 μm thick) was deposited by PECVD everywhere apart from the sensing window. Finally, the wafer was diced into chips. Before the dicing process, a resist layer was spin coated (Olin OiR 908-35) and baked to protect the optical components. A special optical-grade diamond blade was used during dicing. The chips were stored in a wafer box with dust-free tissue to protect chips from damage and scratch. Individual chips were extracted and cleaned only just before performing the experiments. Acetone was used to remove the protective resist layer. Then, sonication was done to eliminate particles from the dicing step. The chips were cleaned in nitric acid for 10 minutes to eliminate any organic residues, rinsed using QDR and spin dried. A summary of the fabrication flow can be seen in Figure 5-6 and the detailed process flow is described in Table 5-4. Figure 5-7 shows an example waveguide chip. The waveguide shown in Figure 5-7(A) is the first waveguide of WP_C10 used for the WGR experiments; the core of this waveguide is Al_2O_3 .

Figure 5-6 Common fabrication process flow to produce Al_2O_3 , Si_3N_4 and TiO_2 . Although each material uses different equipment and has different recipe, the same fabrication process is shared for all waveguide platforms.

Figure 5-7 (A) The 785 nm laser is coupled into an Al_2O_3 waveguide. (B) SEM image of a TiO_2 waveguide showing the end-facet of the waveguide.

Table 5-4 The fabrication recipe for the Si_3N_4 , Al_2O_3 and TiO_2 waveguides

Waveguide core materials		
Si_3N_4	Al_2O_3	TiO_2
Optical layer deposition		
LPCVD	Reactive sputtering	Reactive sputtering
$T = 750^\circ\text{C}$ $Q_{\text{SiH}_2\text{Cl}_2} = 25 \text{ sccm}$ $Q_{\text{NH}_3} = 250 \text{ sccm}$ $Q_{\text{N}_2} = 200 \text{ sccm}$ Power = ?W P = 150 mTorr	AJA ATC 1500 $T = 550^\circ\text{C}$ $Q_{\text{Ar}} = 30 \text{ sccm}$ $Q_{\text{O}_2} = 3.2 \text{ sccm}$ Power = 200W	MESA+ T'Coaty $T = 20^\circ\text{C}$ $Q_{\text{Ar}} = 34 \text{ sccm}$ $Q_{\text{O}_2} = 7.5 \text{ sccm}$ Power = 500 W P = 6e-3 mbar
UV lithography		
PR spin coating	Olin OiR 907/17	4000rpm, 30 s
UV mask aligner	EVG620	$t_{\text{exposure}} = 3.8 \text{ s}$
Post exposure bake	Hot plate	$T = 110^\circ\text{C}$, $t = 60 \text{ s}$
Develop	OPD4264	$T = 55^\circ\text{C}$
Hard bake	Hot plate	$T = 110^\circ\text{C}$, $t = 30 \text{ min}$
Optical layer etching (done by RIE)		
PlasmaTherm 790 $Q_{\text{CHF}_3} = 25 \text{ sccm}$ $Q_{\text{O}_2} = 5 \text{ sccm}$ power = 350Watt pressure = 20mTorr	Oxford PlasmaPro 100 Cobra $Q_{\text{BCl}_3} = 25 \text{ sccm}$ $Q_{\text{HBr}} = 10 \text{ sccm}$ ICP power = 1750 W RF power = 20 W pressure = 3 mTorr	Oxford PlasmaPro 100 Cobra $Q_{\text{SF}_6} = 25 \text{ sccm}$ $Q_{\text{Ar}} = 5 \text{ sccm}$ $Q_{\text{O}_2} = 6 \text{ sccm}$ ICP power = 1500 W RF power = 20 W pressure = 10 mTorr
PR stripping by Oxygen plasma		
Tepla 360 $Q_{\text{O}_2, \text{preheat}} = 0 \text{ sccm}$ $Q_{\text{Ar}, \text{preheat}} = 600 \text{ sccm}$ $t_{\text{preheat}} = 10 \text{ min}$ power _{preheat} = 1000W $Q_{\text{O}_2, \text{etching}} = 360 \text{ sccm}$ $Q_{\text{N}_2, \text{etching}} = 160 \text{ sccm}$ power _{etching} = 800W pressure = 0.6 mbar $t_{\text{etching}} = 10 \text{ mins}$		Tepla 360 $Q_{\text{O}_2, \text{preheat}} = 0 \text{ sccm}$ $Q_{\text{N}_2, \text{preheat}} = 500 \text{ sccm}$ $t_{\text{preheat}} = 10 \text{ min}$ power _{preheat} = 1000 W $Q_{\text{O}_2, \text{etching}} = 500 \text{ sccm}$ $Q_{\text{N}_2, \text{etching}} = 0 \text{ sccm}$ power _{etching} = 800 W pressure = 1 mbar $t_{\text{etching}} = 10 \text{ mins}$
Spin coat protection layer for dicing		
Olin OiR 908-35 = 4000 rpm, 30 s $t_{\text{prebake}} = 120 \text{ s}$ $T_{\text{prebake}} = 95^\circ\text{C}$ $t_{\text{postbake}} = 10 \text{ min}$ $T_{\text{postbake}} = 120^\circ\text{C}$		

5.5 Results and discussion

According to the simulation, the Si_3N_4 waveguide and the TiO_2 waveguide can guide higher-order modes (Si_3N_4 : 2 TE and 2 TM, TiO_2 : 4TE and 3TM). However, the coupling efficiency of the higher modes is a 10^{24} - to 10^{30} -fold smaller than the efficiency of the fundamental TE or TM modes. Thus, the effect of higher modes in the Si_3N_4 and TiO_2 waveguides is negligible.

The first waveguide of the chip design WP_C10 was selected for the WGR experiments. The fabricated thickness of Al_2O_3 , Si_3N_4 and TiO_2 waveguide are described in Table 5-3. Each measurement used 100 sec integration time and 36 mW of incident light power. This long accumulation time was used to obtain a high SNR. A vertical binning was performed to create a line spectrum of 1002 pixels. First, the air clad waveguides were measured to obtain the intrinsic background. Then the waveguide was cladded with toluene and measured in the same way. We performed experiments with p-polarized (parallel to the chip surface) light and s-polarized light (vertical to the chip surface) to excite the waveguide TE and TM modes, respectively.

The data collected with the EMCCD camera consisted of arrays of 1002x1004 pixels. Figure 5-10 shows an image captured by the EMCCD camera. There are two bright pixels at (648,232) and (946,630). These are defective pixels on the EMCDD. The pixels introduced sharp peaks at 1130 cm^{-1} and 1624 cm^{-1} in every measurement. The values of those pixels were removed from the data by replacing them with the mean value of the adjacent pixels. Apart from removing those artifacts, the data is presented as collected without any further signal processing. To compare the waveguide Raman spectra from three different types of waveguides, the Raman spectra were normalized and scaled. Firstly, the spectra were normalized by a coupling efficiency of the waveguides. Because it is very limited to measure the accurate coupling power, it was calculated by Lumerical mode solution. (see Table 5-5) The normalization was followed by amplitude scaling based on a saturation length of the waveguides. Though the length of all waveguides is equal (~ 1.4 cm), the effective waveguide length for WGR can be different due to the different waveguide propagation loss. For the backscattering WGR measurement, a ratio of the Stokes signal power and the pump power can be written as

$$\frac{P_{\text{wg}}}{P_{\text{pump}}} = \rho\sigma(\omega_p, \omega_s) \frac{\eta}{2} \gamma_{\text{in}} \gamma_{\text{out}} \left(\frac{1 - e^{-2\alpha L}}{2\alpha} \right), \quad 5-4$$

where ρ is the molecular density, σ is the Raman cross-section of the molecule, η is the Raman conversion efficiency, γ_{in} and γ_{out} are the in- and out-coupling losses, α is the propagation loss of the waveguide. Here, we assume that the propagation loss for the pump light α_p and the Stokes light α_s are same values ($\alpha_p = \alpha_s$) to simplify the theoretical model. This model shows that the power of the waveguide Raman will be saturated and the saturation length can be determined by the propagation loss of the waveguide.

The propagation losses of all waveguide were measured with the toluene cladding for TE and TM polarization since there is a polarization dependence on the propagation loss, Table 5-6. More details for the propagation loss measurement can be found in the digital supplementary.

The propagation loss was measured from the top of the waveguide. The 785 nm monochrome light was coupled into the waveguide and the image of the waveguide guiding the light was taken from the top using a camera (BFLY-U3-23S6M-C; Point Grey Research Inc., Richmond, BC, Canada). The intensity values along the waveguide were collected.

Since the intensity changes can be described as $I(z) = I_0 e^{-\alpha z}$, an exponential function fitting was performed to extract the propagation loss α .

Figure 5-8 An example of the waveguide propagation loss measurement: The propagation loss was measured from the top of the waveguide. The panel (A) shows the image of a Si₃N₄ waveguide guiding the 785 nm light, (B) shows the intensity along the waveguide and a fitted exponential function. Yellow dots in (A) represent data collection points.

Figure 5-9 shows the background spectrum collected from the Al_2O_3 , Si_3N_4 , and TiO_2 waveguides. An incident pump power of 36 mW was focused on the input of the waveguides. The Raman background was probed for both TE (Figure 5-9A) and TM polarization (Figure 5-9B). The Y-axis represents the averaged counts on the EMCCD camera after 100 sec integration time. In the case of the Al_2O_3 waveguide, the mode profiles of the TE and TM modes exhibit similar overlap with the core region (~ 0.656), as can be seen in Figure 5-11. A similar Raman background is, therefore, expected.

Table 5-5 The calculated coupling efficiencies. The calculation was done by Lumerical mode solution.

	TE mode	TM mode
Al_2O_3	0.0196	0.0229
Si_3N_4	0.0096	0.0191
TiO_2	0.0045	0.0071

Table 5-6 Propagation losses of waveguides were measured with the toluene cladding ($n=1.48$).

	TE mode	TM mode
Al_2O_3	5.1577 ± 4.2981	9.9704 ± 2.8367
Si_3N_4	1.6941 ± 2.7777	2.6242 ± 2.2372
TiO_2	2.5860 ± 3.6288	2.5161 ± 2.5961

Figure 5-9 Inherent background signals for the three waveguide platforms in two different fundamental mode; (A) TE and (B) TM. Dotted vertical lines represent the position of Raman peaks of toluene.

Figure 5-10 Image on the EMCCD, for a toluene measurement on a Si_3N_4 waveguide for the TM mode. The spectrum of toluene is mainly focused on the middle of CCD. As marked in the image, there are two bright spots: the size of each spot is one pixel. The column on the right shows a vertically integrated line spectrum from 1-400 (red), 411-580 (blue) and 581-1004 (green) pixels. There is a sharp peak on the red and green curves that correspond to white dots on the image: The peaks caused by the white dots are marked with black arrows.

Figure 5-11 Mode profiles for the three waveguide materials. The simulations were performed for a toluene cladding. For an air cladding the profiles shift down a little. The dimension of the waveguides are Al_2O_3 : $1.2 \mu\text{m} \times 346 \text{ nm}$, Si_3N_4 : $1.2 \mu\text{m} \times 127 \text{ nm}$ and TiO_2 : $1.2 \mu\text{m} \times 181 \text{ nm}$.

Furthermore, the increased coupling efficiency to the thicker Al_2O_3 waveguides can also explain the higher intensity of the background. When the TE mode of the Al_2O_3 waveguide was excited, we observed ripples in the signal. The spectral periodicity of $\sim 6 \text{ nm}$ indicates some spurious reflection over a distance of 30 mm . Since this is not related to the waveguide (346 nm) or another relevant physical effect, we can ignore these ripples.

Although the Si_3N_4 waveguide is known for a broad fluorescence background in the near-infrared region[34,35], the TM mode gave little fluorescence background and the baseline above 700 cm^{-1} is quite flat. While the TE mode excites a band between $360 - 870 \text{ cm}^{-1}$, the TM mode has less background below 500 cm^{-1} .

A significant reduction of the background signal is observed for the TM polarization in the TiO_2 waveguide, especially below 1000 cm^{-1} . The reduction originates from a combination of a lower overlap of the TM mode with the waveguide core as well as a lower coupling efficiency into the waveguide due to a smaller overlap with the incident input beam. All the waveguides show a sharp peak around 520 cm^{-1} . This location corresponds to a Raman peak of the Si-Si bond of the substrate.

After the background measurements, an analyte was probed by the waveguides for TE and TM polarization. For this experiment, 4 μl of toluene was added onto the sensing window. To avoid the collection of Raman signal from the toluene droplet directly, we placed the droplet $\sim 2 - 3$ mm from the end facet of the waveguide and on top of the spiral. In addition, the 90 deg bend of the input waveguide into the spiral further minimizes the direct excitation of Raman from the toluene by stray light. To further ensure that the collected Raman signal originates solely from the waveguide, we restricted the uncoupled stray light by a physical block placed in close proximity above the input waveguide. The camera was set to continuously record the Raman spectra using a low integration time (1 sec) to monitor spectral changes over time as the toluene was added to the sensing window. Several Raman peaks appeared immediately after adding toluene onto the chip. Again, an incident pump power of 36 mW was used in all experiments. Toluene on the other two waveguides (Al_2O_3 and TiO_2) was probed in the same way. After the toluene measurement, the sharp artifacts at (648, 232) and (964, 630) pixels were removed. No further data processing was carried out.

Figure S-12 Raman spectra of toluene obtained on three waveguide platforms. Panel (A) shows Raman spectra for TE and panel (B) shows spectra for TM. Transparent purple curve represents Raman spectrum of bulk toluene collected on the same experimental setup.

Table 5-7 SNR of Raman spectra of three waveguide materials. The SNR was calculated by the photon count at 1003 cm^{-1} and the averaged photon count at the bottom of the Raman peak

	TE mode	TM mode
Al_2O_3	75.85	119.79
Si_3N_4	52.12	134.41
TiO_2	21.99	39.27

The Raman spectra of toluene are shown in Figure 5-12. Clear peaks near 520 cm^{-1} , 786 cm^{-1} , 1003 cm^{-1} , 1030 cm^{-1} and 1211 cm^{-1} , which are distinct Raman signatures of toluene, are observed for the three waveguides for both polarizations. The peak at 520 cm^{-1} is hard to distinguish as it overlaps with the silicon peak. As expected, the Raman signatures measured for TM excitation exhibit a higher SNR than their TE counterparts. This is due to the larger overlap of the excitation field with the toluene as well as a decreased background. Table 5-7 shows the SNR of the 1003 cm^{-1} peaks for the three material platforms for both TE and TM polarization. The SNR ratio was calculated based on the photon count of the peak and the averaged photon count of the bottom of the selected peak[26]. The SNR can be written as

$$SNR = \frac{C_{sig} - C_{bg}}{\sqrt{C_{sig}}}, \quad 5-5$$

where C_{sig} is the photon count of the Raman peak (1003 cm^{-1}) and C_{bg} is the averaged photon count of a baseline of the selected Raman peak. The calculation also shows that a higher SNR Raman signal was collected for TM polarization than TE polarization. Contrary to what it was expected, the SNR obtained for the TiO_2 waveguide in the TM configuration was lower than for the other two materials, despite the higher refractive index contrast (i.e., higher coupling efficiency of Raman signal back into the waveguide) and similar (or even lower) background than the other two material platforms. The lower coupling efficiency of the pump into the waveguide in combination with higher propagation losses can explain the lower performance of the TiO_2 waveguides. Since TiO_2 has the highest refractive index among the three materials, the thickness of the TiO_2 waveguide can be thinner than the other two waveguides to support single TE mode. An

86 nm thick TiO_2 waveguide showed lower background than the 181 nm thick TiO_2 waveguide used for this study. The thinner TiO_2 waveguide, however, barely supports the TM mode. This causes the Raman collection for the TM to show a lower SNR than the Raman signal for the TE mode (See Figure 5-13). Therefore, the thickness of the TiO_2 waveguide had to be thicker to support the TM mode properly. Improving the coupling efficiency of the pump by developing mode size converters and reducing the propagation losses of the TiO_2 in the near-infrared (i.e., 785-1000 nm range) will improve this platform. The deposition of a fully amorphous TiO_2 without any anatase nanocrystals will help in further reducing the Raman background from the TiO_2 waveguides.

Figure 5-13 Raman measurements of the background and analyte, toluene, on TiO_2 waveguides. (A) and (C) Raman measurements for the TE mode. (B) and (D) TM mode. Blue curves represent the signal collected from a 181 nm thick TiO_2 waveguide and red curves an 86 nm thick TiO_2 waveguide. According to simulation, the thin waveguide is limited at supporting the TM mode. Although the TM mode generates less background than the TE mode in the 86 nm waveguide, the Raman spectrum of toluene for the TM has a lower count than the signal obtained for the TE mode. The thicker waveguide shows a higher background level, which limits the collection of Raman signal.

5.6 Conclusion and Future Work

In summary, we designed and fabricated waveguide chips using three different materials, Al_2O_3 , Si_3N_4 , and TiO_2 , to compare their performance for WGR. Chips containing 14 mm sensing spirals were fabricated using standard microfabrication methods. The waveguide cross-sections were optimized to operate in multimode conditions for TE and TM polarization. Air and toluene cladded spiral waveguides were measured to collect the inherent background and Raman signal from an analyte. The Raman background from the core of the waveguides was reduced using TM polarization. This effect is ascribed to the smaller overlap of the TM mode with the waveguide core material. Although the higher refractive index contrast of the TiO_2 waveguides increases the coupling efficiency of the Raman signal, the increased losses and lower coupling efficiency of the pump light into the waveguide led to a smaller SNR for the TiO_2 platform. For the Al_2O_3 and Si_3N_4 waveguides, TM polarization led to a lower background and Raman spectra with a higher SNR. In this comparison, we conclude that the combination of a Si_3N_4 waveguide and a TM polarized pump beam is best for the WGR. This combination showed the highest SNR Raman spectra of the analyte as well as less background generation and a reasonably flat baseline.

In this chapter, we successfully demonstrated that the evanescent field of the waveguide can probe an analyte, specifically toluene, on the waveguide. We anticipate that the use of WGR can be extended to other applications, such as a drinking water sensor and a biosensor. To make the WGR chip a bit more useful for sensor applications, an additional device for sample handling has to be taken into account. A polydimethylsiloxane (PDMS) microfluidic device can be a promising candidate. PDMS is known as a nontoxic biocompatible material that can remain bio-sample chemically intact. PDMS microchannel devices enable low consumption of samples and agents, preventing the sample from drying out and changing the sample during the measurement.

The use of PDMS microfluidics was reflected in the WGR chip design for further application and PDMS devices were designed and fabricated, as shown in Figure 5-14. The fabrication process of the PDMS microfluidic device will be described in Appendix C. The fabricated PDMS chip, unfortunately, could not be used for the toluene measurement since PDMS is known to swell in the presence of toluene[37] rapidly. However, PDMS can be used for bio-samples handling which is normally suspended in an aqueous solution such as phosphate-buffered saline (PBS) and/or cell culture medium.

To be used as a biosensor, further development of the surface chemistry has to be carried out. The target molecules or particles have to be brought to the waveguide surface. The

functionalization of the waveguide surface will enable the waveguide to capture specific biomolecules.

The combination of the functionalized waveguide and the PMDS microchannel device could be used to measure EVs. In the combined device, EVs would be enriched along the waveguide surface and be probed by the evanescent field generating high SNR Raman signal. Unfortunately, there was not enough time within this project to do these experiments

Figure 5-14 A mask design for PDMS chip fabrication (top) and the fabricated PDMS microchannel device (bottom). The channel is filled with blue colored water to show its structure.

Bibliography

- [1] D. Ahuja, D. Parande, "Optical sensors and their applications", *Journal of Scientific Research and Reviews*, **2012**; *1*, 060.
- [2] N. Sabri, S. Aljunid, M. Salim, S. Fouad, In *Recent Trends in Physics of Material Science and Technology*; Springer: **2015**, p 299.
- [3] V. K. Rai, "Temperature sensors and optical sensors", *Applied Physics B*, **2007**; *88*, 297.
- [4] M. De Goede, M. Dijkstra, R. Obregón, J. Ramón-Azcón, E. Martínez, L. Padilla, F. Mitjans, S. Garcia-Blanco, "Al₂O₃ microring resonators for the detection of a cancer biomarker in undiluted urine", *Optics Express*, **2019**; *27*, 18508.
- [5] R. Horváth, H. C. Pedersen, N. Skivesen, D. Selmeçzi, N. B. Larsen, "Optical waveguide sensor for on-line monitoring of bacteria", *Optics Letters*, **2003**; *28*, 1233.
- [6] B. MacCraith, V. Ruddy, C. Potter, B. O'Kelly, J. McGilp, "Optical waveguide sensor using evanescent wave excitation of fluorescent dye in sol-gel glass", *Electronics Letters*, **1991**; *27*, 1247.
- [7] A. Ksendzov, Y. Lin, "Integrated optics ring-resonator sensors for protein detection", *Optics Letters*, **2005**; *30*, 3344.
- [8] C. A. Barrios, K. B. Gylfason, B. Sánchez, A. Griol, H. Sohlström, M. Holgado, R. Casquel, "Slot-waveguide biochemical sensor", *Optics Letters*, **2007**; *32*, 3080.
- [9] P. V. Lambeck, "Integrated optical sensors for the chemical domain", *Measurement Science and Technology*, **2006**; *17*, R93.
- [10] K. Zinoviev, C. Dominguez, J. A. Plaza, V. J. C. Busto, L. M. Lechuga, "A novel optical waveguide microcantilever sensor for the detection of nanomechanical forces", *Journal of Lightwave Technology*, **2006**; *24*, 2132.
- [11] R. De La Rica, E. Mendoza, L. M. Lechuga, H. Matsui, "Label-free pathogen detection with sensor chips assembled from peptide nanotubes", *Angewandte Chemie International Edition*, **2008**; *47*, 9752.
- [12] E. Li-Chan, "The applications of Raman spectroscopy in food science", *Trends in Food Science & Technology*, **1996**; *11*, 361.

- [13] T. Vankeirsbilck, A. Vercauteren, W. Baeyens, G. Van der Weken, F. Verpoort, G. Vergote, J. P. Remon, "Applications of Raman spectroscopy in pharmaceutical analysis", *TRAC Trends in Analytical Chemistry*, **2002**; *21*, 869.
- [14] D. Bersani, J. M. Madariaga, "Applications of Raman spectroscopy in art and archaeology", *Journal of Raman Spectroscopy*, **2012**; *43*, 1523.
- [15] L. A. Austin, S. Osseiran, C. L. Evans, "Raman technologies in cancer diagnostics", *The Analyst*, **2016**; *141*, 476.
- [16] J. Desroches, M. Jermyn, K. Mok, C. Lemieux-Leduc, J. Mercier, K. St-Arnaud, K. Urmeý, M.-C. Guiot, E. Marple, K. Petrecca, "Characterization of a Raman spectroscopy probe system for intraoperative brain tissue classification", *Biomedical optics express*, **2015**; *6*, 2380.
- [17] W. Lee, A. Nanou, L. Rikkert, F. A. W. Coumans, C. Otto, L. W. M. M. Terstappen, H. L. Offerhaus, "Label-Free Prostate Cancer Detection by Characterization of Extracellular Vesicles Using Raman Spectroscopy", *Analytical Chemistry*, **2018**; *90*, 11290.
- [18] Z. J. Smith, C. Lee, T. Rojalín, R. P. Carney, S. Hazari, A. Knudson, K. Lam, H. Saari, E. L. Ibañez, T. Viitala, "Single exosome study reveals subpopulations distributed among cell lines with variability related to membrane content", *Journal of extracellular vesicles*, **2015**; *4*, 28533.
- [19] R. P. Carney, S. Hazari, M. Colquhoun, D. Tran, B. Hwang, M. S. Mulligan, J. D. Bryers, E. Girda, G. S. Leiserowitz, Z. J. Smith, K. S. Lam, "Multispectral Optical Tweezers for Biochemical Fingerprinting of CD9-Positive Exosome Subpopulations", *Analytical Chemistry*, **2017**; *89*, 5357.
- [20] K. Ajito, K. Torimitsu, "Single nanoparticle trapping using a Raman tweezers microscope", *Applied Spectroscopy*, **2002**; *56*, 541.
- [21] P. Fernandes, P. Salomé, A. Da Cunha, "Growth and Raman scattering characterization of Cu₂ZnSnS₄ thin films", *Thin Solid Films*, **2009**; *517*, 2519.
- [22] J. Park, M. Hwang, B. Choi, H. Jeong, J.-h. Jung, H. K. Kim, S. Hong, J.-h. Park, Y. Choi, "Exosome classification by pattern analysis of surface-enhanced Raman spectroscopy data for lung cancer diagnosis", *Analytical Chemistry*, **2017**; *89*, 6695.
- [23] C. E. Talley, J. B. Jackson, C. Oubre, N. K. Grady, C. W. Hollars, S. M. Lane, T. R. Huser, P. Nordlander, N. J. Halas, "Surface-enhanced Raman scattering from individual Au nanoparticles and nanoparticle dimer substrates", *Nano Letters*, **2005**; *5*, 1569.

- [24] A. Dhakal, A. Z. Subramanian, P. Wuytens, F. Peyskens, N. Le Thomas, R. Baets, "Evanescent excitation and collection of spontaneous Raman spectra using silicon nitride nanophotonic waveguides", *Optics Letters*, **2014**; *39*, 4025.
- [25] N. Schlotter, J. Rabolt, "Raman spectroscopy in polymeric thin film optical waveguides. 1. Polarized measurements and orientational effects in two-dimensional films", *The Journal of Physical Chemistry*, **1984**; *88*, 2062.
- [26] A. Dhakal, P. C. Wuytens, F. Peyskens, K. Jans, N. L. Thomas, R. Baets, "Nanophotonic Waveguide Enhanced Raman Spectroscopy of Biological Submonolayers", *ACS Photonics*, **2016**; *3*, 2141.
- [27] C. Duverger, J.-M. Nedelec, M. Benatsou, M. Bouazaoui, B. Capoen, M. Ferrari, S. Turrell, "Waveguide Raman spectroscopy: a non-destructive tool for the characterization of amorphous thin films", *Journal of Molecular Structure*, **1999**; *480*, 169.
- [28] A. Dhakal, P. Wuytens, F. Peyskens, A. Skirtach, N. Le Thomas, R. Baets In *2016 IEEE Photonics Conference (IPC)*; IEEE: **2016**, p 144.
- [29] S. A. Holmstrom, T. H. Stievater, D. A. Kozak, M. W. Pruessner, N. Tyndall, W. S. Rabinovich, R. A. McGill, J. B. Khurgin, "Trace gas Raman spectroscopy using functionalized waveguides", *Optica*, **2016**; *3*, 891.
- [30] N. F. Tyndall, T. H. Stievater, D. A. Kozak, K. Koo, R. A. McGill, M. W. Pruessner, W. S. Rabinovich, S. A. Holmstrom, "Waveguide-enhanced Raman spectroscopy of trace chemical warfare agent simulants", *Optics Letters*, **2018**; *43*, 4803.
- [31] P. Løvhaugen, B. S. Ahluwalia, T. R. Huser, O. G. Hellesø, "Serial Raman spectroscopy of particles trapped on a waveguide", *Optics Express*, **2013**; *21*, 2964.
- [32] A. Dhakal, P. Wuytens, A. Raza, N. Le Thomas, R. Baets, "Silicon Nitride Background in Nanophotonic Waveguide Enhanced Raman Spectroscopy", *Materials (Basel, Switzerland)*, **2017**; *10*, 140.
- [33] A. Subramanian, P. Neutens, A. Dhakal, R. Jansen, T. Claes, X. Rottenberg, F. Peyskens, S. Selvaraja, P. Helin, B. Du Bois, "Low-loss singlemode PECVD silicon nitride photonic wire waveguides for 532–900 nm wavelength window fabricated within a CMOS pilot line", *IEEE Photonics Journal*, **2013**; *5*, 2202809.
- [34] H. Hao, L. Wu, W. Shen, H. Dekkers, "Origin of visible luminescence in hydrogenated amorphous silicon nitride", *Applied Physics Letters*, **2007**; *91*, 201922.

- [35] C. m. Mo, L. Zhang, C. Xie, T. Wang, "Luminescence of nanometer-sized amorphous silicon nitride solids", *Journal of Applied Physics*, **1993**; 73, 5185.
- [36] C. C. Evans, C. Liu, J. Suntivich, "TiO₂ nanophotonic sensors for efficient integrated evanescent Raman spectroscopy", *Acs Photonics*, **2016**; 3, 1662.
- [37] D. Liu, H. J. Byrne, L. O'Neill, B. Hennelly In *Optical Sensing and Detection V*; International Society for Optics and Photonics: **2018**; Vol. 10680, p 1068027.

Micro Ring Resonator Enhanced Waveguide Raman Spectroscopy

Waveguide Raman spectroscopy has been demonstrated in the previous chapter. Though WGR shows promising results, its performance can be limited by an intrinsic Raman and/or fluorescent background of the waveguide (core) material. This background signal of the material can hinder the acquisition of a high SNR WGR signal.

An optical Micro Ring Resonator (MRR) can be used to improve the Raman signal. When the MRR is on-resonance, the intensity of propagating light will be built up in the ring, which leads to local field enhancement around the ring. This local field enhancement allows using lower pump power, which reduces the background generation in the bus waveguide. The enhanced field will locally generate the background of the waveguide materials where the Raman signal is also generated.

This chapter will cover some theoretical backgrounds on the MMR. After the discussion on the operating concept of the MMR, the design and fabrication of MRRs and the MRR enhanced waveguide Raman spectroscopy will be discussed in depth.

A manuscript has been preparing based on this chapter.

Ring Resonator enhanced Waveguide Raman Spectroscopy

Wooje Lee, Michiel de Goed, Harmen Smedes, Meindert Dijkstra, Sonia M. García-Blanco and Herman L. Offerhaus

6.1 Introduction

Waveguide Raman spectroscopy (WGR) can probe very thin film layers or nanoparticles on top of a waveguide with a higher SNR than a conventional bulk Raman microscope using free-space optics[1-4]. However, WGR suffers from reduced sensitivity due to the intrinsic background of waveguide materials. When light is coupled into the waveguide, it excites the Raman and/or the fluorescent background of the material[5].

There were several attempts to minimize the native background signal of waveguides and to improve the strength of the Raman signal. Background suppression can sometimes be done by a fabrication process optimization. For example, Si_3N_4 is known to have a fluorescent background in the near-infrared range[6,7]. This material property depends on the fabrication process. Process optimization allows the production of a Si_3N_4 waveguide with a low (or no) fluorescent background[8]. These approaches can improve the signal to noise ratio (SNR) of the WGR system, but it cannot enhance the Raman signal.

In this chapter, micro Ring resonator Enhanced WaveGuide Raman spectroscopy (RE-WGR) will be introduced. The optical Micro Ring Resonator (MRR) is a building block of integrated photonics[9] with many different applications such as lasers[10,11], amplifiers[12,13], logic gates[14,15] and filters[16,17]. Since the resonance condition can be changed by some parameters (e.g., temperature or refractive index), the MMR can be used as sensors, and sensor applications have also been demonstrated[18-21].

When the resonator is on-resonance, the light coupled into the ring is combined constructively[9,22]. Therefore, we assume that the use of the MRR will allow probing an analyte on the MRR with lower external laser power. The low laser power generates less background signal power but also generates less Raman signal. The MRR will enhance the Raman signal power locally with only a local generation of background.

6.2 Ring Resonator

There are mainly two types of ring resonators; one is an all-pass ring resonator and the other is an add-drop ring resonator. The all-pass ring resonator consists of one ring and one bus waveguide. The add-drop has one ring and two bus waveguides, see Figure 6-1. In this thesis, the all-pass ring resonator (Figure 6-1A) will be discussed only[9,22,23].

Figure 6-1 Schematic diagram of two types of the micro ring resonators: (A) all-pass ring resonator and (B) add-drop ring resonator.

The all-pass ring resonator can be considered as a four-port system (Figure 6-1A). The relationship between the ports can be described as

$$\begin{pmatrix} E_4 \\ E_2 \end{pmatrix} = \begin{pmatrix} r & it \\ it & r \end{pmatrix} \begin{pmatrix} E_3 \\ E_1 \end{pmatrix}, \quad 6-1$$

where r is a self-coupling coefficient, and t is a cross-coupling coefficient. r^2 and t^2 represent the power splitting ratio of the coupler. If the coupling losses are small enough, they satisfy $r^2 + t^2 = 1$.

Since the field at port 3 is only affected by propagation in the ring, it can simply be expressed as

$$E_3 = ae^{i\phi} E_4, \quad 6-2$$

where a is the single-pass amplitude loss coefficient and ϕ is the single-pass phase shift. The phase shift ϕ which is related to the angular frequency is given by

$$\phi = \omega \tau_{ring} , \quad 6-3$$

where ω is the angular frequency ($\omega = 2\pi c / \lambda$) and τ_{ring} is the transit time of the ring resonator. The transit time is determined by the circumference of the ring, the speed of light and the effective refractive index of the waveguide material. Thus, τ_{ring} can be written as

$$\tau_{ring} = \frac{n_{eff} L}{c} . \quad 6-4$$

6

Here, n_{eff} represents the effective refractive index of the waveguide. Combining Equation 6-4 and the angular frequency, Equation 6-3 can be rewritten as follows

$$\phi = \frac{2\pi n_{eff} L}{\lambda} , \quad 6-5$$

where L ($L = 2\pi R$) is the circumference of the ring. In Equation 6-2, the single-pass amplitude loss coefficient a can be expressed as

$$a^2 = e^{-\alpha L} , \quad 6-6$$

where α is the power attenuation coefficient which is the sum of the waveguide losses (e.g., propagation, bending and coupling losses) in a unit of /m. This value can be converted to dB/m by multiplying a conversion factor $10/\ln(10)$.

The ratio of the transmitted field to the incoming field can be obtained by solving Equation 5-3 and 6-2. The ratio is given by

$$\frac{E_{t1}}{E_{i1}} = \frac{r - ae^{i\phi}}{1 - rae^{i\phi}} = \frac{r - a(\cos\phi + i\sin\phi)}{1 - ra(\cos\phi + i\sin\phi)} . \quad 6-7$$

In Equation 6-7, E_1 and E_2 are replaced by E_{i1} and E_{t1} , respectively. A ratio of the transmitted intensity to the incoming intensity (transmission factor, $\mathcal{T}$) is given by the squared modulus of the transmitted field. Thus, the ratio can be derived as follows

$$\mathcal{T} = \frac{I_{t1}}{I_{i1}} = \left| \frac{E_{t1}}{E_{i1}} \right|^2 = \frac{a^2 - 2ra \cos \phi + r^2}{1 - 2ra \cos \phi + (ra)^2}. \quad 6-8$$

A ratio of the circulating field E_3 (E_3 is replaced by E_{i2}) to the incident field E_{i1} can also be driven by solving Equation 5-3 and 6-2. The ratio can be expressed as follows

$$\frac{E_{i2}}{E_{i1}} = \frac{ita e^{i\phi}}{1 - ra e^{i\phi}} = \frac{ita(\cos \phi + i \sin \phi)}{1 - ra(\cos \phi + i \sin \phi)}. \quad 6-9$$

The buildup factor $\mathcal{B}$ which is defined as the ratio of the circulating intensity to the incident intensity can be derived from Equation 6-9:

$$\mathcal{B} = \frac{I_{i2}}{I_{i1}} = \left| \frac{E_{i2}}{E_{i1}} \right|^2 = \frac{(1 - r^2)a^2}{1 - 2ra \cos \phi + (ra)^2}. \quad 6-10$$

Both Equation 6-8 and 6-10 show that when the phase shift becomes $\phi = 2\pi m$ (m should be an integer), the incident light is resonant with the ring, which is called on resonance. At the on-resonance condition, the intensity at the port 2 (transmission of the bus waveguide) has a minimum value and the intensity is built up in the ring. Since the system goes on resonance when $\phi = 2\pi m$, the resonance wavelength of the ring can be expressed as

$$\lambda_{res} = \frac{n_{eff} L}{m}. \quad 6-11$$

Figure 6-2 shows an example of the transmitted intensity spectrum of the MRR. The figure shows that the intensity is dropped rapidly when $\phi = 2\pi m$. Some important parameters of the MRR are marked on the figure, which will be discussed in the next section.

Figure 6-2 Transmission spectrum of the all-pass ring resonator. When phase shift becomes $2\pi m$, the transmitted intensity is dropped sharp. The figure also shows some important spectral characteristics of a MRR.

6

6.2.1 Spectral Characteristics of a Micro-Ring Resonator

Figure 6-2 shows an example of a transmitted intensity spectrum. In the figure, some important spectral characteristics are indicated. There are two most important parameters to determine the ability of the ring resonator; a quality factor Q and a finesse $\mathcal{F}$.

The FWHM is a parameter showing the width of a peak at a certain resonant wavelength. The FWHM is given by

$$FWHM = \frac{\lambda_{res}^2}{\pi n_g L} \left(\frac{1-ra}{\sqrt{ra}} \right), \quad 6-12$$

where n_g represents the group refractive index. The group index can be defined as a ratio of the velocity of the light in the vacuum to the group velocity in a medium. The group index n_g is given by

$$n_g = n_{eff} - \lambda_0 \frac{dn_{eff}}{d\lambda_0}, \quad 6-13$$

where λ_0 is the wavelength of the light in a vacuum.

Another parameter is the FSR. The FSR represents the spectral spacing between adjacent resonant peaks. The FSR is expressed as

$$FSR = \frac{\lambda_{res}^2}{n_g L} . \quad 6-14$$

Based on the FSR and FWHM, the finesse $\mathcal{F}$ and the quality factor Q can be defined. The finesse $\mathcal{F}$ and Q factor are given by

$$\mathcal{F} = \frac{FSR}{FWHM} . \quad 6-15$$

$$Q = \frac{\lambda_{res}}{FWHM} . \quad 6-16$$

The finesse characterizes the spectral shape of the resonant peaks. The value represents the sharpness of a resonant peak as a ratio of the resonant wavelength to the width of the peak. The Q factor represents the sharpness of a resonant peak as a ratio of the distance between adjacent resonant peaks to the width of the peak.

Figure 6-3 Waveguide mode profiles of the $1.4\ \mu\text{m} \times 350\ \text{nm}$ (width $\times$ height) Al_2O_3 waveguide. (A) shows the cross section of the waveguide structure used for the simulation. (B) is the quasi-TE mode and (C) is the quasi-TM mode of the Al_2O_3 .

6.3 Design consideration

Several simulation runs were performed to predict the transmission spectrum of the MRRs. Two types of simulations were carried out using Lumerical MODE solutions (v7.11.1541, Lumerical Inc., Vancouver, BC, Canada) and MATLAB R2016b (v9.1.0, MathWorks Inc., Natick, MA, USA). Lumerical MODE solutions was used to calculate eigenmode, effective refractive index, and group refractive index of the waveguides. MATLAB was used for the theoretical prediction of the transmission spectrum of the MRRs.

Figure 6-3A shows the cross-section of the Al_2O_3 waveguide used for the waveguide mode simulation. Since an analyte, toluene, will be used to demonstrate the MRR enhanced waveguide Raman spectroscopy (RE-WGR), the simulation was performed with toluene upper cladding. As usual, the TE mode is confined more in the Al_2O_3 core than the TM

mode. The effective refractive index (n_{eff}) and the group refractive index (n_g) at the pump wavelength of 785 nm were also calculated by Lumerical mode solutions, Table 6-1.

Table 6-1 Calculated effective index and group index of the Al_2O_3 waveguide

Polarization	Effective index (n_{eff})	Group index (n_g)
quasi-TE	1.54097	1.64198
quasi-TM	1.53002	1.63280

Based on n_{eff} and n_g , the transmission spectrum of the ring resonator was predicted using custom software implemented in MATLAB. The resonance wavelengths of the MRRs for various radii were calculated using Equation 6-11. Figure 6-4 shows the resonance wavelengths for each radius. Since the effective refractive index of the waveguide for TE and TM polarization are not the same value, the resonance wavelengths were calculated for TE and TM polarization separately. As can be seen in Figure 6-4, the larger the ring radius, the more resonance wavelengths appear. The FSR is also reduced by increasing the radius of the ring.

A simulation to predict the transmission spectrum was performed for a 200 μm ring. To simplify the calculation, the incoming field was chosen to be 1. The propagation losses of 4.84 dB/cm and 6.52 dB/cm for TE and TM polarization were used for this calculation. The propagation loss measurement will be shown in Section 6.6. The simulation was performed based on the assumption that the cross-coupling value is 10% ($t^2 = 0.1$).

Since the measured center wavelength of the laser is 783.03 nm, FWHM and Q-factor of the 200 μm Al_2O_3 MRR were calculated at the closest resonance wavelength to the center wavelength of the laser which is 783.032 nm for TE_{00} and 782.850 nm for TM_{00} polarization, respectively (Table 6-2). The simulation also shows that the MRR for the TE mode has a higher Q-factor (~ 67000) than for the TM mode (~ 56000). The intensity buildup in the ring for the TE mode is higher than the TM mode. The intensity in the MRR for the TE mode can be ~ 6 times higher than the intensity in the bus waveguide. For the TM mode, the intensity in the ring can be ~ 4 times higher than the intensity in the bus waveguide.

Figure 6-4 Resonance wavelengths of MRRs for different radii; (A) for TE polarization and (B) for TM polarization. The inset shows the wavelength range of interest (782.5-785.5 nm). The thick black in the figure represents the wavelength of the laser.

Figure 6-5 The transmission spectrum and intensity buildup spectrum of the 200 μm MRR, (A) for TE and (B) for TM polarization.

Table 6-2 Calculated FWHM, FSR, and Q-factor of the 200 μm MRR for TE and TM mode.

Mode	FWHM (μm)	FSR (nm)	Q-factor
TE ₀₀	11.6	0.298	4.87×10^4
TM ₀₀	14.0	0.299	5.59×10^4

6.4 Mask Design of Micro Ring Resonators

The Al_2O_3 microring resonators with different dimensions (e.g., a separation gap between a bus waveguide and a ring, a waveguide width and a radius of a ring) were designed to cover the tolerance of the fabrication process. The ring resonator was designed to guide light having a wavelength of 785 nm. Figure 6-6 shows a mask design for the waveguides and ring resonator. The mask was designed using CleWin 5.0 Layout Editor (v5.0.12, WieWeb software, Hengelo, the Netherlands) with MATLAB scripts.

The chip A2 (Figure 6-6B) was mainly used for this experiment. The width of all the bus waveguides in this chip is 1.4 μm . There are five different groups of MRRs on A2. The MRRs can be grouped by the radius; 50, 75, 100, 150, 200 and 300 μm (at the center of the waveguide). In each group, there are five different separation gaps between the bus waveguide and the ring. The gaps are 600, 700, 800, 900 and 1000 nm. Every MRRs has a two digits number that identifies the MRR. The tens digit shows the radius and the ones-digits provide information about the separation gap of the MRR. For example, ring 51 means MRR in group no.5 (radius of 200 μm) having a separation gap of 600 nm.

In this mask, there is a chip named C0 Raman (Figure 6-6C). This chip only has straight waveguides and spirals. This chip has been included in this design for the waveguide Raman experiments (WGR) and studying the RE-WGR by comparing the Raman signal of spirals to the signal from the MRRs.

Figure 6-6 The mask design for MRR fabrication. The entire mask is shown in (A). (B) shows the MRR chip and (C) shows the spiral waveguides. The location of chip A2 and C0 is marked with red boxes in panel A.

6.5 Fabrication of Micro Ring Resonator

All the MRRs were fabricated in the MESA+ Nanolab at the University of Twente. A silicon wafer with an 8 μm thick thermally grown SiO_2 was used as a substrate. Aluminum oxide (Al_2O_3) was used as the core material for the MRR. A 350 nm thick Al_2O_3 layer was deposited on the SiO_2 layer by a reactive sputtering process using the AJA ATC 1500. The process chamber was pumped down to 3.7 mTorr. The sample chuck was heated to a set temperature of 550 $^\circ\text{C}$ by two 1 kW halogen lamps. The sputtering machine has three sputtering guns with three different 2-inch targets: Aluminum (Al, 99.9995%), Erbium (Er, 99.95%), and Ytterbium (Yb, 99.9%). The sputtering guns are controlled by either DC power or RF power source. The reactive sputtering was performed in Argon (Ar, 25 sccm) and Oxygen (O_2 , 2.5 sccm) ambient environment. After temperature stabilization, a 200 W RF power was applied to the Al target to start the deposition process.

The sputtering deposition process was followed by photolithography to transfer the waveguide and MRR patterns on the Al_2O_3 layer. The transferred layer serves as a mask

layer for the etching process. For photolithography, a positive photoresist (OiR907-17) was spin-coated on the Al_2O_3 layer. The thickness of the photoresist layer was 1.2 μm after the spin-coating. The MRR and waveguide patterns were transferred to the photoresist layer using UV contact aligner (Electronic Vision Group, EVG 620) which uses 'i-line' (365 nm) for the exposure. After the UV exposure, the wafer was immersed in a solution (OPD4262) to develop the structures.

With the photoresist mask, a reactive ion etching (RIE) was performed to transfer the MRR and waveguide patterns to the Al_2O_3 layer. The RIE process was performed on the Oxford PlasmaPro 100 Cobra using BCl_3 and HBr gas mixture at a ratio of 5 BCl_3 to 2 HBr where BCl_3 was 25 sccm and HBr was 10 sccm. An RF power of 25 W, ICP power of 1750 W, the total gas flow of 35 sccm and pressure of 12 mTorr were used. We obtained the etching rate of ~ 50 nm/min under these process parameters.

The process flow of the Al_2O_3 MRR fabrication is briefly summarized in Figure 6-7, and Figure 6-8 shows SEM images of an end facet of a waveguide and the gap between a bus waveguide and a ring structure.

Figure 6-7 Fabrication process flow for the Al_2O_3 waveguides and micro ring resonators

Figure 6-8 SEM images of a fabricated Al_2O_3 MRR. Panel A shows an end facet of the chip and panel B shows the gap between a bus waveguide and a ring structure.

6.6 Result and discussion

The propagation losses of the Al_2O_3 waveguides for both TE and TM polarization were measured for a more accurate theoretical prediction. To measure the propagation losses, a 785 nm beam was coupled into the waveguide and images of the light-guiding waveguide were taken from the top of the waveguide using a camera (BFLY-U3-23S6M-C; Point Grey Research Inc., Richmond, BC, Canada). The intensity values along the waveguide were collected and an exponential function ($I(z) = I_0 e^{-\alpha z}$) was fitted to determine the propagation loss α .

$$\alpha_{/cm} = \alpha_{dB/cm} \frac{\ln 10}{10} \quad 6-17$$

The propagation losses of the waveguides are shown in the legend of Figure 6-9, where the value of α in the figure is the propagation loss in the unit of /cm. Since the unit of ‘/cm’ and the unit of ‘dB/cm’ have a relationship described in Equation 6-17, to convert the extracted value to the unit of dB/cm, a conversion factor $10/\ln 10$ has to be multiplied to the exponent value of the base of the Euler’s number. The measured propagation losses shown in Table 6-3 were used for better prediction of the performance parameters of the MRRs such as FWHM, FSR, and Q-factor.

Figure 6-9 The propagation losses of the waveguide for (A) TE and (B) TM polarization. The exponent of the base e is the loss coefficient α .

Table 6-3 The extracted propagation losses for TE and TM polarization.

Mode	loss, α (dB/cm)
TE ₀₀	4.84±1.27
TM ₀₀	6.52±2.02

The setup described in Chapter 5 was used for the RE-WGR experiments. The laser source of the setup was replaced by a semiconductor laser (OBIS 785; Coherent Inc., Santa Clara, CA, USA) to achieve a narrower laser bandwidth. As described in the earlier chapter (see Figure 5-1A), the Al₂O₃ waveguides chip was mounted on a sample stage and a monochrome light (~785 nm) was coupled into the bus waveguide or spiral using a 40×/0.60 microscope objective (Nachet, France).

Figure 6-10 Incident monochrome light coupled into the ring (A) 31, (B) 41, (C) 51 and (D) 61. The separation gap of the MRRs are 600nm. The radii of the rings are (A) 100, (B) 150, (C) 200 and (D) 300 μm .

Due to the fabrication tolerance, there were discrepancies in the shape of the waveguide structures between the design target and the realized MRRs. Therefore, all the MRRs on the chip A2 were measured to find a stable MRR for the RE-WGR experiments. The rings lighting up in Figure 6-10 indicate that the MRRs are on-resonance. As a result of the experiments, it has been shown that the larger MRRs with the smallest separation gap performed best in this chip.

Toluene was probed on the MRR 51 (1.4 $\mu\text{m} \times 350 \text{ nm}$, the radius of 200 μm and 600 nm gap between the bus waveguide and the ring). Toluene was also measured on a spiral waveguide (1.4 $\mu\text{m} \times 350 \text{ nm}$, $w \times h$ in the chip C0 Raman) to compare the Raman signal enhancement induced by the MRR. Both the RE-WGR and WGR experiments were done for TE polarization.

As already discussed in Chapter 5, for the backscattering WGR measurement, a ratio of the Stokes Raman signal power to the incident pump power can be defined as [1]

$$\frac{P_{\text{wg}}}{P_{\text{pump}}} = \rho \sigma(\omega_p, \omega_s) \frac{\eta}{2} \gamma_{\text{in}} \gamma_{\text{out}} \left(\frac{1 - e^{-2\alpha L}}{2\alpha} \right) \quad 6-18$$

Here, we are interested in the α and the last term of Equation 6-18 $(1 - e^{-2\alpha L})/2\alpha$. In this equation, the α is the propagation loss in a unit of /cm. We define the last term of the equation as an effective length factor. Figure 6-11 shows the calculated length factor of the Al_2O_3 waveguides for both TE and TM mode. The curve and the saturation length were calculated using the measured propagation loss for TE (1.1136 /cm) and TM mode (1.502 /cm). A length of the MRR 51 on Chip A2 and a length of the spiral on Chip C0 were marked in the figure. We define the length of the MRR as the sum of the length of the bus waveguide from the end facet to the center of the ring and the circumference of the ring.

As can be seen in Figure 6-11, the length of the MRR and the spiral do not yield the same saturation length. Thus, the measurement data were scaled to the saturation length for better comparison between the RE-WGR and WGR signal and between the different polarizations.

Figure 6-11 The calculated length factor based on the propagation loss of the Al_2O_3 waveguide for TM polarization. The saturation length (L_s) of the waveguide for TE polarization is about 0.90 cm and TM is about 0.67 cm.

Figure 6-12 The RE-WGR measurement of toluene. (A) is the RE-WGR measurement for TM and (B) is for TE mode.

Figure 6-13 The RE-WGR and the WGR measurements of toluene for TE mode. (A) is the RE-WGR data shown in Figure 6-12B. (B) is the WGR measurement for TE polarization.

The signal intensity of the measurement is affected by several parameters such as the coupling efficiency, the propagation loss and the length of the waveguide. Thus, the raw data has to be scaled to compare the results fairly. All the measurement data was normalized by a coupling efficiency and scaled to the saturation lengths. Because it is very difficult to determine the exact coupling efficiency of the waveguide, the coupling efficiencies were calculated by a theoretical model using Lumerical mode solutions. Table 6-4 shows the coupling efficiencies and the saturation length of the waveguide.

Table 6-4 the saturation length and the coupling efficiency of the waveguide for TE and TM polarization

Mode	saturation length, L_s (cm)	coupling efficiency
TE ₀₀	0.8980	0.0193
TM ₀₀	0.6658	0.0226

Figure 6-12 shows the RE-WGR signal of toluene for TE and TM mode at different incident power. Gray dotted lines on the figure represent the location of toluene peaks (520, 786, 1003, 1030 and 1211 cm^{-1}). In Chapter 5, it has been demonstrated that the TM mode is advantageous for WGR measurement. However, the result here shows that the RE-WGR signal for the TE mode is slightly better than the signal for TM mode. We attribute this difference to the mode profiles in the MRRs. The RE-WGR for the TM mode (Figure 6-12A) shows a stronger peak around 520 cm^{-1} . Figure 6-3B and C show that for the TM mode the electric field extends more to the upper and the bottom cladding than TE mode. Therefore, the guided-light has more interaction with silicon and this is reflected in the Raman spectra.

Toluene RE-WGR signal and toluene WGR signal were also compared in Figure 6-13. Figure 6-13A is the same toluene RE-WGR signal shown in Figure 6-12B. The range of the y-axis was changed to show the intensity difference between the RE-WGR and WGR signals.

The chip A2 and C0 used for the RE-WGR and WGR experiments came from the same wafer and they have the same waveguide dimensions (1.4 $\mu\text{m} \times 350 \text{ nm}$). This means that the two chips have an Al_2O_3 layer with the same material properties and the almost same layer thickness; the layer thickness can be slightly different depending on the uniformity of the deposition process. However, the WGR signal measured on the spiral shows higher

photon counts than the RE-WGR signal, Figure 6-13. There is also a noticeable difference in the spectra profile. WGR signal shows a more intense broadband background around 800 to 1000 cm^{-1} .

To compare the spectra obtained under different experimental conditions, the spectra have to be modified to make the comparison numerically fair. The signal to background ratio has been used for the comparison. The signal to background ratio (SNB) is defined as C_{sig} / C_{bg} , where C_{sig} is the signal count and C_{bg} is the background count which is the averaged counts near the bottom of a signal peak. C_{sig} and C_{bg} are graphically described in Figure 6-14. For the SNB calculation, we extracted the signal photon counts C_{sig} at 1003 cm^{-1} and the background photon counts C_{bg} near the bottom of the peak at 1003 cm^{-1} .

Figure 6-14 An example of the signal count C_{sig} and the background count C_{bg} .

Table 6-5 The signal to background ratio of toluene Raman signal at 1003 cm^{-1}

incident optical power (mW)	signal to background ratio (C_{sig} / C_{bg})	
	RE-WGR	WGR
13.5	64.23	55.57
16.9	132.16	50.40
20.2	83.06	53.14
22.5	95.69	27.78

The extracted SNB in Table 6-5 shows that the performance of the RE-WGR is slightly better than the performance of WGR.

Originally, we expected that the MRR enhances the Raman signal and the RE-WGR can probe an analyte with a lower incident power due to the resonance. To study the signal enhancement induced by the MRR, a series of measurements with the different incident optical powers were conducted.

Interestingly, the SNB of the WGR signal seems to be independent of the incident power. The coupled optical power contributes equally to the background and the signal. Therefore, the change in incident power does not affect the SNB of the signal. On the other hand, the SNB of the RE-WGR was increased by the incident optical power. This shows a small signal improvement induced by the MRR.

6.7 Conclusion and Future works

WaveGuide Raman spectroscopy (WGR) was demonstrated to probe an analyte on top of the waveguide. However, the signal quality of WGR is limited by the inherent Raman and/or fluorescent background of the material. To improve the signal quality of WGR, micro Ring resonator (MRR) Enhanced WaveGuide Raman spectroscopy (RE-WGR) was suggested. When the MRR is on-resonance, the MRR can build up the intensity in the ring. Thus, we assumed that the WGR signal can be locally enhanced by the MRR when it is on-resonance. Al_2O_3 MRRs were designed and fabricated for the RE-WGR experiment.

An analyte, Toluene, was measured with the MRR for TE and TM polarization. In contrast to the result discussed in Chapter 5, the RE-WGR for TE polarization shows slightly better signal quality than for TM polarization. We believe this depends on the waveguide design parameters.

To study the signal enhancement, we measured toluene on the MRR and the spiral waveguide and the results were compared. The result shows that the SNB of WGR is independent of laser power. The SNB of the RE-WGR shows a trend that the SNB is increased by raising the laser power, but the increase of the SNB is small. It shows that the waveguide Raman signal can be improved by using a microring resonator.

To improve the RE-WGR result, the RE-WGR experiments have to be done with a narrower band tunable laser. Because we conducted the experiment with a wavelength fixed laser, we were not able to characterize the MRR in a better way. By tuning the laser, many parameters of the MRR (e.g., cross-coupling coefficient and Q factor) can be extracted. Furthermore, the waveguide fabrication process is known to contain process tolerance. Process tolerance can change the shape of the final waveguide. These changes will prevent MRR from working as designed. The same issues can be compensated by using a tunable laser. Thus, a tunable laser having a narrower bandwidth will have to be required for the RE-WGR experiment.

In Chapter 5 and 6, use of WGR and RE-WGR were proposed to measure the Raman signal of EVs. The waveguide approaches require surface chemistry to functionalize the waveguide surface to capture EVs. However, surface chemistry cannot be done within a given period so that only proof of the concept experiments were performed by measuring toluene. Toluene has a very distinctive Raman fingerprint so that toluene was used for a feasibility check. Therefore, proposed approaches have to be tested with a model system such as measuring polystyrene rin beads, and surface functionalization to bring EVs to the waveguide surface has to be done as a next step.

In this chapter, we performed an indirect comparison of the effect of the resonance due to the absence of the narrowband tunable laser. A tunable laser with a narrower bandwidth will allow a more accurate study of the signal enhancement by the MRR. With this type of laser, on- and off-resonance can be precisely controlled. Probing an analyte on- and off-resonance will give more insight into the effect of the resonance on the signal enhancement and will allow a direct comparison between the two conditions.

Bibliography

- [1] A. Dhakal, A. Z. Subramanian, P. Wuytens, F. Peyskens, N. Le Thomas, R. Baets, "Evanescent excitation and collection of spontaneous Raman spectra using silicon nitride nanophotonic waveguides", *Optics Letters*, **2014**; *39*, 4025.
- [2] A. Dhakal, P. C. Wuytens, F. Peyskens, K. Jans, N. L. Thomas, R. Baets, "Nanophotonic waveguide enhanced Raman spectroscopy of biological submonolayers", *Acs Photonics*, **2016**; *3*, 2141.
- [3] S. A. Holmstrom, T. H. Stievater, D. A. Kozak, M. W. Pruessner, N. Tyndall, W. S. Rabinovich, R. A. McGill, J. B. Khurgin, "Trace gas Raman spectroscopy using functionalized waveguides", *Optica*, **2016**; *3*, 891.
- [4] N. F. Tyndall, T. H. Stievater, D. A. Kozak, K. Koo, R. A. McGill, M. W. Pruessner, W. S. Rabinovich, S. A. Holmstrom, "Waveguide-enhanced Raman spectroscopy of trace chemical warfare agent simulants", *Optics Letters*, **2018**; *43*, 4803.
- [5] A. Raza, S. Clemmen, P. Wuytens, M. de Goede, A. S. Tong, N. Le Thomas, C. Liu, J. Suntivich, A. G. Skirtach, S. M. Garcia-Blanco, "High index contrast photonic platforms for on-chip Raman spectroscopy", *Optics Express*, **2019**; *27*, 23067.
- [6] C. m. Mo, L. Zhang, C. Xie, T. Wang, "Luminescence of nanometer-sized amorphous silicon nitride solids", *Journal of Applied Physics*, **1993**; *73*, 5185.
- [7] H. Hao, L. Wu, W. Shen, H. Dekkers, "Origin of visible luminescence in hydrogenated amorphous silicon nitride", *Applied Physics Letters*, **2007**; *91*, 201922.
- [8] A. Subramanian, P. Neutens, A. Dhakal, R. Jansen, T. Claes, X. Rottenberg, F. Peyskens, S. Selvaraja, P. Helin, B. Du Bois, "Low-loss singlemode PECVD silicon nitride photonic wire waveguides for 532–900 nm wavelength window fabricated within a CMOS pilot line", *IEEE Photonics Journal*, **2013**; *5*, 2202809.
- [9] J. Heebner, R. Grover, T. Ibrahim, T. A. Ibrahim *Optical microresonators: theory, fabrication, and applications*; Springer Science & Business Media, **2008**; Vol. 138.
- [10] W. M. Green, M. J. Rooks, L. Sekaric, Y. A. Vlasov, "Optical modulation using anti-crossing between paired amplitude and phase resonators", *Optics Express*, **2007**; *15*, 17264.
- [11] P. J. Pauzauskie, D. J. Sirbully, P. Yang, "Semiconductor nanowire ring resonator laser", *Physical Review Letters*, **2006**; *96*, 143903.

- [12] X. Yang, C. W. Wong; Google Patents: **2009**.
- [13] D. G. Rabus, "Realization of optical filters using ring resonators with integrated semiconductor optical amplifiers in GaInAsP/InP", **2002**.
- [14] Q. Xu, M. Lipson, "All-optical logic based on silicon micro-ring resonators", *Optics Express*, **2007**; *15*, 924.
- [15] J. K. Rakshit, J. N. Roy, "Micro-ring resonator based all-optical reconfigurable logic operations", *Optics Communications*, **2014**; *321*, 38.
- [16] D. Rabus, M. Hamacher, U. Troppenz, H. Heidrich, "High-Q channel-dropping filters using ring resonators with integrated SOAs", *IEEE Photonics Technology Letters*, **2002**; *14*, 1442.
- [17] M. Matsuo, H. Yabuki, M. Makimoto, "Dual-mode stepped-impedance ring resonator for bandpass filter applications", *IEEE Transactions on Microwave Theory and Techniques*, **2001**; *49*, 1235.
- [18] A. Ksendzov, Y. Lin, "Integrated optics ring-resonator sensors for protein detection", *Optics Letters*, **2005**; *30*, 3344.
- [19] M. de Goede, L. Chang, J. Mu, M. Dijkstra, R. Obregón, E. Martínez, L. Padilla, F. Mitjans, S. M. Garcia-Blanco, "Al₂O₃:Yb³⁺ integrated microdisk laser label-free biosensor", *Optics Letters*, **2019**; *44*, 5937.
- [20] Y. Sun, X. Fan, "Optical ring resonators for biochemical and chemical sensing", *Analytical and Bioanalytical Chemistry*, **2011**; *399*, 205.
- [21] G.-D. Kim, H.-S. Lee, C.-H. Park, S.-S. Lee, B. T. Lim, H. K. Bae, W.-G. Lee, "Silicon photonic temperature sensor employing a ring resonator manufactured using a standard CMOS process", *Optics Express*, **2010**; *18*, 22215.
- [22] D. G. Rabus *Integrated ring resonators*; Springer, **2007**.
- [23] W. Bogaerts, P. De Heyn, T. Van Vaerenbergh, K. De Vos, S. Kumar Selvaraja, T. Claes, P. Dumon, P. Bienstman, D. Van Thourhout, R. Baets, "Silicon microring resonators", *Laser & Photonics Reviews*, **2012**; *6*, 47.

Summary

Chapter 1 and **Chapter 2** provided an introduction to this thesis. In **Chapter 1**, research motivation and outline were covered. In **Chapter 2**, some theoretical background of EVs, Raman spectroscopy, and optical trapping were introduced.

In **Chapter 3**, Raman spectra of four EVs subtypes were obtained and classified based on Principal Component Analysis (PCA). As a first step, we harvested EVs from four different types of cells: red blood cell, platelet, PC3, and LNCaP. For the blood product-derived EVs (RBC and platelet), we obtained concentrated RBCs and platelets from a blood bank in the Netherlands (Sanquin, Amsterdam, the Netherlands) and isolated EVs by centrifuging them three times. For the prostate cancer cell line-derived EVs (PC3 and LNCaP), we cultured cell lines, took cell culture medium and centrifuged collated supernatant to isolate EVs released from the cell lines. The prepared samples were validated by Nanoparticle Tracking Analysis NTA and Transmission Electron Microscopy (TEM).

Recording the Raman spectra of such a tiny particle is challenging since the size of the particle (~ 130 nm) is smaller than the diffraction limit of the system (~ 340 nm). Thus, we used Raman microscopy combined with optical tweezers. We trapped EVs in the focus of the laser beam and recorded Raman spectra of EVs using the same laser.

The Raman spectra that were obtained from the four EVs subtypes do not show very distinct spectral differences. PCA was used to classify the Raman spectra of EVs. Baseline correction and the background removal from the Raman spectra were performed before the PCA. The multivariate analysis was carried out on three different spectral areas to find the best spectral range for classification: the fingerprint ($400\text{-}1800\text{ cm}^{-1}$), high frequency ($2700\text{-}3050\text{ cm}^{-1}$) or the full spectrum ($400\text{-}3050\text{ cm}^{-1}$). PCA on the fingerprint region

showed a clear separation of each EV subtypes. Two PCA score plots (PComp3 vs. PComp1 and PComp3 vs. PComp2) separate the prostate cancer EV group from the healthy EV group with 94.67% and 98% respectively of the data being accurately classified. This result indicates the clear discrimination of these two groups based on their spectral fingerprints.

In **Chapter 4**, we developed an automated classifier based on a machine learning algorithm called a Convolutional Neural Network (CNN). Although the EV spectra classifier based on PCA showed quite good performance, it required data preprocessing. The new classifier was compared to the classifier based on PCA. CNN has a feature extracted layer called a convolution layer prior to the fully connected layer. The convolution layer extracts features from the given spectra. Then, the fully connected network is trained on the extracted data. CNN is a supervised machine learning algorithm. In the learning process, the network propagates the input spectrum through the network and compares the output with the given label that corresponds to the spectrum. Then, the network adjusts its parameters based on distance, also known as cost, between the calculated result and given label. This cycle is iteratively performed to reduce the cost of the network.

Prior to the algorithmic analysis, Raman spectra of the four EVs subtypes were collected by the same method described in Chapter 2. We obtained 300 spectra from four EVs subtypes (75 spectra from each EVs subtype). CNN requires a large volume of input data for the learning process. Since the 300 spectra have to be cleft into three subsets for training and testing the trained model, i.e., training, validation, and testing dataset, the volume of input data was not sufficient to train the network effectively. To solve the shortage of training data, we carried out data augmentation by adding randomly generated white Gaussian noise to the original data. As a result, we had 300 Raman spectra for each subtype (in total 1200 spectra for the entire dataset). 240 out of the 300 Raman spectra of each subtype were randomly selected to form a training dataset and the rest of the spectra were also randomly divided into two groups to form the testing and the validation set.

Although the prediction model based on PCA showed a high prediction accuracy (95%, PCA-QDA performed in the fingerprint region), the PCA based classifier is not suitable for raw data handling. The CNN based classifier showed a classification performance with an accuracy of >90% overall. The CNN classifier showed an even better performance for the

raw data compared to the preprocessed data indicating that preprocessing might remove interesting features.

Chapter 5 discussed the intrinsic background of different waveguide core materials, Al_2O_3 , Si_3N_4 and TiO_2 .

In this chapter, waveguide Raman (WGR) spectroscopy was proposed. WGR uses an evanescent field to excite an analyte on a waveguide. Since the evanescent field exponentially decays from the waveguide surface, the field can excite an analyte in close proximity to the waveguide. Long interaction length with the analyte can be achieved by increasing the length of the waveguide. These features make WGR a promising technique for studying thin films, surfaces or nanoparticles.

There are two main parameters that limit the performance of WGR, namely the propagation losses and the intrinsic background of waveguide material. In this chapter, we focused on the background of three waveguide materials, Al_2O_3 , Si_3N_4 and TiO_2 . We designed and fabricated waveguides. The dimension of fabricated waveguides is $1.2\ \mu\text{m}$ wide and $\sim 14\ \text{mm}$ long with a spiral in the middle of the chip. Since the waveguide materials have different refractive indices (Al_2O_3 : 1.66, Si_3N_4 : 2.0 and TiO_2 : 2.30), different core thicknesses were selected, i.e., 346 nm, 127 nm and 181 nm for Al_2O_3 , Si_3N_4 and TiO_2 , respectively. A near-infrared laser at 785 nm was coupled into the waveguides for the WGR with an optical power of 36 mW. The incident beam was p-polarized (horizontally polarized) or s-polarized (vertically polarized) to excite the quasi-TE or quasi-TM mode, respectively.

First, we measured the air clad waveguides to investigate the fluorescence or Raman background of the waveguide materials. An analyte, toluene, was then loaded on the waveguide chip. The background from Al_2O_3 , Si_3N_4 and TiO_2 for TE and TM were compared. In contrast to the TE mode, the TM mode for the Si_3N_4 and TiO_2 waveguides show a reasonably flat baseline above 1000cm^{-1} . The TE and TM background from the Al_2O_3 waveguide has a similar background level. We attributed this to the size of the waveguide. The Al_2O_3 has a lower refractive index (1.66) than the other two core materials. The thickness of the Al_2O_3 waveguide core, therefore, is thicker than for the others. Moreover, the TE and TM mode overlap with the core shows a similar value (~ 0.656). Due to the similar mode overlap, the waveguide shows a similar background level for the TE and TM modes. The Si_3N_4 is known to have a fluorescence background which can be excited in the near-infrared. The TM mode has less fluorescence background for Si_3N_4 than the TE

mode between $360 - 870 \text{ cm}^{-1}$. TiO_2 showed the biggest background change among the three core materials.

Chapter 6 suggests the use of an integrated microring resonator (MRR) for the WGR. Although WGR is a promising technique to characterize nano-structures or thin films, its performance is restricted by the inherent backgrounds of the materials. To minimize the background contribution, it is demonstrated that the TM polarization can minimize the Raman or fluorescent background of the waveguide core materials in Chapter 5.

In the WGR system, an incident optical power equally contributes to the background generation and the Raman excitation. The WGR requires studies of low background material and new fabrication processes to improve the performance of the WGR. Thus, this chapter approached the background issue from a different angle.

7

We assume that the field enhancement induced by an MRR may improve the Raman signal power rather than the background generation. If this assumption is true, a high SNR waveguide Raman signal can be achieved for a lower incident laser power. The low laser power generates a low level of the background signal, but the Raman signal can be enhanced due to the field enhancement occurred at the resonance.

Al_2O_3 MRRs were designed and fabricated for the RE-WGR experiments that guide a quasi TE and a quasi TM mode. The fabricated Al_2O_3 MRR has a waveguide structure of $1.4 \mu\text{m}$ width and 350 nm height and a ring with a $200 \mu\text{m}$ radius and a 600 nm gap between the bus waveguide and the ring. Spiral waveguides were also fabricated with the same waveguide dimension ($1.4 \mu\text{m} \times 350 \text{ nm}$) to study the effect of resonance. As an analyte, toluene was measured on both of the MRR and the spiral waveguide.

It was observed that the RE-WGR improved the Raman signal of toluene. Though the WGR did not show big changes in the signal to background ratio (SNB) with the incident power change, the RE-WGR shows an increasing trend of the SNB by raising the incident laser power.

HOOFDSTUK 8

Samenvatting

Hoofdstuk 1 en **Hoofdstuk 2** gaven een introductie van dit proefschrift. **Hoofdstuk 1** behandelde de motivatie van het onderzoek en een overzicht van dit proefschrift. **Hoofdstuk 2** introduceerde de theoretische achtergrond betreffende EVs, Raman spectroscopie en optische pincetten.

Hoofdstuk 3 besprak de acquisitie van Raman spectra van vier EV subtypes en diens classificatie met hoofdcomponentenanalyse. Eerst werden de EVs verzameld van vier celtypes: rode bloedcellen, trombocyten, PC3 en LNCap. De EVs uit de eerste twee celtypes waren verkregen uit geconcentreerde rode bloedcellen en trombocyten van een Nederlandse bloedbank (Sanquin, Amsterdam, Nederland) en waren vervolgens geïsoleerd door het monster driemaal te centrifugeren. De EVs uit de laatste twee celtypes zijn verkregen vanuit gekweekte cellijnen, waarvan het supernatant van het cel medium gecentrifugeerd werd om de EVs te isoleren. De verkregen monsters waren geïnspecteerd en gevalideerd aan de hand van analyse van de beweging van de EVs en transmissie elektron microscopie.

Het verkrijgen van een Raman spectrum van een EV is lastig omdat de afmetingen (~ 130 nm) kleiner zijn dan de diffractielimiet van het gebruikte systeem. Daarom is Raman spectroscopie gecombineerd met optische pincetten. EVs werden gevangen in de focus van een laser, welke ook gebruikt werd om de Raman spectra vast te leggen.

De Raman spectra verkregen van de vier EV subtypes vertonen niet grote spectrale verschillen. Hoofdcomponentenanalyse werd toegepast om de Raman spectra van de EVs te classificeren. Eerst werd op het signaal de basislijncorrectie uitgevoerd samen met de verwijdering van de achtergrond. De multivariate analyse werd uitgevoerd voor drie spectrale bereiken om het beste bereik voor classificatie te vinden: de vingerafdruk (400-

1800 cm^{-1}), hoge frequenties (2700-3050 cm^{-1}) of het volledige spectrum (400-3050 cm^{-1}). Een duidelijk onderscheid tussen de EV subtypes was zichtbaar voor hoofdcomponentenanalyse uitgevoerd op de vingerafdruk. De score diagrammen van deze analyse (PComp3 vs PComp1 en PComp3 vs PComp2) tonen een duidelijk verschil tussen de prostaatkanker EVs en de gezonde EVs met een klassificatiescore van respectievelijk 94.67% en 98%. Dit resultaat toont aan dat er een onderscheid gemaakt kan worden tussen deze twee groepen EVs gebaseerd op hun spectrale vingerafdruk.

In **Hoofdstuk 4** werd een automatische beoordelaar ontwikkeld gebaseerd op het machinaal leren algoritme convolutioneel neurale netwerk (CNN). Hoewel de classificatie van EV spectra gebaseerd op hoofdcomponentenanalyse goede resultaten opleverde, moest de data voorbereid worden voordat de analyse uitgevoerd werd. De nieuwe beoordelaar was vergeleken met die gebaseerd op hoofdcomponentenanalyse. CNN bevat een convolutielaag die kenmerken extraheert van de spectra. Erna is het netwerk getraind met de geëxtraheerde data. CNN is een gecontroleerde machinaal leren algoritme. Gedurende het leerproces propageert het netwerk input spectra door het netwerk en vergelijkt het resultaat met het label van het spectrum. Dan past het netwerk zijn parameters gebaseerd op de afstand, of de kost, tussen het berekende resultaat en het gegeven label. Deze cyclus wordt iteratief uitgevoerd om de kost van het netwerk te verminderen.

Voordat de algoritmische analyse werd uitgevoerd, werden Raman spectra van de vier EV subtypes verzameld volgens dezelfde methode zoals beschreven in Hoofdstuk 2. Van de vier EV subtypes zijn elk 75 spectra verkregen. CNN vereist een grote hoeveelheid input data voor het leerproces. Aangezien de data opgesplitst moest worden in drie subgroepen voor het trainen en testen van het model, was de hoeveelheid input data niet voldoende om het netwerk effectief te trainen. Dit tekort werd opgelost door de hoeveelheid data te vermeerderen door ruis volgens een Gaussische verdeling toe te voegen aan de oorspronkelijke spectra. Zodoende zijn 300 spectra verkregen voor elk subtype. Daarvan zijn 240 spectra willekeurig gekozen voor een trainingsdataset en de overige spectra zijn willekeurig verdeeld in twee groepen: een voor het testen en een voor de validatie.

Hoewel het voorspellingsmodel gebaseerd op de hoofdcomponentenanalyse een hoge voorspellingsnauwkeurigheid vertoonde (95%, PCA-QDA in vingerafdruk bereik), is deze niet geschikt voor het omgaan met ruwe data. De beoordelaar gebaseerd op de CNN had een classificatienauwkeurigheid beter dan 90%. Deze bood ook een beter resultaat voor de

ruwe data vergeleken met de bewerkte data, wat suggereert dat de vooranalyse kenmerken verwijdt.

Hoofdstuk 5 behandelde de intrinsieke achtergrond van de verschillende golfgeleider materialen Al_2O_3 , Si_3N_4 en TiO_2 .

In dit Hoofdstuk werd golfgeleider Raman spectroscopie voorgesteld. Hiervoor wordt gebruikt gemaakt van een evanescente golf om een analyt op een golfgeleider te exciteren. Omdat het evanescente golf exponentieel afneemt buiten de golfgeleider kan het elektrische veld enkel een analyt in de nabijheid van de golfgeleider exciteren. Een lange interactielengte met het analyt kan bereikt worden door de golfgeleider te verlengen. Deze eigenschap maakt golfgeleider Raman spectroscopie een veelbelovende techniek voor het bestuderen van dunne lagen, oppervlakken en nanodeeltjes.

Twee parameters limiteren het gebruik van golfgeleider Raman spectroscopie, namelijk het optische verlies en de intrinsieke achtergrond van het materiaal. Dit Hoofdstuk richtte zich op de bepaling van de achtergrond van de drie golfgeleider materialen Al_2O_3 , Si_3N_4 en TiO_2 . Golfgeleiders werden ontworpen en gefabriceerd. De golfgeleiders waren $1.2\text{ }\mu\text{m}$ breed en $\sim 14\text{ mm}$ lang. Omdat de gebruikte materialen elk een andere brekingsindex hebben (Al_2O_3 : 1.66, Si_3N_4 : 2.0 en TiO_2 : 2.3), waren verschillende golfgeleider hoogtes gebruikt. Al_2O_3 had een hoogte van 364 nm , Si_3N_4 had een hoogte van 127 nm en TiO_2 had een hoogte van 181 nm . Een laser met golflengte van 785 nm werd in de golfgeleiders gekoppeld met een optisch vermogen van 36 mW . Het binnenkomende licht had een horizontale of verticale polarisatie om respectievelijk de quasi-TE of quasi-TM golf in de golfgeleider aan te slaan.

Eerst werden de fluorescentie en Raman achtergrond van de golfgeleiders zonder bedekking gemeten. Erna werden de golfgeleiders in tolueen ondergedompeld. De achtergrond van de drie golfgeleider materialen werd met elkaar vergeleken voor zowel de TE als TM polarisatie. De TM golf toonde een nagenoeg vlakke achtergrond boven 1000 cm^{-1} , in tegenstelling tot de TE golf. Verder werd gemeten dat de TE en TM golf een soortgelijke achtergrond hebben in de Al_2O_3 golfgeleider. Dit werd gelinkt aan de afmetingen van de golfgeleider. Al_2O_3 heeft een lagere brekingsindex waardoor een grotere golfgeleiderkern vereist is. Verder is de overlap van de zowel de TE als TM golf met de golfgeleiderkern ongeveer hetzelfde (~ 0.66), waardoor de achtergrond voor beide polarisaties hetzelfde is. Van Si_3N_4 is bekend dat het een fluorescentie achtergrond heeft die geëxciteerd kan worden in het bijna infrarood. In Si_3N_4 heeft de TM golf een lagere

fluorescentie achtergrond in het bereik $360\text{--}870\text{ cm}^{-1}$. TiO_2 heeft de grootste achtergrond van de drie golfgeleider materialen.

Hoofdstuk 6 introduceert het gebruik van een microring resonator (MRR) voor golfgeleider Raman spectroscopie. Hoewel golfgeleider Raman spectroscopie een veelbelovende techniek is voor het karakteriseren van nanostructuren of dunne lagen, wordt het beperkt door de intrinsieke achtergrond van de gebruikte materialen. Om de bijdrage van de Raman of fluorescentie achtergrond te reduceren, werd in Hoofdstuk 5 aangetoond dat TM polarisatie gebruikt kan worden.

In golfgeleider Raman spectroscopie draagt het invallende licht evenveel bij aan het opwekken van het achtergrond signaal als de Raman excitatie. Om de prestaties van deze methode te verbeteren moeten nieuwe materialen en fabricatietechnieken bestudeerd worden om de achtergrond te verlagen.

Het wordt verondersteld dat de veldversterking van een MRR enkel het Raman signaal versterkt in plaats van de achtergrond. Dat zou een hogere SNR opleveren en het geconsumeerde laser vermogen verminderen. Een lager laser vermogen levert een lagere achtergrond op, maar het Raman signaal kan versterkt worden vanwege de veldversterking die plaatsvindt op de resonantie.

Al_2O_3 MRRs die zowel een TE als TM golf ondersteunen werden ontworpen en gefabriceerd. De geproduceerde golfgeleider heeft een breedte van $1.4\text{ }\mu\text{m}$ en een hoogte van 350 nm . De MRR heeft een straal van $200\text{ }\mu\text{m}$ en een input waveguide 600 nm daarvan verwijderd. Golfgeleider spiralen zijn ook gerealiseerd met dezelfde golfgeleider dimensies. Tolueen werd gebruikt als analyt voor zowel de MRR als de spiraal golfgeleiders.

Het was geobserveerd dat de MRR resonantie het Raman signaal van tolueen versterkte. Hoewel golfgeleider Raman spectroscopie geen veranderingen toonde van de signaal-tot-achtergrond verhouding als functie van invallend optisch vermogen, nam de signaal-tot-achtergrond verhouding van MRR golfgeleider Raman spectroscopie toe met invallend optisch vermogen.

APPENDIX A

Convolution Neural Network for Classification of Raman spectra of EVs

A.1 The Python script coded for realization of the CNN

```
import tensorflow as tf
import numpy as np
import matplotlib.pyplot as plt
from tkinter import *
import pickle

tf.reset_default_graph()

#Define functions to count time taken for training process
def tic():
    import time
    global startTime_for_tictoc
    startTime_for_tictoc = time.time()

def toc():
    import time
    if 'startTime_for_tictoc' in globals():
        time_counted = "Elapsed time is " + str(
            '{:.2f}'.format(time.time() - startTime_for_tictoc)) + "
sec" + " (" + str(
            '{:.2f}'.format((time.time() - startTime_for_tictoc) / 60))
+ " min)"
        print(time_counted)
    else:
        print("Toc: start time not set")
    np.savetxt("time.txt", [time_counted], fmt="%s")

#Define Leaky ReLU function
def leaky_relu(x, alpha):
    return tf.nn.relu(x) - alpha * tf.nn.relu(-x)

tic()
```

```

gpu_options = tf.GPUOptions(per_process_gpu_memory_fraction=0.8)
tf.device('/gpu:0')

w = 0.03

# select activation function for convolution layer and fully connected
layer
# 1: ReLU, 2: Leaky ReLU, 3: tanh
act_fuction_selector_CL = 2
act_fuction_selector = 3

if act_fuction_selector == 1:
    print("ReLU is selected as an activation function!!!\n")
elif act_fuction_selector == 2:
    print("Leaky ReLU is selected as an activation function!!!\n")
elif act_fuction_selector == 3:
    print("tanh is selected as an activation function!!!\n")

# set file path for saving model and session for validation and
# tensorboard, respectively
DataNorm = 0 # Normalize chopped data before training

# File path for learning structure and data for input
FilePath_learn = "C:/Users/" # File path to save trained model
FilePath_data = "C:/Users/" # File path for training and validation
# data set will be used for training

# <structure info>: Tensor bard, trained model, structural information
# for testing and cost change over time
Filepath_board = FilePath_learn + "board"
Filepath_model = FilePath_learn + "model.ckpt"
Filepath_structure = FilePath_learn + "structure.dat"
Filepath_history = FilePath_learn + "history.txt"

# <Training and validation dataset>
Input_training = FilePath_data + "/Training.csv"
Input_validation = FilePath_data + "/Validation.csv"

# <Label for the training and validation>
Input_label = FilePath_data + "/label_train.csv"
Input_validation_label = FilePath_data + "/label_validation.csv"

structure_exp1 = 'Number of convolutional layers : 3'
structure_exp2 = 'Number of fully connected layers : 5'
output_exp = 'Number of classes: 4'

fileType_ErAc = ".csv"
fileName_ErAc1 = "Training_accuracy"
fileName_ErAc2 = "Validation_accuracy"
fileName_ErAc3 = "Training_cost"
fileName_ErAc4 = "Validation_cost"
fileName_ErAc5 = "Patient_count"

fullFile_ErAc1 = FilePath_learn+fileName_ErAc1+fileType_ErAc
fullFile_ErAc2 = FilePath_learn+fileName_ErAc2+fileType_ErAc

```

```

fullFile_ErAc3 = FilePath_learn+fileName_ErAc3+fileType_ErAc
fullFile_ErAc4 = FilePath_learn+fileName_ErAc4+fileType_ErAc
fullFile_ErAc5 = FilePath_learn+fileName_ErAc5+fileType_ErAc

# set training condition (size of batch, traning epochs, number of
classes, learning rate)
training_epoch = 1000
divider = int(1)
batch_size = 32
nb_class = 4
learning_rate = 0.0001

# Set condition for early stopping
patience = 200
min_delta = 0.02

## Select spectral range for the training
## Values are not wavenumber
## Fingerprint      : 0-655
## Full spectrum   : 0-1152
## High freq       : 1021-1152

spectral_band_start = type_the_number
spectral_band_end = type_the_number

# Define structure of convolutional layers
# 1st convolution layer
filter_size1 = 5
num_of_filter1 = 30
pool_size1 = 5
mp1_stride = 1
# 2nd convolution layer
filter_size2 = 5
num_of_filter2 = 30
pool_size2 = 5
mp2_stride = 2
# 3rd convolution layer
filter_size3 = 5
num_of_filter3 = 30
pool_size3 = 10
mp3_stride = 3

# Define structure of fully connected layers
# Number of neurons in each hidden layer.
hd_layer1 = 1000
hd_layer2 = 500
hd_layer3 = 200
hd_layer4 = 200

# offset for Leaky ReLU
alpha1 = 0.01
alpha2 = 0.01
alpha3 = 0.01

# Load and reshape input and label data

```



```

with tf.name_scope('Input') as scope:
    x_data = np.loadtxt(Input_training, delimiter=',', dtype=np.float32)
    y_data = np.loadtxt(Input_label, delimiter=',', dtype=np.string_)
    y_data = tf.one_hot(y_data, depth=nb_class)

    y_data =
y_data.eval(session=tf.Session(config=tf.ConfigProto(gpu_options=gpu_opt
ions)))
    x_data = x_data[:, spectral_band_start:spectral_band_end]

#####
# Input data Normalization for training
if DataNorm == 1:
    Normalized_x_data = np.zeros(np.shape(x_data))
    for i in range(len(Normalized_x_data)):
        Normalized_x_data[i, :] = (x_data[i, :]-
min(x_data[i, :]))/(max(x_data[i, :])-min(x_data[i, :]))
    x_data = Normalized_x_data
    del Normalized_x_data
#####

data_length = x_data.shape[1]

X = tf.placeholder(tf.float32, [None, data_length])
Y = tf.placeholder(tf.float32, [None, nb_class])
keep_prob = tf.placeholder(tf.float32)

X_spectra = tf.reshape(X, [-1, data_length, 1])

print('Input dimension: ', data_length)

#####
# Convolutional layers
# 1st convolution layer
with tf.name_scope('Convolutional_layer') as scope:
    #####
    with tf.name_scope('Convolutional_layer1'):
        with tf.name_scope('Filter'):
            F1 = tf.Variable(tf.random_normal([filter_size1, 1,
num_of_filter1, stddev=0.01], name='filter1')
            with tf.name_scope('Convolution'):
                CL1 = tf.nn.conv1d(X_spectra, F1, stride=1, padding='SAME')
            with tf.name_scope('Activation'):
                if act_fuction_selector_CL == 1 :
                    CL1 = tf.nn.relu(CL1)
                elif act_fuction_selector_CL == 2 :
                    CL1 = leaky_relu(CL1, alpha1)
                elif act_fuction_selector_CL == 3 :
                    CL1 = tf.tanh(CL1)
            with tf.name_scope('Max_pooling'):
                CL1 = tf.layers.MaxPooling1D(pool_size=pool_size1,
strides=mp1_stride, padding='SAME')(CL1)
            #####
# 2nd convolution layer
            with tf.name_scope('Convolutional_layer2'):
                with tf.name_scope('Filter'):
                    F2 = tf.Variable(tf.random_normal([filter_size2,

```

```

num_of_filter1, num_of_filter2], stddev=0.01),
    name='filter2')
with tf.name_scope('Convolution'):
    CL2 = tf.nn.conv1d(CL1, F2, stride=1, padding='SAME')
with tf.name_scope('Activation'):
    if act_fuction_selector_CL == 1 :
        CL2 = tf.nn.relu(CL2)
    elif act_fuction_selector_CL == 2 :
        CL2 = leaky_relu(CL2, alpha2)
    elif act_fuction_selector_CL == 3 :
        CL2 = tf.tanh(CL2)
with tf.name_scope('Max pooling'):
    CL2 = tf.layers.MaxPooling1D(pool_size=pool_size2,
strides=mp2_stride, padding='SAME')(CL2)
#####
# 3rd convolution layer
with tf.name_scope('Convolutional_layer3'):
    with tf.name_scope('Filter'):
        F3 = tf.Variable(tf.random_normal([filter_size3,
num_of_filter2, num_of_filter3], stddev=0.01),
            name='filter3')
    with tf.name_scope('Convolution'):
        CL3 = tf.nn.conv1d(CL2, F3, stride=1, padding='SAME')
    with tf.name_scope('Activation'):
        if act_fuction_selector_CL == 1 :
            CL3 = tf.nn.relu(CL3)
        elif act_fuction_selector_CL == 2 :
            CL3 = leaky_relu(CL3, alpha3)
        elif act_fuction_selector_CL == 3 :
            CL3 = tf.tanh(CL3)
    with tf.name_scope('Max pooling'):
        CL3 = tf.layers.MaxPooling1D(pool_size=pool_size3,
strides=mp3_stride, padding='SAME')(CL3)

FCN_input_dimension = int(CL3.shape[1]) * int(CL3.shape[2])
CL3 = tf.reshape(CL3, [-1, FCN_input_dimension])

print('Fully connected layer input dimension: ', FCN_input_dimension)
#####
# Fully connected layers
# 1st hidden layer
with tf.name_scope('Fully_connected_layer') as scope:
    with tf.name_scope('Hidden_layer1'):
        with tf.name_scope('Weight'):
            W1 = tf.get_variable("weight1", shape=[FCN_input_dimension,
hd_layer1], initializer=tf.contrib.layers.xavier_initializer())
            W1_hist = tf.summary.histogram("weight1", W1)
        with tf.name_scope('Bias'):
            b1 = tf.Variable(tf.random_normal([hd_layer1]),
name='bias1')
        with tf.name_scope('Activation'):
            if act_fuction_selector == 1:
                FCNL1 = tf.nn.relu(tf.matmul(CL3, W1) + b1)
            elif act_fuction_selector == 2:
                FCNL1 = leaky_relu(tf.matmul(CL3, W1) + b1, 0.1)
            elif act_fuction_selector == 3:
                FCNL1 = tf.nn.tanh(tf.matmul(CL3, W1) + b1)

```

```

        with tf.name_scope('Drop_out'):
            FCNL1 = tf.nn.dropout(FCNL1, keep_prob=keep_prob)
#####
# 2nd hidden layer
        with tf.name_scope('Hidden_layer2'):
            with tf.name_scope('Weight'):
                W2 = tf.get_variable("weight2", shape=[hd_layer1,
hd_layer2], initializer=tf.contrib.layers.xavier_initializer())
                W2_hist = tf.summary.histogram("weight2", W2)
            with tf.name_scope('Bias'):
                b2 = tf.Variable(tf.random_normal([hd_layer2]),
name='bias2')
            with tf.name_scope('Activation'):
                if act_fuction_selector == 1:
                    FCNL2 = tf.nn.relu(tf.matmul(FCNL1, W2) + b2)
                elif act_fuction_selector == 2:
                    FCNL2 = leaky_relu(tf.matmul(FCNL1, W2) + b2, 0.1)
                elif act_fuction_selector == 3:
                    FCNL2 = tf.nn.tanh(tf.matmul(FCNL1, W2) + b2)
            with tf.name_scope('Drop_out'):
                FCNL2 = tf.nn.dropout(FCNL2, keep_prob=keep_prob)
#####
# 3rd hidden layer
        with tf.name_scope('Hidden_layer3'):
            with tf.name_scope('Weight'):
                W3 = tf.get_variable("weight3", shape=[hd_layer2,
hd_layer3], initializer=tf.contrib.layers.xavier_initializer())
                W3_hist = tf.summary.histogram("weight3", W3)
            with tf.name_scope('Bias'):
                b3 = tf.Variable(tf.random_normal([hd_layer3]),
name='bias3')
            with tf.name_scope('Activation'):
                if act_fuction_selector == 1:
                    FCNL3 = tf.nn.relu(tf.matmul(FCNL2, W3) + b3)
                elif act_fuction_selector == 2:
                    FCNL3 = leaky_relu(tf.matmul(FCNL2, W3) + b3, 0.1)
                elif act_fuction_selector == 3:
                    FCNL3 = tf.nn.tanh(tf.matmul(FCNL2, W3) + b3)
            with tf.name_scope('Drop_out'):
                FCNL3 = tf.nn.dropout(FCNL3, keep_prob=keep_prob)
#####
# 4th hidden layer
        with tf.name_scope('Hidden_layer4'):
            with tf.name_scope('Weight'):
                W4 = tf.get_variable("weight4", shape=[hd_layer3,
hd_layer4], initializer=tf.contrib.layers.xavier_initializer())
                W4_hist = tf.summary.histogram("weight4", W4)
            with tf.name_scope('Bias'):
                b4 = tf.Variable(tf.random_normal([hd_layer4]),
name='bias4')
            with tf.name_scope('Activation'):
                if act_fuction_selector == 1:
                    FCNL4 = tf.nn.relu(tf.matmul(FCNL3, W4) + b4)
                elif act_fuction_selector == 2:
                    FCNL4 = leaky_relu(tf.matmul(FCNL3, W4) + b4, 0.1)
                elif act_fuction_selector == 3:
                    FCNL4 = tf.nn.tanh(tf.matmul(FCNL3, W4) + b4)

```

```

        with tf.name_scope('Drop_out'):
            FCNL4 = tf.nn.dropout(FCNL4, keep_prob=keep_prob)
#####
# Output layer connected to softmax
    with tf.name_scope('Hidden_layer5'):
        with tf.name_scope('Weight'):
            W5 = tf.get_variable("weight5", shape=[hd_layer4, nb_class],
initializer=tf.contrib.layers.xavier_initializer())
            W5_hist = tf.summary.histogram("weight5", W5)
        with tf.name_scope('Bias'):
            b5 = tf.Variable(tf.random_normal([nb_class]), name='bias5')
        with tf.name_scope('Hypothesis'):
            logits = tf.matmul(FCNL4, W5) + b5
            hypothesis = tf.nn.softmax(logits)

#####
# Cost evaluation and taining
    with tf.name_scope('Cost') as scope:
        cost_i = tf.nn.softmax_cross_entropy_with_logits(logits=logits,
labels=Y)
        cost = tf.reduce_mean(cost_i)
        cost_hist = tf.summary.scalar('cost', cost)
    with tf.name_scope('Train') as scope:
        optimizer =
tf.train.AdamOptimizer(learning_rate=learning_rate).minimize(cost)

#####
# save the trained model
saver = tf.train.Saver()

sess = tf.Session(config=tf.ConfigProto(gpu_options=gpu_options))
sess.run(tf.global_variables_initializer())

writer = tf.summary.FileWriter(Filepath_board)
summary = tf.summary.merge_all()
writer.add_graph(sess.graph)

#####
# save the structural details of the trained model
# (number of hidden layers, number of neurons in each hidden layer,
activation function used for training, etc.)
file = open(Filepath_structure, 'wb')

pickle.dump([structure_exp1, structure_exp2, output_exp, nb_class,
spectral_band_start, spectral_band_end,
FilePath_data, DataTypeLearn, DataTypeLabel,
filter_size1, num_of_filter1, pool_size1, mp1_stride,
filter_size2, num_of_filter2, pool_size2, mp2_stride,
filter_size3, num_of_filter3, pool_size3, mp3_stride,
hd_layer1,
hd_layer2,
hd_layer3,
hd_layer4,
act_fuction_selector_CL, act_fuction_selector,
alpha1, alpha2, alpha3,
DataNorm
], file)

```

```
file.close()

print('Structure information is saved.')

#####
# Load data set for validation
correct_prediction = tf.equal(tf.argmax(hypothesis, 1), tf.argmax(Y, 1))
accuracy = tf.reduce_mean(tf.cast(correct_prediction, tf.float32))

x_data_train = x_data
y_data_train = y_data

x_data_validation = np.loadtxt(Input_validation, delimiter=',',
dtype=np.float32)
y_data_validation = np.loadtxt(Input_validation_label, delimiter=',',
dtype=np.string_)
y_data_validation = tf.one_hot(y_data_validation, depth=nb_class)

y_data_validation =
y_data_validation.eval(session=tf.Session(config=tf.ConfigProto(gpu_opti
ons=gpu_options)))
x_data_validation = x_data_validation[:,
spectral_band_start:spectral_band_end]

if DataNorm == 1:
    Normalized_x_data_validation = np.zeros(np.shape(x_data_validation))
    for i in range(len(Normalized_x_data_validation)):
        Normalized_x_data_validation[i, :] = (x_data_validation[i, :]-
min(x_data_validation[i, :]))/(max(x_data_validation[i, :])-
min(x_data_validation[i, :]))
        x_data_validation = Normalized_x_data_validation
    del Normalized_x_data_validation

# Script for learning
print('\nLearning started. It takes some time....\n')

train_size = x_data.shape[0]

# variables to record tracing cost, training accuracy, validation cost,
validation accuracy...
hist_loss = []
hist_c = []
hist_validation_acc = []
hist_train_acc = []
hist_validation_cost = []
hist_patience_cnt = []
hist_slope = []

patience_cnt = 0 #set the early stopping count to '0'

step = 0
for epoch in range(training_epoch):
    avg_cost = 0
    total_batch = int(train_size / batch_size)
    validation_cnt = 0

    for i in range(total_batch):
```

```

        validation_cnt += 1
        step += 1
        batch_mask = np.random.choice(train_size, batch_size)
        x_batch = x_data[batch_mask]
        y_batch = y_data[batch_mask]
        s, c, _ = sess.run([summary, cost, optimizer], feed_dict={X:
x_batch, Y: y_batch, keep_prob: 0.5})
        writer.add_summary(s, global_step=step)
        avg_cost += c / total_batch
        hist_c.append(c)

    ## Cost monitor (averaged)
    hist_loss.append(avg_cost)

    ## Accuracy eveluation after each epoch
    train_acc = sess.run(accuracy, feed_dict={X: x_data_train, Y:
y_data_train, keep_prob: 1}) # Accuracy with train dataset
    hist_train_acc.append(train_acc)
    validation_acc = sess.run(accuracy, feed_dict={X: x_data_validation,
Y: y_data_validation, keep_prob: 1}) # Accuracy with validation dataset
    hist_validation_acc.append(validation_acc)
    validation_cost = sess.run(cost, feed_dict={X: x_data_validation, Y:
y_data_validation, keep_prob: 1})
    hist_validation_cost.append(validation_cost)

    ##Early stopping algorithm
    if epoch < 3:
        slope = 0
        hist_slope.append(slope)
    elif epoch >= 3:
        slope = (hist_validation_cost[epoch] -
hist_validation_cost[epoch - 3])/3
        hist_slope.append(slope)
    if epoch > 50 and slope > 0:
        patience_cnt += 1
    elif epoch > 300 and abs(slope) < 0.002:
        patience_cnt += 1
    elif epoch <= 300 and slope < 0:
        patience_cnt = 0

    hist_patience_cnt.append(patience_cnt)

# show the progress of the learning procedure.
    if patience_cnt >= patience and validation_acc >= 0.9:
        print('Epoch:', '%04d' % (epoch + divider), '      cost :',
' {:.4f}'.format(avg_cost),
'      Training accuracy:', ' {:.3f}'.format(train_acc), '
Validation accuracy:',
' {:.3f}'.format(validation_acc),
'      val_cost:', ' {:.3f}'.format(validation_cost), '
patient_cnt:', '%02d' % patience_cnt)
        print("Early stopping...")
        break
    if epoch % divider == 0:
        print('Epoch:', '%04d' % (epoch + divider), '      cost :',
' {:.4f}'.format(avg_cost),

```

```

        '          Training accuracy:', '{:.3f}'.format(train_acc), '
Validation accuracy:',
        '{:.3f}'.format(validation_acc),
        '          val_cost:', '{:.3f}'.format(validation_cost), '
patient_cnt:', '%02d' % patience_cnt, '      Slope:',
        '{:.5f}'.format(slope))

save_path = saver.save(sess, Filepath_model)
print('\nLearning Finished!!')
print("Model saved in path: %s" % save_path)
print('Structure information is saved.\n')

#####End of Learning#####

toc()

# Save training history (training accuracy, validation accuracy,
training cost, validation cost
# and early stopping count)
np.savetxt(fullFile_ErAc1, hist_train_acc, fmt='%.5f', delimiter=',')
np.savetxt(fullFile_ErAc2, hist_validation_acc, fmt='%.5f',
delimiter=',')

np.savetxt(fullFile_ErAc3, hist_loss, fmt='%.5f', delimiter=',')
np.savetxt(fullFile_ErAc4, hist_validation_cost, fmt='%.5f',
delimiter=',')

np.savetxt(fullFile_ErAc5, hist_patience_cnt, fmt='%.5f', delimiter=',')

####Save the history files to visualize result later
save_train_hist = open(Filepath_history, 'wb')
pickle.dump([hist_train_acc,
             hist_validation_acc,
             hist_loss
             ], save_train_hist)
save_train_hist.close()
SaveResult = FilePath_learn + "Learning_Result.txt"
file = open(SaveResult, 'wb')

pickle.dump([train_acc,
            validation_acc
            ], file, protocol=0)
file.close()
print('Structure information is saved.')

sess.close()

```

APPENDIX B

Wavenumber Calibration

B.1 The Matlab script coded for wavenumber calibration

The software has three separated scripts to convert pixel number to corresponding wavenumber: main script, a function for finding toluene peaks and a function for finding ArHg peaks.

B.1.1 Main script: “Pixel2Wvn.m”

```
%% Get Toluene measerment files
fprintf('Loading dataset ...\n');
[FileNameToluene, PathNameToluene, FilterIndexToluene]=uigetfile({'*.txt'}
, 'MultiSelect', 'on', 'Select Toluene measurement');

if iscellstr(FileNameToluene)
    FileNameToluene=cellstr(FileNameToluene);
else FileNameToluene=cellstr(FileNameToluene);
end
NmeasureToluene=size(FileNameToluene,2);

for N=1:NmeasureToluene
    fprintf('Loading "%s"\n', char(FileNameToluene(N)));
    DataNameTol(N)=fullfile(PathNameToluene, FileNameToluene(N));
end;
fprintf('from "%s"\n', char(PathNameToluene));

%% Get ArHg measuremnet files
fprintf('Loading dataset ...\n');
[FileNameArHg, PathNameArHg, FilterIndexArHg]=uigetfile({'*.txt'}, 'MultiSe
lect', 'on', 'Select the ArHg measurement', PathNameToluene);

if iscellstr(FileNameArHg)
    FileNameArHg=cellstr(FileNameArHg);
else FileNameArHg=cellstr(FileNameArHg);
```



```
end
NmeasureArHg=size(FileNameArHg,2);

for N=1:NmeasureArHg
    fprintf('Loading "%s"\n',char(FileNameArHg(N)));
    DataNameArHg(N)=fullfile(PathNameArHg,FileNameArHg(N));
end;
fprintf('from "%s"\n', char(PathNameArHg));

%% Find and get toluene and ArHg peak locations on EMCCD
% Toluene peak and ArHg peak selection are performed in separated
function
% ToluenePeakFinder.m
% ArHgPeakFinder.m
[TolPeakVal, TolPeakLoc, TolSpect, TolSpectNorm] =
ToluenePeakFinder(DataNameTol,NmeasureToluene);
[ArHgPeakVal, ArHgPeakLoc, ArHgSpect, ArHgSpectNorm] =
ArHgPeakFinder(DataNameArHg,NmeasureArHg);

%% Measured wavelength of excitation beam in nm
excitation_beam = 784.849;

%% Absolute wavelength of the Toluene peaks for exciation of 785nm
if NmeasureToluene == 1 || NmeasureToluene == 2
    ToluenePeakWL=[784.849;
    818.4743742;...
    825.3921322;...
    836.6833085;...
    852.0897093;...
    854.0545833;...
    867.4641812;...
    880.2930423;...
    898.2423057];
elseif NmeasureToluene == 3
    ToluenePeakWL=[784.849;...
    818.4743742;...
    825.3921322;...
    836.6833085;...
    852.0897093;...
    854.0545833;...
    867.4641812;...
    880.2930423;...
    898.2423057;...
    1018.396985;...
    1032.817411
    ];
end;

%% Emission lines of ArHg lamp (wavelength)
if NmeasureToluene == 1
    ArHgPeakWL=[794.818;...
    800.616;...
    811.531;...
    826.452;...
```

```

        842.465;...
        852.144;...
        866.794
    ];
elseif NmeasureToluene == 2 || NmeasureToluene == 3
    ArHgPeakWL=[794.818;...
        800.616;...
        811.531;...
        826.452;...
        842.465;...
        852.144;...
        866.794;...
        912.297;...
        922.450;...
        935.422;...
    ];
end;

%% Convert wavelength to wavenumber
WaveLengthTol=[ToluenePeakWL];
absWaveNumTol=1./ (WaveLengthTol*10^-7);
reciprocalWaveNumTol=1/(excitation_beam*10^-7)-absWaveNumTol;
reciprocalWaveNumTol=sort(reciprocalWaveNumTol);

WaveLengthArHg=[ArHgPeakWL];
absWaveNumArHg=1./ (WaveLengthArHg*10^-7);
reciprocalWaveNumArHg=1/(excitation_beam*10^-7)-absWaveNumArHg;
reciprocalWaveNumArHg=sort(reciprocalWaveNumArHg);

reciprocalWaveNum= [reciprocalWaveNumTol; reciprocalWaveNumArHg];
reciprocalWaveNum=sort(reciprocalWaveNum);

NTolPeak = size(reciprocalWaveNumTol,1);
NArPeak = size(reciprocalWaveNumArHg, 1);

pixels=1:9;
x=1:0.001:9;

%% Lorenz curve fitting and Interpolation
for n=1:NTolPeak
    Tol_select=TolSpectNorm(TolPeakLoc(n)-4:TolPeakLoc(n)+4);
    Tol_select=Tol_select';

    [sigma,mu,A]=myLorenzfit(pixels,Tol_select,0.1);
    Tol_y(n,:)=A*1./((x-mu).^2+sigma);

    Tol_interpolate_pix(n,1)=find(Tol_y(n,:)==max(Tol_y(n,:)));
    Tol_pixels(n,1)=(Tol_interpolate_pix(n,1)-1)/1000+TolPeakLoc(n)-4;
end

for n=1:NArPeak
    ArHg_select=ArHgSpectNorm(ArHgPeakLoc(n)-4:ArHgPeakLoc(n)+4);
    ArHg_select=ArHg_select';

    [sigma,mu,A]=myLorenzfit(pixels,ArHg_select,0.1);
    ArHg_y(n,:)=A*1./((x-mu).^2+sigma);

```

```
ArHg_interpolate_pix(n,1)=find(ArHg_y(n,:)==max(ArHg_y(n,:)));
ArHg_pixels(n,1)=(ArHg_interpolate_pix(n,1)-1)/1000+ArHgPeakLoc(n)-
4;
end

peakpixels = [Tol_pixels; ArHg_pixels];
peakpixels = sort(peakpixels);
CCDpixelsnumber = size(ArHgSpectNorm, 1);
CCDpixels=1:CCDpixelsnumber;

%% Polynomial fitting
N =5;
p=polyfit(peakpixels,reciprocalWaveNum,N);
for a=1:N+1
    CCDwvn_poly(a,:)=p(a)*CCDpixels.^(N+1-a);
    fit_poly(a,:)=p(a)*peakpixels.^(N+1-a);
end
fit=sum(fit_poly);
CCDwvn=sum(CCDwvn_poly);

% CCDwvn is calibrated wavenumber.

%% Function, myLorenzfit
function [sigma,mu,A]=myLorenzfit(x,y,h)

%% threshold
if nargin==2, h=0.2; end

%% cutting
ymax=max(y);
xnew=[];
ynew=[];
for n=1:length(x)
    if y(n)>ymax*h;
        xnew=[xnew,x(n)];
        ynew=[ynew,y(n)];
    end
end

%% fitting

Y=1./ynew;

p=polyfit(xnew,Y,2);
A2=p(1);
A1=p(2);
A0=p(3);

A=1./A2;
mu=-A1/2/A2;
sigma=(A0)/A2-mu^2;
end
```

B.1.2 Toluene peak finding function: “ToluenePeakFinder.m”

```

%% Function to find the emission lines of Toluene
function [TolPeakVal, TolPeakLoc, TolSpect, TolSpectNorm] =
ToluenePeakFinder(DataName, Nmeasurements)

for N=1:Nmeasurements
    I(:, :, N)=importdata(char(DataName(N)));
end;

%% Stitching and Reshape the spectrum depending on spectral region to be
calibrated
window_size = 342;

I = reshape(I, size(I,1), Nmeasurements);
if Nmeasurements == 1
    StRaman = I;
elseif Nmeasurements == 2
    SpectrumA = I(:,1);
    SpectrumB = I(window_size:size(I,1),2);

    IntDiffAB = mean(SpectrumA(980:1000,1)) - mean(SpectrumB(1:20,1));

    StRaman = [SpectrumA; SpectrumB+IntDiffAB];
elseif Nmeasurements ==3
    SpectrumA = I(:,1);
    SpectrumB = I(window_size:size(I,1),2);
    SpectrumC = I(window_size:size(I,1),3);

    IntDiffAB = mean(SpectrumA(980:1000,1)) - mean(SpectrumB(1:20,1));
    SpectrumB = SpectrumB+IntDiffAB;
    IntDiffBC = mean(SpectrumB(638:658,1)) - mean(SpectrumC(1:20,1));

    StRaman = [SpectrumA; SpectrumB; SpectrumC+IntDiffBC];
end;

Spectrum = StRaman;
NSpectrum=(Spectrum-min(Spectrum))/(max(Spectrum)-min(Spectrum));

%% smooth the spectrum to minimize error in peak selection
Spectrum = whittakerSmoother(Spectrum, 1);
NSpectrum = whittakerSmoother(NSpectrum, 1);

%% Find peaks in Fingerprint region/ High frequency region
MidRange = 620;
NumPeakLaser = 1;
NumPeakLow = 5;
NumPeakMid = 3;

SizeLaser = 200;
SizeRaman = 1002;

DumLaser = zeros(SizeLaser,1);
DumSignal = zeros(SizeRaman-SizeLaser,1);

TolSignal = [DumLaser; NSpectrum(SizeLaser+1:SizeRaman)];
DumSigMid(1:MidRange,1) = 0;

```

```

SizeDummyMid = size(DumSigMid,1);
DumSigLow(1:(SizeRaman-SizeDummyMid),1)=0;

Laser = [NSpectrum(1:200,1); DumSignal];
sigLow = [TolSignal(1:SizeDummyMid); DumSigLow];
sigMid = [DumSigMid; TolSignal(SizeDummyMid+1:SizeRaman)];

PixelArea = 1:SizeRaman;
PixelArea = PixelArea';

[PeakValLaser, PeakLocLaser] =
findpeaks(Laser,PixelArea,'MINPEAKHEIGHT',0.02,'MINPEAKDISTANCE',20);
PeakLocLaser = PeakLocLaser(NumPeakLaser);
PeakValLaser = PeakValLaser(NumPeakLaser);
PeakValSigLow=findpeaks(sigLow,PixelArea,'MINPEAKHEIGHT',0.02,'MINPEAKDI
STANCE',5);
PeakSeleLow=PeakValSigLow(1:NumPeakLow);
PeakValSigMid=findpeaks(sigMid,PixelArea,'MINPEAKHEIGHT',0.02,'MINPEAKDI
STANCE',40);
PeakSeleMid=PeakValSigMid(1:NumPeakMid);
if Nmeasurements == 3;
    DumSigHigh = zeros(1499,1);
    sigHigh = [DumSigHigh; NSpectrum(1500:size(NSpectrum,1),1)];

    HighArea = 1:size(sigHigh, 1);
    HighArea = HighArea';
    [PeakValHigh, PeakLocHigh] =
findpeaks(sigHigh,HighArea,'MINPEAKHEIGHT',0.02,'MINPEAKDISTANCE',100,
'MINPEAKWIDTH', 10);
    PeakRatio = PeakValHigh*100;
end

%% Select peaks which we are intrested in
for n=1:(length(PeakValSigLow)-NumPeakLow)
    if PeakValSigLow(n+NumPeakLow) > min(PeakSeleLow)
        loc=find(PeakSeleLow == min(PeakSeleLow));
        PeakSeleLow=[PeakSeleLow(1:loc-1);PeakSeleLow(loc+1:end)];
        PeakSeleLow=[PeakSeleLow;PeakValSigLow(n+NumPeakLow)];
    end
end
PeakValSigLow=PeakSeleLow;
for n=1:NumPeakLow
    PeakLocLow(n)=find(sigLow == PeakValSigLow(n));
end

for n=1:(length(PeakValSigMid)-NumPeakMid)
    if PeakValSigMid(n+NumPeakMid) > min(PeakSeleMid)
        loc=find(PeakSeleMid == min(PeakSeleMid));
        PeakSeleMid=[PeakSeleMid(1:loc-1);PeakSeleMid(loc+1:end)];
        PeakSeleMid=[PeakSeleMid;PeakValSigMid(n+NumPeakMid)];
    end
end

PeakValSigMid=PeakSeleMid;
for n=1:NumPeakMid
    PeakLocMid(n)=find(sigMid == PeakValSigMid(n));
end

```

```

PeakLocAll = [PeakLocLaser PeakLocLow PeakLocMid];
PeakValAll = [PeakValLaser; PeakValSigLow; PeakValSigMid];

if Nmeasurements == 3;
    j = 1;
    for i = 1: size(PeakLocHigh,1);
        if PeakRatio(i,1)>3
            PeakValHigh2(j,1) = PeakValHigh(i,1);
            PeakLocHigh2(j,1) = PeakLocHigh(i,1);
            j = j+1;
        end
    end
    PeakLocAll = [PeakLocAll PeakLocHigh2'];
    PeakValAll = [PeakValAll; PeakValHigh2];
end;

PixNumMatrix = 1:size(NSpectrum,1);
PixNumMatrix = PixNumMatrix';

%% Return selected peaks (pixel location and normalized peak value)
TolPeakVal = PeakValAll;
TolPeakLoc = PeakLocAll';
TolSpect = Spectrum;
TolSpectNorm = NSpectrum;
end

%% Smoothing function, whittaker smoothing algorithm is used
function z = whittakerSmoother(y,lambda)
    m = length(y);
    I = eye(m);
    D = diff(I);
    z = (I + lambda * D' * D)\y;
end

```

B.1.3 ArHg peak finding function: “ArHgPeakFinder.m”

```

%% Function to find the emission lines of ArHg
function [ArHgPeakVal, ArHgPeakLoc, ArHgSpect, ArHgSpectNorm] =
ArHgPeakFinder(DataName, Nmeasurements)

for N=1:Nmeasurements
    I(:, :, N)=importdata(char(DataName(N)));
end;

%% Stitching and Reshape the spectrum depending on spectral region to be
calibrated
window_size = 342;

I = reshape(I, size(I,1), Nmeasurements);
if Nmeasurements == 1
    StRaman = I;
elseif Nmeasurements == 2
    SpectrumA = I(:,1);
    SpectrumB = I(window_size:size(I,1),2);

    IntDiffAB = mean(SpectrumA(980:1000,1)) - mean(SpectrumB(1:20,1));

```

```
    StRaman = [SpectrumA; SpectrumB+IntDiffAB];
elseif Nmeasurements ==3
    SpectrumA = I(:,1);
    SpectrumB = I(window_size:size(I,1),2);
    SpectrumC = I(window_size:size(I,1),3);

    IntDiffAB = mean(SpectrumA(980:1000,1)) - mean(SpectrumB(1:20,1));
    SpectrumB = SpectrumB+IntDiffAB;
    IntDiffBC = mean(SpectrumB(638:658,1)) - mean(SpectrumC(1:20,1));

    StRaman = [SpectrumA; SpectrumB; SpectrumC+IntDiffBC];
end;

Spectrum = StRaman;
NSpectrum=(Spectrum-min(Spectrum))/(max(Spectrum)-min(Spectrum));

%% smooth the spectrum to minimize error in peak selection
Spectrum = whittakerSmoother(Spectrum, 1);
NSpectrum = whittakerSmoother(NSpectrum, 1);

%% Find peaks in Fingerprint region
PeakArea = 1:size(NSpectrum,1);
PeakArea = PeakArea';

[PeakVal, PeakLoc, PeakWidth, PeakPro] =
findpeaks(NSpectrum,PeakArea,'MINPEAKDISTANCE', 40, 'MAXPEAKWIDTH', 10);
PeakRatio = PeakVal*100;
j = 1;

%% Select peaks which we are intrested in
for i = 1: size(PeakLoc,1);
    if PeakRatio(i,1)>5
        PeakVal2(j,1) = PeakVal(i,1);
        PeakLoc2(j,1) = PeakLoc(i,1);
        j = j+1;
    end
end

PeakValORG = PeakVal;
PeakLocORG = PeakLoc;

clear PeakVal PeakLoc

if Nmeasurements ==1
    for i = 1:7
        PeakVal(i,1) = PeakVal2(i,1);
        PeakLoc(i,1) = PeakLoc2(i,1);
    end;
elseif Nmeasurements == 2 || Nmeasurements ==3
    for i = 1:10
        PeakVal(i,1) = PeakVal2(i,1);
        PeakLoc(i,1) = PeakLoc2(i,1);
    end;
end

%% Return selected peaks (pixel location and normalized peak value)
```

```
ArHgPeakVal = PeakVal;
ArHgPeakLoc = PeakLoc;
ArHgSpect = Spectrum;
ArHgSpectNorm = NSpectrum;
end

%% Smoothing function, whittaker smoothing algorithm is used
function z = whittakerSmoother(y,lambda)
    m = length(y);
    I = eye(m);
    D = diff(I);
    z = (I + lambda * D' * D)\y;
end
```


APPENDIX C

Fabrication Process Flow for PDMS microchannel device

C.1 Mask Design

Masks for the PDMS fabrication were designed using AutoCAD 2010 (Autodesk Inc., CA, US). The design files were converted to extension of ‘cif’ file using CleWin 5.0 Layout Editor (v5.0.12, WieWeb software, Hengelo, the Netherlands).

Figure C-1 An example mask design which is designed for the waveguide Raman experiment

C.2 Device fabrication process

A mold has to be fabricated prior to PDMS device fabrication. The molds were fabricated using a photolithography technique using a positive photoresist (AZ-40-XT, KAYAKU Advanced Materials Co., LTD, Westborough, MA, USA). The fabrication procedure is as follows:

1. Spin coating the positive photoresist onto a 4-inch silicon wafer
2. Baking the spin coated wafer on a hot plate at 127 °C for 7 min
3. Exposure to UV light (i-line, $\lambda \sim 365$ nm) with a quartz mask
4. Post exposure baking on a hot plate at 105 °C for 90 sec
5. Develop the wafer for 4 min
6. Performing silane treatment using chlorotrimethylsilane (CTMS) for 10 min

PDMS microfluidic devices were fabricated using the mold. The process flow is as follows:

C

1. Mixing PDMS elastomer and cross-linker (RTV 615, GE Silicones, Boston, MA, USA). Mixing ratio of the PDMS is elastomer(A) : cross-linker(B) = 7:1
2. Performing degassing in a desiccator by applying vacuum. The degassing is conducted to remove bubbles in the PDMS mixture
3. Pouring the PDMS mixture onto the mold and leaving it on a flat surface for 1 min
4. Curing the PDMS in a convection oven at 80 °C for 45 min
5. Cutting the PDMS microfluidic device from the mold and punching inlets and outlets

Scientific Output

Peer-reviewed Paper

- W. Lee, M. de Goede, H. Smedes, M. Dijkstra, S.M. García-Blanco, H.L. Offerhaus, “Ring Resonator enhanced Waveguide Raman Spectroscopy” (in preparation)
- L.G. Rikkert, P. Beekman, J. Caro, F.A.W. Coumans, A. Enciso Martinez, G. Jenster, S. Le Gac, W. Lee, T. G. van Leeuwen, G. B. Loozen, A. Nanou, R. Nieuwland, H. L. Offerhaus, C. Otto, D. M. Pegtel, M. C. Piontek, E. van der Pol, L. de Rond, W. H. Roos, R. B. M. Schasfoort, M. H. M. Wauben, H. Zuilhof, L.W.M.M. Terstappen, “Cancer-ID: towards identification of cancer by tumor-derived extracellular vesicles in blood”, *Frontiers in Oncology*, **2020**; *10*, 608
- W. Lee, P. Muñoz-Galindo, I. Hegeman, Y.S. Yong, M. Dijkstra, S.M. García-Blanco, H.L. Offerhaus, “Study on Multiple Waveguide Platforms for Waveguide Integrated Raman Spectroscopy”, *OSA continuum*, **2020**; *3*, 1322
- I. Hegeman, M. Dijkstra, F.B. Segerink, W. Lee, S.M. García-Blanco, “Development of Low-loss TiO₂ Waveguides”, *Optics Express*, **2020**; *28*
- W. Lee, A.T.M. Lenferink, C. Otto, H.L. Offerhaus, “Classifying Raman Spectra of Extracellular Vesicles based on Convolutional Neural Networks for Prostate Cancer Detection”, *Journal of Raman Spectroscopy*, **2020**; *51*, 293
- W. Lee, A. Nanou, L. Rikkert, F. A. W. Coumans, C. Otto, L. W. M. M. Terstappen, H. L. Offerhaus, "Label-Free Prostate Cancer Detection by Characterization of Extracellular Vesicles Using Raman Spectroscopy", *Analytical Chemistry*, **2018**; *90*, 11290
- T. Ha, W. Lee, S.J. Hong, “Process Integration of Ion Sensitive Field Effect Transistor Bio-Sensor Array Platform”, *Journal of Nanoscience and Nanotechnology*, **2017**; *17*, 8321

Conference contribution

- W. Lee, A.T.M. Lenferink, C. Otto, H.L. Offerhaus, “Convolutional Neural Networks for Raman Spectra Analysis”, *EKC 2019*, **Jul 2019**, Vienna, Austria
- W. Lee, A.T.M. Lenferink, C. Otto, H.L. Offerhaus, “Convolutional Neural Networks for Classification of Tumor Derived Extracellular Vesicles”, *ISEV 2019*, **Apr 2019**, Kyoto, Japan
- W. Lee, P. Muñoz-Galindo, Y.S. Yong, M. Dijkstra, E. Vissers, S.M. García-Blanco, H.L. Offerhaus “Raman Spectroscopy, Free Space and On-Chip”, *MESA+ day 2018*, **Sep 2018**, Enschede, the Netherlands
- W. Lee, H.L. Offerhaus, “Spectroscopic way to detect early stage cancer: Extracellular vesicles, Raman Spectroscopy and Machine Learning”, *MESA+ day 2018*, **Sep 2018**, Enschede, the Netherlands
- W. Lee, H.L. Offerhaus, “Classifying Raman Spectra of Extracellular Vesicles using a Convolutional Neural Network”, *ICORS 2018*, **Aug 2018**, Jeju, Korea
- W. Lee, H.L. Offerhaus, “Prostate Cancer Diagnosis using Raman Spectroscopy and Convolutional Neural Network”, *EKC 2018*, **Aug 2018**, Glasgow, UK
- W. Lee, A. Nanou, L. Rikkert, F.A.W. Coumans, C. Otto, L.W.M.M. Terstappen, H.L. Offerhaus, “Characterization of Extracellular Vesicles using Raman Spectroscopy for Label-free Cancer Detection”, *ISEV 2018*, **May 2018**, Barcelona, Spain
- W. Lee, A. Nanou, L. Rikkert, A. Enciso-Martinez, F.A.W. Coumans, C. Otto, A. Lenferink, L.W.M.M. Terstappen, H.L. Offerhaus, “Characterization of Extracellular Vesicles using Raman Spectroscopy for Label-free Prostate Cancer Detection”, *NLSEV kick-off meeting*, **Nov 2017**, Utrecht, the Netherlands
- W. Lee, A. Nanou, L.W.M.M. Terstappen, H.S. Rho, S. le Gac, H.L. Offerhaus, “Characterization of Extracellular Vesicles using Raman Spectroscopy”, *FOM Veldhoven*, **Jan 2017**, Veldhoven, the Netherlands
- A. Nanou, A. Enciso-Martinez, P. Beekman, W. Lee, H.S. Rho, S. le Gac, C.J.M. van Rijn, H.L. Offerhaus, C. Otto, L.W.M.M. Terstappen, “Cancer-ID: Isolation and characterization of tumor-derived extracellular vesicles”, *MIRA Day*, **Nov 2016**, Enschede, the Netherlands

Acknowledgements

Finally, I am at the end of a very long sailing, that is my Ph.D. program. I started my Ph.D. to be an independent researcher. However, it was also a big part of my life and allowed me to make a lot of beautiful memories. During this journey, I have more family members, found a lot of good friends and colleagues, and more opportunities to make my life better. Of course, I faced a lot of challenging moments. When I cleft through these difficulties, they, however, became nutrients to be stronger. All of my experiences allowed me to develop my personal life and professional career.

First of all, I would like to thank my supervisor, Herman L. Offerhaus, for giving me the great opportunity to be a member of Optical Sciences, University of Twente. It would not have been possible to start my journey in the Netherlands without his positive decision. After I was onboard OS, I was able to be grown in the field of optics and spectroscopy with his guidance. He also shared his wisdom with me for my personal life. All his support has made my professional and personal life fruitful.

I would to also thank OS. Within OS, I had a lot of great discussions to find the right solutions and got tremendous support from all the group members. Plus, I had great memories during the scientific retreat to Prague, the group excursion to Holten, and all the group events and activities. Thank you all who organize these events. Without support from the OS, it would not be possible to finish the Ph.D. program. Thanks, Sonia, Jose, Karen, Meindert, Frans, Jeroen, Lantian, Sandra, Carlijn, Raimond, Ward, Simen, Pablo, Ivo, Jinfeng and Michiel. And I also thank all the group members who are not listed here.

Special thanks go to Pablo and Yean-sheng. It would not be possible to build the waveguide Raman setup without your help. I learned a lot from Yensheng when I started building the setup. He transferred a lot of knowledge to me about proper optics, cleaning the optics, and many useful tricks. Sadly, Yean-Sheng left us too soon. He will be remembered forever.

During my Ph.D., I have worked on the Cancer-ID project. Without the Cancer-ID project, it would not be possible to write this thesis. I truly appreciate all the Cancer-ID project members. I also have very nice colleagues from the Cancer-ID project. Luckily, I started working on this project with other Ph.D. candidates; Linda, Leonie, Agustin, Pepijn, Afroditi, Melissa and Gyllion. I was pleased to learn many things from you guys and to help you too. We are not only good colleagues but also very good friends. I cannot forget all the

memories that we have made for the last five years. We also traveled together to Rotterdam, Barcelona, Jeju and Kyoto. We also had unforgettable team-building activities. Especially in Kyoto, when I had a hard time at the end of my Ph.D., Linda, Agustin and Pepijn carefully listened to my complaints and gave me a big warm hug. It made me going forward.

To bring my idea into the real world, I got numerous supports from our research technicians, Jeoron, Frans, Meindert and Aufrid. It would not be possible to do the all the experiments without your support. I really appreciate all the fruitful discussion and advice of you.

Dear my office mates Carlijn and Pablo. I also want to thank you. I still remembered the very first conversation which I made with Pablo. In the first conversation, you told me a lot about the shops in Enschede. It was very helpful to get used to the city. Carlijn, we have never worked directly together on the same topic. But, all the discussion and chat with you was very nice and sometimes gave me new ideas. When I needed some help from Dutch people for some phone conversation, you were willing to help me.

OS drinking team! We haven't drink that much, though. I am delighted to be a member of OS drinking team. Thanks, Michiel, Simen, Jinfeng and Chrysostomos. We made a lot of fun activities and memories together. I was my great pleasure to spend time with you guys.

I would like to express my appreciation to my family and Korean friends. 저를 낳아 주시고 늘 제 결정을 믿고 지지해 주신 부모님께 감사드립니다. 먼 길 떠나는 저를 믿고 진화와 함께 하는 것을 허락해 주신 어머니, 아버님께도 감사드립니다. 양가 부모님의 지지 와 지원 없이 이 자리까지 오는 건 불가능 했을 거라고 생각합니다.

처음에 네덜란드에 와서 막막할 때 큰 도움을 주신 사촌누나 부부, 김명옥님, 최봉준님께 감사의 인사를 드립니다. 처음 두 분께서 해 주신 따듯 한 말씀과, 네덜란드 전반에 대한 정보 덕분에 큰 무리없이 타향 살이에 적응하고, 지금까지 무탈하게 지낼 수 있었던 큰 밑거름이 되었습니다.

제 불평 불만을 같이 들어주고 응원해준 가장 친한 친구들 최보근, 신용훈, 김형호, 권영해 에게도 진심으로 감사의 인사를 전합니다. 비록 멀리 떨어져 있어 자주 만나진 못했지만, 친구들의 응원과 격려 덕분에 힘든 시기도 잘 넘어갈 수 있었습니다. 앞으로도 좋은 친구관계를 유지하고, 같이 성장해 나갈 수 있기를 바랍니다.

제가 학업을 이어 나갈 수 있게 아낌없는 지원을 해 주신 홍상진 교수님께 감사의 인사를 드립니다. 무척 대고 찾아갔던 학부 2 학년 학생이었던 저에게 비전을 제시해 주셨습니다. 교수님의 조언과 비전이 제가 더 큰 세상을 볼 수 있는 밑 거름이 되었습니다. 이 후 학부연구생

그리고 석사 과정을 거치는 동안 받은 교수님의 아낌없는 가르침이 박사 과정 중에 맞닥뜨린 어려움들을 헤쳐 나갈 수 있는 원동력이 되었습니다.

엔스헤데에서 만나고 많은 도움을 주신 분들께도 감사의 인사를 전합니다. 엔스헤데에서 가족 같이 따뜻하게 대해주신 노훈석 박사님, 양윤선 박사님 부부께 진심으로 감사드립니다. 두 분이 가족같이 대해주신 덕분에 엔스헤데의 삶에 빠르게 적응할 수 있었습니다. 더불어 우연히 같은 프로젝트에 참여하게 된 노훈석 박사님의 많은 조언 덕분에, 새로운 기술을 배우고 제 연구의 방향을 찾아 앞으로 나아 갈 수 있었습니다. 엔스헤데에서 많은 시간을 함께 하지는 못했지만 늘 따뜻하게 대해주신 김재복 박사님, 김은혜님 께도 감사의 인사를 전합니다. 덕분에 유트렉과 엔스헤데를 오가며 즐거운 기억을 많이 만들었습니다. 가까운 거리에서 저와 많은 시간 함께 보내주신 진정민 박사님, 김선근님, 김찬미님, 김규태님께도 깊은 감사의 인사를 전합니다.

일면식도 없던 저에게 많은 도움을 주신 김효욱 박사님, 송화명님께 큰 감사의 인사를 전합니다. 두 분의 도움과 응원 덕분에, Covid-19 이 창궐하는 이 어려운 시기에 원하는 회사에 취업을 할 수 있었습니다. 회사의 동료로 그리고 이웃으로 큰 힘이 되어 주는 이정태님, 최예진님, 최정원님, Marco Mueller 님께 감사의 인사를 전합니다.

Lastly, special thanks must go to my family, Jinhwa, Philip and Hannah. 진화, 지난 4 년 8 개월 동안 묵묵하게 나를 믿고 지지해줘서 고마워. 진화의 지지와 도움 없이는 이 자리까지 오는 일은 불가능 했을 꺼야. 바쁜 남편을 위해 굶은 집안일을 도맡아 하게 해서 늘 미안하고 감사한 마음이 커. 지난 4 년 8 개월보다 더 행복하게 살자! 고마워 사랑해! 나의 아들 필립. 아빠가 바빠서 많이 못 놀아줘서 늘 미안해. 아주 많은 시간을 함께 보내지는 못했지만, 걱정과는 달리 건강하게 씩씩하게 자라줘서 고마워. 앞으로는 아빠 랑 더 많은 시간을 함께 보내자. 사랑해! 마지막으로 힘들고 어려운 시기에 우리에게 찾아와 큰 기쁨과 행복이 되어준 딸 한나야 고마워. 앞으로 행복하게 즐겁게 지내자. 사랑해!

September 2020

Wooje Lee
From Eindhoven...