On-Chip Carbon Dioxide Sensor Based on Non-Dispersive Infrared Spectroscopy

Xiaoning Jia

Doctoral dissertation submitted to obtain the academic degree of
Doctor of Photonics Engineering

Supervisors
Prof. Roel Baets, PhD - Prof. Günther Roelkens, PhD

Department of Information Technology
Faculty of Engineering and Architecture, Ghent University

April 2021
Members of the Examination Board

Chair

Prof. Filip De Turck, PhD, Ghent University

Other members entitled to vote

Prof. Bart Kuyken, PhD, Ghent University
Prof. Jeroen Missinne, PhD, Ghent University
Prof. Delphine Morini, PhD, Université Paris-Saclay, France
Joris Roels, PhD, Melexis Technologies

Supervisors

Prof. Roel Baets, PhD, Ghent University
Prof. Günther Roelkens, PhD, Ghent University
Acknowledgements

Doing a PhD is a challenging but undeniably exciting and fruitful experience, and it’s been a quite unforgettable 4-year journey in my life. Looking back, I now feel so happy and fortunate that I had this chance to pursue a PhD degree in a world-renowned research group and a lovely city like Gent. This PhD wouldn’t have been possible without the help and support from many people: my colleagues, friends, and my family. First of all, I would like to give my deepest (and deepest) gratitude to my PhD promoters: Prof. Roel Baets, Prof. Gunther Roelkens, and my industrial advisor Dr. Joris Roels. Roel, thank you very much for taking me in when I knocked the door of your office with the ambition of doing a PhD in the Photonics Research Group. Your warm smile was the beginning of my wonderful 4-year PhD journey, and it has been filled with your kind guidance and numerous advice ever since. The countless discussions I had with you really inspired me, and I hope one day I can have the same mind as deep and broad as yours. Gunther, thank you for your help and patient guidance when I came to you with problems that I can’t solve, your ‘no problem, that’s what I’m here for’ really eased my anxiety. I deeply appreciate your recognition of my work and your kind reminder when I was about to deviate from the right path. Joris, thank you for providing precious thoughts and ideas from the industrial point of view, you have been a very good mentor as well as a friend. I would also like to thank Melexis and Vlaio for funding this PhD project, this PhD would not have been possible without your financial support.

Second, I would like to express my gratitude to the people who provided help and support with their expertise and precious time in the course of this PhD. Jasper, it was very nice of you that you didn’t even hesitate to help when I asked you to layout the PCBs and to make some cables. Steven and Muhammad, thank you both very much for answering my numerous processing and training requests, and for the countless useful discussions regarding clean room processing. Liesbet, thank you for helping me take the nice pictures with the SEM and FIB. Kristof Dhaenens, Sheila Dunphy and Stijn Duynslager, thank you for your help in the film lamination and wafer dicing. Joris Van Kerrebrouck, I sincerely appreciate your efforts in helping with the amplifier design and solving all kinds of electronics problems. Michael Vanslembrourck and Peter Guns, your help in setting up my measurement are greatly appreciated. Of course, many thanks go to the supporting
and administrative staff: Ilse Van Royen, Kristien De Meulder, Ilse Meersman, Mike Van Puyenbroeck, and Bert Coryn, my work and life in Gent went a lot smoother with your help.

I would also like to express my thanks to my colleagues in the Photonics Research Group, thank you for making my work and life at Gent joyful because of your help and delightful conversations. Anton, I learned a lot from you in our useful discussions when I just started my PhD. Jeroen, thank you for being a good friend with whom I could share my worries and joy, and I’m grateful that we could cheer each other up when we were in the valleys of our PhDs. My lovely office mates: Kristof, Stephane, Khannan, Ivo, Ana, Iman, thank you for making a pleasant working environment. Kristof, you’ve been a very good neighbor, I can still recall our common sighs and laughter when we were stressed.

2020 is a tough year due to the Corona outbreak, and we were forced to work from home and be isolated from each other. I would like to give my special thanks to my Chinese colleagues for your company in the year of 2020 and the years before it: Yuting, Xiaomin, Chonghuai, Xiangfeng, Zuyang, Yang, Jing, Haolan, Yufei, Yanlu, Chupao, Zhongtao, Chao, Mi and Hong. My four years of PhD would have been a quite lonely experience if it were not for you guys.

Finally, I would like to sincerely thank my parents, my brother, and my sister. As the youngest family member, I’ve grown up with your whole-hearted love and support. Although we do not speak out our love and feelings that much, I know that we all have enormous love for each other. To my lovely parents: I’ve worked hard towards a master and a PhD degree, I really appreciate your support and understanding in the past years, I hope I have made you very proud.

Gent, December 2020
Xiaoning Jia
# Table of Contents

Acknowledgements  
Nederlandse samenvatting  
English summary  

1 Introduction  
1.1 Background and motivation  
1.1.1 Mobile air conditioning and refrigeration systems  
1.1.2 Demand controlled ventilation  
1.1.3 Medical diagnosis  
1.2 Overview of gas sensing technology  
1.2.1 Non-optical gas sensors  
1.2.2 Optical sensors  
1.2.2.1 Indirect methods  
1.2.2.2 Direct methods  
1.3 State-of-the-art NDIR CO₂ sensors on the market  
1.4 Thesis objective and outline  
1.5 List of publications  
1.5.1 Patent  
1.5.2 Publications in international journals  
1.5.3 Publications in international conferences  
References  

2 On-Chip Optical Gas Sensing  
2.1 Optical Absorption Spectroscopy  
2.1.1 Absorption spectra and the mid-infrared  
2.1.2 Absorption line shape and line broadening  
2.1.3 Beer-Lambert law  
2.2 Mid-infrared Silicon Photonics  
2.3 Free-space and guided-wave optical sensing  
2.4 Conclusion  
References  

5 NDIR CO₂ Sensor Configurations 3-1
   3.1 Introduction ........................................ 3-1
   3.2 Sensor system ....................................... 3-2
   3.3 Multi-slot waveguide .................................. 3-3
      3.3.1 Waveguide design ................................ 3-4
      3.3.2 Optical source and detector ...................... 3-7
      3.3.2.1 Optical source ................................ 3-7
      3.3.2.2 Optical detector ................................ 3-9
      3.3.2.3 Transition loss ................................ 3-10
   3.4 Metallic hollow waveguide ............................. 3-10
      3.4.1 Waveguide design ................................ 3-11
      3.4.2 Optical source and detectors ..................... 3-13
   3.5 Metallic integrating cylinder ......................... 3-13
      3.5.1 Integrating cylinder analysis and simulation ... 3-14
      3.5.2 Optical source and detectors ..................... 3-15
   3.6 Comparison ........................................... 3-15
      3.6.1 Power budget considerations ..................... 3-15
      3.6.2 Fabrication complexity and footprint ............. 3-18
   3.7 Conclusion ........................................... 3-18
References .................................................. 3-20

6 On-Chip CO₂ Sensing Based on Integrating Cylinder 4-1
   4.1 Introduction ........................................... 4-1
   4.2 Theoretical analysis ................................... 4-3
   4.3 Simulations ............................................ 4-6
      4.3.1 Total loss ........................................ 4-7
      4.3.2 Equivalent path length ............................ 4-7
      4.3.3 Design optimization ................................ 4-8
      4.3.4 Sidewall roughness and source emission pattern ... 4-10
      4.3.5 Response time simulation .......................... 4-14
   4.4 Device fabrication ...................................... 4-15
      4.4.1 Deep reactive ion etching and sidewall scattering 4-15
      4.4.2 Gold deposition .................................... 4-16
      4.4.3 Wafer bonding ...................................... 4-17
   4.5 Experimental settings .................................. 4-18
      4.5.1 Experimental setup ................................ 4-19
      4.5.2 Source and optical filters ......................... 4-20
      4.5.3 Photodiodes and transimpedance amplifiers ....... 4-21
      4.5.4 Gas generation ...................................... 4-22
   4.6 Proof-of-concept experimental results .................. 4-23
      4.6.1 Allan deviation measurement ....................... 4-23
      4.6.2 CO₂ sensing measurement ........................... 4-26
      4.6.3 Response Time ...................................... 4-27
   4.7 Conclusions ............................................ 4-28
References .................................................. 4-30
## 5 Fully Integrated NDIR CO₂ Sensor

5.1 Introduction ................................................. 5-1
5.2 LED and photodiode optopairs ............................... 5-2
  5.2.1 Mid-IR III-V LEDs ..................................... 5-2
      5.2.1.1 LED used in this thesis work ....................... 5-3
  5.2.2 Mid-IR photodiode ..................................... 5-3
      5.2.2.1 Performance metrics .............................. 5-3
      5.2.2.2 Photodiode equivalent circuit ................. 5-5
      5.2.2.3 Noise characteristics ........................... 5-6
      5.2.2.4 Photodiode operation mode .................... 5-7
      5.2.2.5 Photodiodes used in this thesis work ........ 5-8
  5.3 Estimation of coupling efficiencies ...................... 5-11
  5.4 Full sensor fabrication ................................ 5-12
  5.5 Trans-impedance amplifier design ....................... 5-14
  5.6 Experimental results ................................... 5-16
      5.6.1 Allan deviation .................................. 5-16
      5.6.2 CO₂ sensing measurement ......................... 5-19
      5.6.3 Response time .................................... 5-22
      5.6.4 Water interference ................................ 5-23
      5.6.5 Long term stability .............................. 5-25
  5.7 Conclusions .............................................. 5-27
References .................................................. 5-28

## 6 Conclusions and Perspectives .......................... 6-1

6.1 Conclusions .............................................. 6-1
6.2 Perspectives ............................................. 6-3
      6.2.1 Improving the signal to noise ratio ............. 6-3
      6.2.2 Temperature compensation ......................... 6-4
      6.2.3 The shape of the integrating cavity ............ 6-4
      6.2.4 Gas membrane ..................................... 6-5
      6.2.5 LED driver, TIA, and lock-in amplifier .......... 6-5
1. (a) Specifieke meetopstelling voor het karakteriseren van de integreerde cilindrische gassensor. (b) De transmissiespectra van de optische filters in het detectiekanaal (rood) en het referentiekanaal (blauw). (c) Emissiespectra van de lichtbron in het golflengtegebied tussen 3 en 5 µm. ........................................ xxiv

2. (a) Signaalrespons van zowel het detectiekanaal (rood) als het referentiekanaal (blauw) voor verschillende opgelegde CO₂ concentraties. (b) De genormaliseerde signaalrespons, verkregen door de normalisatie van het detectiekanaalsignaal ten opzichte van het referentiekanaalsignaal. Een legende van de verschillende opgelegde CO₂ concentraties is weergegeven in de tabel aan de rechterzijde. xxv

3. Een 3D schematische voorstelling van de volledig geïntegreerde NDIR CO₂ sensor. De sensor bestaat uit twee verschillende substraten die met elkaar verbonden worden. In het bovenste substraat wordt de geïntegreerde cilindrische caviteit gevormd. Op het onderste substraat zijn de lichtbron en fotodiodes aangebracht. xxvi

4. (a) Signaalrespons van het detectiekanaal (rood) en het referentiekanaal (blauw) van de volledig geïntegreerde sensor aan verschillende CO₂ gasconcentraties. (b) De genormaliseerde signaalrespons, verkregen door de normalisatie van het detectiekanaalsignaal ten opzichte van het referentiekanaalsignaal. Een legende van de verschillende opgelegde CO₂ concentraties is weergegeven in de tabel aan de rechterzijde. xxvii

1. (a) Dedicated gas sensing setup to characterize the integrating cylinder. (b) Transmission spectra of the sensing and reference filters. (c) Emission spectrum of the light source in the 3µm - 5µm wavelength range. ........................................ xxxi

2. (a) Response of both sensing signal and reference signal to CO₂ concentration steps. (b) Normalized transmission, obtained by dividing the sensing signal with the reference signal. The CO₂ concentration steps are listed on the right side. xxxii
3D schematic of the fully integrated NDIR CO\textsubscript{2} sensor. The sensor consists of two silicon substrates. On the top substrate, the integrating cylinder is defined. On the bottom substrate, the LED and photodiodes are integrated on-chip. 

(a) Response of both the sensing signal and reference signal to CO\textsubscript{2} concentration steps. (b) Normalized transmission, obtained by dividing the sensing signal by the reference signal. The CO\textsubscript{2} concentration steps are listed on the right side.

CO\textsubscript{2} concentration at the driver’s place in the vehicle with four passengers. With a sudden leak, all the CO\textsubscript{2} is released in 60 seconds, the C Feet, C head and C mean denote the CO\textsubscript{2} concentration at the driver’s feet, head and the average value, respectively. (a) With no ventilation and no re-circulation. (b) With maximum ventilation and re-circulation. Adapted from [37].

Simulation of the indoor CO\textsubscript{2} concentration in a bedroom (V = 21 m\textsuperscript{3}) with one person during an 8-hour period. The initial CO\textsubscript{2} concentration is set to be 400ppm. When there is no ventilation, the CO\textsubscript{2} concentration increases linearly as a function of time. Increased ventilation rate leads to reduced steady-state CO\textsubscript{2} concentration and the steady state is reached faster. Adapted from [39].

Comparison of CO\textsubscript{2} concentration waveforms of a normal individual and a patient with asthma. The waveform consists of four phases: (1) expiratory upstroke (segment PQ); (2) Alveolar plateau (segment QR); (3) Inspiratory downstroke (segment RS); (4) respiratory baseline (segment ST). Adapted from [46].

Evolution of NDIR gas sensors over the past three decades, and the vision for a fully integrated on-chip NDIR gas sensor in the future. Adapted from [90].

The electromagnetic spectrum and the mid-IR ‘finger print’ region, the absorption spectra of trace gases are for 100% concentration at atmospheric pressure. Adapted from [1].

(a) Symmetric stretching and (b) antisymmetric stretching of a CO\textsubscript{2} molecule. In symmetric stretching, the two C=O bonds stretch in opposite directions while in antisymmetric stretching one of the C=O bond stretches while the other compresses.

Absorption spectrum of 100\% CO\textsubscript{2} gas at atmospheric pressure, calculated from the Hitran database [6].

Optical transparency window of several common materials used in mid-IR silicon photonics. In the yellow regions the material loss is negligible while the red regions denote heavy material absorption. Adapted from [20].

Reported suspended waveguide structures for the mid-IR wavelength range. Images taken from [25, 28, 30].
3.1 Schematic of the NDIR sensor system. (a) Single-beam configuration. (b) Dual-beam configuration. 3-2

3.2 Schematic of a multi-slot silicon waveguide, the buried oxide (BOX) layer is partly under-etched to reduce the absorption loss of the optical mode, periodic SiO$_2$ pillars are formed for mechanical support. 3-4

3.3 (a) Cross-section of the multi-slot waveguide. (b) Horizontal components of the electric fields of the fundamental TE and TM modes in the middle of the waveguide. (c) 2D field profiles of the first three TE and TM modes. 3-5

3.4 Two types of optical sources for the multi-slot waveguide. (a) Mid-IR III-V LED heterogeneously integrated on the silicon waveguide. (b) Thermal source realized by heating the (doped) waveguide to a target temperature. 3-7

3.5 (a) Bandgap energy and lattice constants of the III-V semiconductor alloys. (b) Layer stack of the transfer-printable mid-IR LED proposed in this work, extrapolated from [9]. 3-8

3.6 Schematic of the metallic hollow waveguide, the waveguide is formed by bonding two metalized silicon substrates. The waveguide is etched on one substrate and the other one is just a planar substrate. 3-10

3.7 Metallic hollow core waveguide simulation. (a) Propagation loss of the hollow waveguide with various widths, the height of the waveguide is 300 $\mu$m, an isotropic point source (at 4.26 $\mu$m) is used and gold coatings are applied on the sidewalls as reflective mirrors. (b) Propagation loss for various cone angles of the source for a 300 $\mu$m $\times$ 300 $\mu$m square waveguide, a cone angle of 90$^\circ$ corresponds to half a sphere. 3-12

3.8 Reflectance of gold at 4.26 $\mu$m against angle of incidence. 3-12

3.9 Hollow core waveguide with an LED (or a photodiode) integrated on-chip. The LED and photodiode can be integrated on-chip using the same processes. 3-14

3.10 Schematic of the integrating cylinder. 3-15

4.1 (a) 3D schematic of the integrating cylinder with access waveguides. The integrating cylinder is formed by water bonding of two silicon substrates. (b) 2D analysis of the integrating cylinder, with $R$ being the radius of the circle and $d$ being the width of all access waveguides. $P$ is the input power, $P_{WG}$, $P_{rad}$ and $P_{prop}$ are the power coupled to the output waveguides, the power absorbed/radiated at the cavity sidewall, and the power lost during propagation in the cavity, respectively. 4-3
4.2 (a) 2D schematic of the simulation setup. An isotropic source is used to launch the rays, and two detectors are placed at the input/output waveguides to calculate the transmission. (b) Total transmission loss from the input waveguide to the output waveguide for various cylinder radius and waveguide width combinations.

4.3 Equivalent path length from the input waveguide to the output waveguide for various cylinder radius and waveguide width combinations.

4.4 Sensitivity (in absorbance unit, or AU) due to the absorption of 1000 ppm CO$_2$, for various cylinder radius and access waveguide width combinations.

4.5 Simulation setup to study the angular distributions of the reflected light over (a) a roughened surface; (b) a lambertian surface.

4.6 (a)-(c) Angular distributions of the directions of the reflected light over a rough surface with various roughness amplitudes. (d) Angular distribution of the directions of the reflected over a lambertian surface.

4.7 Three types of integrating cylinders. (a) A specular cylinder. (b) A roughened cylinder. (c) A lambertian cylinder.

4.8 Sensitivity of the integrating cylinder for 1000 ppm CO$_2$. (a) A specular cylinder. (b) A lambertian cylinder. (c) A roughened cylinder with roughness amplitude of $r_m = 5 \mu m$. (d) A roughened cylinder with roughness amplitude of $r_m = 10 \mu m$.

4.9 (a) CO$_2$ concentration distribution inside the integrating cylinder at $t = 2.5s$. (b) Time-dependent CO$_2$ concentration at the right edge of the integrating cylinder.

4.10 Fabrication process flow of the on-chip integrating cylinder.

4.11 (a) Silicon chip after DRIE, the chips consists of two cylinders. (b) Microscope image of the roughness on the sidewalls. (c) Amplitude distribution of the sidewall roughness.

4.12 (a) Propagation loss of the hollow waveguide at 4.26\mu m as a function of the gold coating thickness on the sidewalls. (b) Thickness of gold at different locations of the hollow waveguide in the gold sputtering test.

4.13 (a) Bonding recipe. (b) Fabricated chip with a EURO 10 cent coin, the chip consists of two integrating cylinders. (c) Scanning electron microscope (SEM) image of the gold-gold interface after bonding.


4.15 (a) Emission spectrum of the source in the 3 - 5\mu m wavelength range, the whole emission spectrum is shown in the inset. (b) Transmission spectra of the active and reference filters, superimposed on the CO$_2$ absorption spectrum in 3-5\mu m wavelength range.
4.16 (a) Circuit schematic of the designed TIA, the soldered PCB with the TIA is shown in the inset. (b) The simulated transient response of the TIA. (c) The simulated AC response of the TIA.

4.17 (a) Baseline of the sensing $I_{A0}$ and reference signals $I_{R0}$, measured over a period of ~3 hours, with 100ms lock-in integration time. (b) Normalized signal $I_{A0}/I_{R0}$.

4.18 Allan deviation of the sensing and reference signals, as well as the normalized signal (in pure N$_2$ environment).

4.19 (a) Response of both sensing signal and reference signal to CO$_2$ concentration steps. (b) Normalized transmission, obtained by dividing the sensing signal with the reference signal. The CO$_2$ concentration steps are listed on the right side.

4.20 Normalized absorbance of the integrating cylinder at different CO$_2$ concentrations. The response at CO$_2$ = 100 ppm is in the inset, with 10 measurements superimposed in one figure.

4.21 Response time measurement of the integrating cylinder, the integration time of the lock-in amplifier is 100ms. The gas flow mechanism when the 50% CO$_2$ is abruptly shut off is shown in the inset.

5.1 (a) Picture of the LED bare chips, the LED has a top ring contact and a planar bottom contact. (b) Layer stack of the LED. Pictures from [17].

5.2 Equivalent circuit of a photodiode [21]. $I_{ph}$ is the photocurrent, $I_d$ is the dark current, $C_j$ is the junction capacitance, $R_{sh}$ is the shunt resistance and $R_s$ is the series resistance.

5.3 (a) Photodiode working at photoconductive mode. (b) Photodiode working at photovoltaic mode. Figures taken from [22].

5.4 Microscope images of the sensing photodiode PD42 and the reference photodiode PD36.

5.5 Normalized emission spectrum of the LED [17] and spectral responses of the photodiodes [18, 23], with the absorption spectrum of CO$_2$ in the 3µm - 5µm wavelength range [24].

5.6 Simulation setup for the coupling efficiencies. (a) From LED to the input waveguide. (b) From output waveguide to the photodiode.

5.7 3D schematic of the fully integrated NDIR CO$_2$ sensor. The sensor consists of two silicon substrates. On the top substrate, the integrating cylinder is defined. On the bottom substrate, the LED and photodiodes are integrated on-chip. PD42: photodiode at the sensing arm, with spectral response aligned with the CO$_2$ absorption band. PD36: photodiode at the reference arm, with spectral response cutoff at 4µm.
5.8 Fabrication process flow of the bottom substrate containing the LED/PD chips. (a) Blank silicon substrate. (b) Three trenches are formed on the substrate via selective etching of silicon by KOH. (c) Gold deposition and lift-off. In this step, two contact pads (anode and cathode) are formed for the LED/PD, the deposited gold on the bottom substrate acts as a reflector for the integrating cylinder. (d) LED/PD bonding and wire bonding. The LED/PD chips are bonded in the trench using flip-chip bonding, after which the top contacts of the LED/PD are wire-bonded to the anode pads.

5.9 (a) Patterned photoresist film (in purple) at the input side, the resist film is hanging over the edge of the trench. (b) Contact pads after gold deposition and lift-off. (c) Flip-chip bonded LED in the trench, with a wire bond connecting the top contact of the LED and the contact pad on the substrate.

5.10 (a) Circuit schematic of the TIA for the sensing photodiode (PD42). The TIA for the reference photodiode (PD36) only differs in the feedback resistor and capacitor. (b) Transient response of both TIAs. (c) AC response of both TIAs.

5.11 PCB consisting of the TIAs for the sensing photodiode and the reference photodiode. The fabricated sensor chip is mounted on the PCB and the contacts pads of the LED/PDs are connected to the board with wire bonds.

5.12 The sensing signal and the reference signal measured over a period of 3 hours, the measurement interval is 100 ms.

5.13 (a) Allan deviations of the sensing signal, the reference signal and the normalized signal.

5.14 (a) Response of both the sensing signal and reference signal to CO₂ concentration steps, the signals are measured with an integration time of 30s. (b) Normalized transmission, obtained by dividing the sensing signal with the reference signal. The CO₂ concentration steps are listed on the right side.

5.15 Normalized absorbance of the full sensor at different CO₂ concentrations.

5.16 Response time measurement of the fully integrated sensor.

5.17 Water vapor and CO₂ absorption spectra in 3.0μm - 5.5μm wavelength range, together with the emission/responsivity spectra of the LED/PDs. The amplitude of the water vapor absorption spectrum is magnified by a factor of 100.

5.18 (a) Water interference measurement. During the measurement, the RH level is swept from 0 to 99% while the CO₂ concentration is kept at 2500ppm. (b) Simulation of the water vapor absorption at various RH levels, the absorption of water is added on top of the absorption of 2500ppm CO₂.
5.19 (a) Long-term stability measurement of the sensor for a period of
∼ five days. (b) Ambient temperature profile measured with a
thermopile detector. ................................. 5-26

5.20 Temperature-dependence of the LED/PD characteristics. (a) Emission spectrum of the LED. (b) Spectral response of the sensing
photodiode. (c) Spectral response of the reference photodiode.
Figures taken from [17], [23], and [18], respectively. ........... 5-27

6.1 Gas membranes used in commercial gas sensors. The two gas sen-
sors are CozIR-LP3 CO\(_2\) sensor from Gas Sensing Solutions, and
PAC3500 CO sensor from Draeger, respectively. ................. 6-5
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Overview of state-of-the-art commercial NDIR CO₂ sensors from several major manufacturers.</td>
<td>1-17</td>
</tr>
<tr>
<td>3.1</td>
<td>Optical constants and calculated reflectance of gold, silver, aluminum, and copper at 4.26 μm.</td>
<td>3-13</td>
</tr>
<tr>
<td>3.2</td>
<td>Power budget breakdown for the three sensor configurations.</td>
<td>3-17</td>
</tr>
<tr>
<td>4.1</td>
<td>Optimal cylinder radius and access waveguide width combinations, and the corresponding equivalent path length and total transmission loss.</td>
<td>4-10</td>
</tr>
<tr>
<td>4.2</td>
<td>Specifications of the photodiodes used in the experimental setup.</td>
<td>4-21</td>
</tr>
<tr>
<td>5.1</td>
<td>Major specifications of the LED used in this work.</td>
<td>5-3</td>
</tr>
<tr>
<td>5.2</td>
<td>Noise contributions and the total noise current when the photodiode works at photoconductive and photovoltaic modes.</td>
<td>5-8</td>
</tr>
<tr>
<td>5.3</td>
<td>Specifications of the photodiodes used in this thesis. PD42BS: photodiode at the sensing channel. PD36-03: photodiode at the reference channel.</td>
<td>5-10</td>
</tr>
</tbody>
</table>
# List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFM</td>
<td>Atomic Force Microscope</td>
</tr>
<tr>
<td>ASIC</td>
<td>Application-Specific Integrated Circuit</td>
</tr>
<tr>
<td>BOX</td>
<td>Buried Oxide</td>
</tr>
<tr>
<td>CO2</td>
<td>Carbon Dioxide</td>
</tr>
<tr>
<td>CMOS</td>
<td>Complementary Metal Oxide Semiconductor</td>
</tr>
<tr>
<td>DRIE</td>
<td>Deep Reactive Ion Etching</td>
</tr>
<tr>
<td>DCV</td>
<td>Demand Controlled Ventilation</td>
</tr>
<tr>
<td>EIC</td>
<td>Electrical integrated Circuit</td>
</tr>
<tr>
<td>EMI</td>
<td>Electromagnetic Interference</td>
</tr>
</tbody>
</table>
FWHM  Full Width at Half Minimum

GaSb  Gallium Antimonide
GWP  Global Warming Potential
GC  Gas Chromatography

IoT  Internet of Things
IC  Integrating Cylinder
IR  Infra Red

LOD  Limit of Detection
LED  Light Emitting Diode

MAC  Mobile Air Conditioning
MEMS  Micro-Electro-Mechanical System
MIR  Mid-Infrared
MOS  Metal Oxide Semiconductors
MZI  Mach–Zehnder Interferometer
NDIR Non-Dispersive Infrared
NEP Noise Equivalent Power
NIR Near Infra Red

P

PAS Photoacoustic Spectroscopy
PCB Printed Circuit Board
PD Photodiode
PIC Photonic Integrated Circuit
PMMA Poly MethylMethAcrylate
PTS Photothermal Spectroscopy

Q

QCW Quasi-Continuous Wave

R

RH Relative Humidity
RI Refractive Index
RMS Root Mean Square

S

SEM Scanning Electron Microscope
SNR Signal to Noise Ratio
SOI Silicon on Insolator

T

TDLAS Tunable Diode Laser Absorption Spectroscopy
TE Transverse-electric
TIA Trans-Impedence Amplifier
TM Transverse-magnetic
De marktvraag naar een kleine en goedkope CO$_2$-gassensor is snel groeiend. Een dergelijke CO$_2$-sensor is van groot belang voor veel toepassingen zoals het monitoren van de luchtkwaliteit binnen gebouwen, het monitoren tuinbouwserres waarin gewassen gekweekt worden en ook het monitoren van allerlei fabricageprocessen in de algemene industrie. In het bijzonder heeft de Europese Unie een gradueel verbod geïntroduceerd op fluorgassen in mobiele airconditioningssystemen (MAC) omwille van het milieu en brandveiligheid. Het is daardoor duidelijk dat CO$_2$ een geschikte kandidaat is om de fluorgassen te vervangen als koelgas in de MAC koelsystemen. Echter, het risico op een CO$_2$ lek bestaat, wat zich vertaalt naar een risico voor de personen in de wagen, en bijgevolg ook voor de andere voertuigen in het verkeer. Vandaar de nood aan een heel responsieve CO$_2$ sensor die een mogelijk lek kan detecteren en bijgevolg de inzittenden kan waarschuwen voor dit lek. De afzetmarkt voor snelle en goedkope CO$_2$ sensoren in mobiele MAC systemen is dan ook enorm groot.

Verschillende technologieën zijn vandaag beschikbaar op de markt als een CO$_2$ sensor, gaande van elektrochemische detectie tot katalytische detectie en spectroscopische methodes. Elektrochemische en katalytische methodes kunnen voordelig zijn in termen van kost, maar zijn nadelig op vlak van stabiliteit op lange termijn, lage duurzaamheid en kruisreactiviteit met andere gassen. Daarentegen bieden spectroscopische technieken een betere stabiliteit op lange termijn, zijn ze ook nauwkeuriger en kunnen ze gemakkelijk een onderscheid maken tussen verschillende gassen. Spectroscopische gasdetectie is gebaseerd op de ‘vingerafdruk’ analyse van de absorptielijnen van het specifiek gas in het infrarode golflengtebereik (op ongeveer 4.26µm voor CO$_2$). Tussen de verschillende spectroscopische methodes onderscheidt niet-dispersieve infrarood (NDIR) spectroscopie zich als de beste kandidaat voor CO$_2$ detectie op basis van de relatief eenvoudige implementatie, de compactheid van de sensor en het veel lagere kostplaatsje van de gehele sensor. Daarbovenop valt een lage kruissensitiviteit op te merken omwille van de zeer hoge CO$_2$ absorptiecoëfficiënt bij 4.26µm ten opzichte van de laag tot zeer lage absorptie coëfficiënten van andere gassen bij dezelfde golflengte. Omwille van de bovenstaande redenen wordt de markt van CO$_2$ sensoren voornamelijk bezet door de NDIR-sensoren. Echter, de huidige generatie NDIR-sensoren zijn groot, met een optische interactielengte van enkele centimeters om tot een detectielimiet te komen van 100 moleculen per miljoen. Bovendien zijn ze ook duur.
omdat ze geassembled zijn met discrete optische elementen, wat wederom bijdraagt tot hun omvang. Dit alles voorkomt het gebruik van de huidige generatie NDIR CO\textsubscript{2} sensoren in prijsgroeiende markten zoals de automotive markt.

De doelstelling van dit doctoraatsonderzoek is het ontwikkelen van een drastisch kleinere en goedkopere NDIR CO\textsubscript{2} sensor die gefabriceerd kan worden op grote schaal, op basis van wafer-processen zoals in de elektronica gebeurt. Er zijn drie paden onderzocht om een dergelijke grote sensor te verkleinen tot een grootte die past op een silicium chip, namelijk: de ‘multi-slot’ veelvoudige-gleufgolfgeleider, de holle metallische golfgeleider en de holle geïntegreerde cilinder. De holle geïntegreerde cilinder is uiteindelijk gekozen als de beste kandidaat op basis van een lager stroomverbruik, eenvoudigere fabricagetechnieken en lagere productiekosten ten opzichte van de andere kandidaten. De geïntegreerde cilinder kan beschouwd worden als het 2D equivalent van een 3D integrerende sfeer die op een chip past. Zoals bij een integrerende sfeer, kan het licht veelvoudig reflecteren in de cilinder, wat ertoe leidt dat het licht een lange interactielengte kan hebben met het gas. In dit doctoraatsonderzoek is de integrerende cilinder zowel numeriek als experimenteel onderzocht. De eerste CO\textsubscript{2} gasmetingen zijn uitgevoerd met externe optische lichtbronnen en detectoren. Naderhand zijn de lichtbron en detectoren ook geïntegreerd op dezelfde silicium chip als de sensor. In dit werk worden de volgende eigenschappen bestudeerd: de sensorreactie naar CO\textsubscript{2}, de reactietijd van de sensor, de kruis-gevoeligheid van de sensor naar water en de lange-termijnstabiliteit van de sensor. Enkele praktische aspecten zoals de compensatie voor signaaldrift door temperatuurschommelingen en bescherming door een gasmembraan worden ook kort besproken.

**Niet-dispersieve infraroodspectroscopie**

Een typische NDIR optische sensor bestaat uit twee kanalen: een detectiekanaal en een referentiekanaal. Het uitgezonden licht (met emissie rond 4.26\(\mu\)m) wordt spectraal gefilterd en belicht vervolgens twee fotodetectoren. Voor het detectiekanaal overlap het gefilterde lichtspectrum met de absorptieband van CO\textsubscript{2}, zodat de lichtintensiteit afneemt bij toenemende CO\textsubscript{2}-concentratie. De hoeveelheid absorptie van het licht wordt goed beschreven door de wet van Beer-Lambert:

\[
I = I_0 \exp(-\alpha_{\text{gas}} c_{\text{gas}} L) \quad (1)
\]

Hier is \(\alpha_{\text{gas}}\) de absorptiecoëfficiënt van het te onderzoeken gas per ppm gas, hier CO\textsubscript{2}. \(C_{\text{gas}}\) is de gasconcentratie in ppm en \(L\) is de lengte van het optische pad in het detectiekanaal. Bij het referentiekanaal verandert de lichtintensiteit niet wanneer de concentratie van het te onderzoeken gas verandert. Dit kan worden bereikt door een andere golflengte te gebruiken, waarbij geen van de gewoonlijke gassen absorberen (in het geval van CO\textsubscript{2} sensors is dit gewoonlijk 3.9\(\mu\)m), of door het referentiekanaal te isoleren van de omgeving. Het gebruik van een referentieka-
naal is voornamelijk bedoeld om invloeden zoals fluctuaties in de intensiteit van de lichtbron en andere gedeelde ruis te elimineren.

**On-Chip CO\textsubscript{2}-detectie op basis van een integrerende cilinder**

De integrerende cilinder bestaat uit een cilindrische caviteit met een metallisch laagje op de wanden. De caviteit wordt afgesloten door het samenbinden van twee silicium substraten onder hoge druk. Om de optimale configuratie te vinden van de integrerende cilinder wordt er gebruik gemaakt van simulatiesoftware die lichtstralen met reflecties kan volgen in een 3D ruimte. De simulatieresultaten tonen aan dat voor een gegeven cilinder met een bepaalde straal (R), hoogte (h) en breedte van de toegangsgolfgeleiders (d) kunnen wordt geoptimaliseerd om de gevoeligheid van de integrerende cilinder te maximaliseren. Voor de experimentele uitvoering kiezen we bijgevolg voor een cilinder met straal \( R = 2.0 \text{mm} \), hoogte \( h = 300 \mu\text{m} \) en breedte van de toegangsgolfgeleider \( d = 200 \mu\text{m} \). Een dergelijke geïntegreerde cilinder levert een equivalente optische padlengte op van 3.5 cm, en een netto vermogensverlies van -7.4 dB. De fabricage van de sensor maakt gebruik van een diepe reactieve-ionen-etsstap (DRIE) om het silicium te etsen, de depositie van goud op de wanden van de caviteit en het aan elkaar binden van tweesubstraten om de caviteit te sluiten. Dit zijn allemaal processtappen die compatibel zijn met het proces van elektronische / MEMS wafers. Daardoor is het mogelijk om de sensor te fabriceren op grote schaal, en bijgevolg met een lage kost.

Om de CO\textsubscript{2} gassensor experimenteel te testen is er een specifieke meetopstelling gebouwd zoals voorgesteld in Figuur 1(a). Het infrarode licht dat wordt uitgestraald door een breedbandige thermische bron (waarvan het emissiespectrum getoond wordt in Figuur 1(c)) koppelt in aan de toegangspoort van de golfgeleider via een optische vezel. Het licht wordt gemoduleerd door een optomechanische modulator die periodiek de doorgang van licht tot de golfgeleider blokkeert. De modulator is geplaatst tussen de optische vezel en de toegangspoort van de golfgeleider. De twee golfgeleiders die als uitgangspoorten dienen van de integrerende cilinder worden gebruikt als het detectiekanaal en het referentiekanaal. In het detectiekanaal wordt het licht gedetecteerd door een fotodiode op kamertemperatuur. De opgewekte fotostroom wordt eerst versterkt door een transimpedantieversterker (TIA) en vervolgens door een lock-in versterker. Het enige verschil tussen het referentiekanaal en het detectiekanaal is de gekozen golflengte: 4.26\(\mu\text{m} \) en 3.9\(\mu\text{m} \). De selectie van de gewenste golflengte (spectrale filtering) in elk kanaal wordt bereikt door twee optische banddoorlaatfilters. Voorts zijn de twee kanalen identiek.

Eerst wordt een Allan-deviatie meting uitgevoerd voor het bepalen van het niveau van de systeemruis en het vinden van de optimale integratietijd van de lock-in versterkers. De detectielimiet van de sensor is bij benadering 2x10\textsuperscript{-4} AU (absorbance units), over een integratietijd van 1 seconde. De respons van de sensor op
Figuur 1: (a) Specifieke meetopstelling voor het karakteriseren van de integrerende cilindrische gassensor. (b) De transmissiespectra van de optische filters in het detectiekanaal (rood) en het referentiekanaal (blauw). (c) Emissiespectra van de lichtbron in het golflengtegebied tussen 3 en 5 µm.

Variërende CO₂ concentraties wordt getest door de geïntegreerde sensor te kop-\(\text{pelen aan een externe gastoevoer. Met deze externe gastoevoer kan op een gecontro-}\)\(\text{leerde manier CO₂ toegevoerd worden waarvan de concentratie extern regelbaar is. Figuur 2 (a) toont zowel het detectiesignaal als het referentiesignaal wanneer de CO₂-concentratie wordt gevarieerd.}

Het is duidelijk dat wanneer de CO₂-concentratie toeneemt, het detectiesignaal afneemt door absorptie van het licht, terwijl het referentiesignaal relatief stabiel blijft. Om de gemeinschappelijke ruis op beide kanalen te elimineren, wordt het detectiesignaal genormaliseerd ten opzichte van het referentiesignaal en wordt het genormaliseerde signaal getoond in Figuur 2 (b). Men kan zien dat na de normalisatie de gemeenschappelijke fluctuaties op beide signalen worden verwijderd, waardoor het genormaliseerde signaal van hogere kwaliteit is. De detectielimiet van de sensor is ongeveer 100 ppm, wat consistent is met de voorspellingen van de Allan-afwijkingssmeting. De reactietijd van de sensor wordt gemeten door de integrerende cilinder eerst bloot te stellen aan een gasmengsel met 50% CO₂ en vervolgens de gasstroom abrupt af te sluiten. Uit deze meting is duidelijk dat de sensor een korte reactietijd heeft van slechts 2.6 seconden, dit dankzij de kleine oppervlakte van de geïntegreerde cilinder (5 mm bij 5 mm).
Figuur 2: (a) Signaalrespons van zowel het detectiekanaal (rood) als het referentiekanaal (blauw) voor verschillende opgelegde CO₂ concentraties. (b) De genormaliseerde signaalrespons, verkregen door de normalisatie van het detectiekanaalsignaal ten opzichte van het referentiekanaalsignaal. Een legende van de verschillende opgelegde CO₂ concentraties is weergegeven in de tabel aan de rechterzijde.

Volledig geïntegreerde NDIR CO₂-sensor

Een combinatie van een LED en fotodiode, die spectraal is afgestemd op de absorptieband van CO₂, wordt gebruikt als de optische bron en detector voor de volledig geïntegreerde sensor. Het gebruik van de relatief goedkope componenten elimineert de noodzaak van de opto-mechanische modulator en de optische filters, waardoor de CO₂-sensor zowel compact als goedkoop te maken valt. Voor het referentiekanaal wordt een fotodiode gebruikt die licht detecteert tot een golflengte van 4 µm, maar niet hoger. Op deze wijze verandert het referentiesignaal niet wanneer de CO₂-concentratie wijzigt, aangezien de CO₂ absorptie plaatsvindt voor golflengtes groter dan 4 µm. De LED en de fotodiodes zijn geïntegreerd op het on-
derliggende siliciumsubstraat door middel van een flipchip bindingsmechanisme. Vervolgens wordt de half-open cilindrische caviteit aangesloten en de uiteindelijke sensor gevormd door middel van het aan elkaar binden van beiden substraten. Het schema van de volledig geïntegreerde sensor wordt getoond in Figuur 3. Na de fabricage is de sensor gemonteerd op een printplaat (PCB) met twee TIA’s om beide kanalen uit te lezen. De sensor werkt in quasi continu bedrijf (CW). In deze modus levert de LED het meeste licht, maar verbruikt hij ook het meeste elektrische vermogen. De signaalrespons van de sensor in de CW-modus is weergegeven in Figuur 4. In tegenstelling tot eerder metingen, waarbij de lichtbron en detectoren niet geïntegreerd waren op de chip, is hier nog een merkbare ruis waar te nemen op het genormaliseerde signaal. Dit valt toe te kennen aan de onbalans tussen de beide kanalen. Ook zijn er geen optische filters aanwezig, waardoor de bandbreedte van het licht in het detectiekanaal groter is dan voordien. Beide oorzaken dragen ertoe bij dat de detectielimiet van de geïntegreerde sensor licht verslechtert is ten opzichte van de eerste configuratie. Ook is de kruisgevoeligheid met waterdamp naderzocht en er is geen noemenswaardige kruisgevoeligheid vastgesteld. Verdere simulaties bevestigen dat deze kruisgevoeligheid onder de CO₂-detectielimiet van de sensor ligt. Voor het nagaan van de stabiliteitseigenschappen van de sensor op lange termijn, werd deze gedurende vijf dagen blootgesteld aan de omgevingslucht. Uit deze resultaten blijkt dat de stabiliteit afhankelijk is van temperatuurfluctuaties van de lucht in het laboratorium. Deze afhankelijkheid aan de temperatuur kan in toekomstige iteraties uit de signaalrespons genormaliseerd worden door de sensor te pre-kalibreren en het apart monitoren van de temperatuur van de chip.

Figuur 4: (a) Signaalrespons van het detectiekanaal (rood) en het referentiekanaal (blauw) van de volledig geïntegreerde sensor aan verschillende CO$_2$ gasconcentraties. (b) De genormaliseerde signaalrespons, verkregen door de normalisatie van het detectiekanaalsignaal ten opzichte van het referentiekanaalsignaal. Een legende van de verschillende opgelegde CO$_2$ concentraties is weergegeven in de tabel aan de rechterzijde.
English summary

The need for a miniaturized, low-cost CO$_2$ gas sensor is rapidly growing. Such a CO$_2$ sensor is of great importance in many applications such as air quality monitoring, greenhouse farming, and industry process control. In particular, as the European Union introduced a gradual ban on the usage of fluorinated gases in mobile air conditioning systems (MAC) for environmental and fire safety reasons, it is becoming clear that CO$_2$ is a suitable substitute for MAC refrigerants. However, concerns about the potential leakage of CO$_2$ on board arises, as the CO$_2$ concentration in the cabin can quickly reach up to a few percent and it imposes a high risk for the people on board as well as for the traffic. To address this problem, a fast responsive CO$_2$ sensor is needed on board to identify the leakage and thus removing the safety concerns. The application in MAC systems opens a huge new market for fast, low-cost CO$_2$ sensors.

Different technologies are available today for CO$_2$ sensing, including electrochemical, catalytic and spectroscopic techniques. Electrochemical and catalytic techniques can be advantageous in terms of cost, yet they suffer from disadvantages such as short-term stability, low durability and cross-response to other gases. In contrast, spectroscopic techniques offer long-term stability, high accuracy and high gas specificity. Spectroscopic gas sensing is based on probing the ‘fingerprint’ absorption lines of the gas with optical radiation in the infrared wavelength range (4.26µm for CO$_2$ sensing). Among the various spectroscopic techniques, non-dispersive infrared (NDIR) spectroscopy stands out as the most appropriate method for CO$_2$ sensing due to its simple implementation, compactness, and low cost. Furthermore, the large absorption coefficient of CO$_2$ at 4.26µm and the negligible overlap with the absorption bands of common gases provide high sensitivity and low cross-sensitivity. For these reasons, the CO$_2$ sensor market is largely dominated by NDIR sensors. However, current NDIR sensors are bulky, as a long interaction length (a few cm) is required to achieve a sub-100ppm limit of detection. They are also expensive as they are based on discrete co-assembled optical elements. This all prevents their use in price- and size-sensitive markets.

The goal of this thesis is to develop a miniaturized, low-cost NDIR CO$_2$ sensor that can be realized on wafer-scale. To integrate a few cm-long interaction length on a silicon chip with a small footprint, three different approaches are proposed in this work: the multi-slot silicon waveguide, the hollow metallic waveguide, and the hollow integrating cylinder. The integrating cylinder is chosen as the optimal approach due to power budget considerations, as well as fabrication complexity and production cost. The integrating cylinder can be considered as the 2D ver-
sion of an integrating sphere implemented on-chip. As in an integrating sphere, the multiple reflections of the incident light inside the cylindrical cavity provides a long interaction length with the gas. In this work, the integrating cylinder is investigated both numerically and experimentally. CO\textsubscript{2} sensing measurements are performed using external optical source and detectors. After the proof-of-concept demonstration of the integrating cylinder, the optical source and detectors are integrated on-chip. The characterization of the fully integrated sensor, including CO\textsubscript{2} sensing response, response time, cross-sensitivity to water vapor, and the long-term stability of the sensor are presented. Practical aspects such as temperature drift compensation and gas membrane protection are briefly discussed.

Non-dispersive infrared spectroscopy

A typical NDIR optical sensor consists of two channels: the sensing channel and the reference channel. The light emitted from the optical source (with emission around 4.26\(\mu\text{m}\)) is spectrally filtered and coupled to two photodetectors. For the sensing channel, the filtered light spectrum overlaps with the absorption band of CO\textsubscript{2}, such that the light intensity decreases with increasing CO\textsubscript{2} concentration. The absorption is well described by the Beer-Lambert law:

\[ I = I_0 \exp(-\alpha_{gas} c_{gas} L) \]  

(2)

where \(\alpha_{gas}\) is the absorption coefficient of the target gas per ppm, \(c_{gas}\) is the gas concentration in ppm, and \(L\) is the optical path length. At the reference channel, the light intensity does not change when the concentration of the target gas changes, which can be achieved either by using a reference wavelength at which no common gases absorb (typically at 3.9\(\mu\text{m}\) for CO\textsubscript{2} sensing), or isolating the reference channel from the ambient. The presence of the reference channel is mainly to eliminate influences such as source fluctuations and other common-mode noise.

On-Chip CO\textsubscript{2} sensing based on an integrating cylinder

The integrating cylinder consists of a metallized cylindrical cavity with access waveguides, formed by wafer bonding of two silicon substrates. To determine the optimal parameters of the integrating cylinder, 3D ray tracing simulations are carried out. The simulation results show that for a given cylinder with a certain radius (\(R\)) and height (\(h\)), the widths (\(d\)) of the access waveguides can be optimized to maximize the sensitivity of the integrating cylinder. For experimental characterization, we choose an integrating cylinder with a radius \(R = 2.0\text{mm}\), height \(h = 300\mu\text{m}\) and access waveguide widths \(d = 200\mu\text{m}\). Such an integrating cylinder yields an optical path length of 3.5\text{cm} and a total propagation loss of 7.4dB. The fabrication of the integrating cylinder involves deep reactive ion etching (DRIE) of silicon, gold deposition and lift-off, and wafer bonding, which are
all wafer scale fabrication techniques and therefore the sensor can be fabricated in large volume at low cost. To experimentally evaluate the fabricated integrating cylinder, a dedicated gas sensing setup as shown in Figure 1(a) is built. The infrared light emitted from a broadband thermal source (with emission spectrum shown in Figure 1(c)) is coupled into the input waveguide of the integrating cylinder by an optical fiber. An optical chopper is placed in between the source and the fiber to modulate the light. The two output waveguides of the integrating cylinder are used as the sensing channel and the reference channel. At the sensing channel, the output light is detected by an uncooled photodiode, and the photocurrent is first amplified by a trans-impedance amplifier (TIA) and then acquired by a lock-in amplifier. The only difference between the reference channel and the sensing channel is the working wavelength: $\lambda_{\text{sens}} = 4.26\mu m$ and $\lambda_{\text{ref}} = 3.8\mu m$, the wavelength selection is achieved by two optical band pass filters. The two channels are otherwise identical.

An Allan deviation measurement is first performed to determine the system noise and the optimal integration time of the lock-in amplifiers. The limit of detection of the sensing system is about $2 \times 10^{-4}$ (in absorbance units, AU) at an integration time of 1 second. The response of the integrating cylinder to changes in CO$_2$ concentration is measured by feeding the integrating cylinder with a series of sample gases containing various concentrations of CO$_2$. Figure 1(a) shows both the sensing signal and the reference signal when the CO$_2$ concentration is varied. It can
Figure 2: (a) Response of both sensing signal and reference signal to CO\textsubscript{2} concentration steps. (b) Normalized transmission, obtained by dividing the sensing signal with the reference signal. The CO\textsubscript{2} concentration steps are listed on the right side.

be seen that when the CO\textsubscript{2} concentration increases, the sensing signal decreases due to absorption, while the reference signal stays relatively stable. To eliminate the common mode noise present on both channels, the sensing signal is normalized w.r.t the reference signal and the normalized signal is shown in Figure 2(b). One can see that after normalization, the common fringes present on both signals are removed and the normalized signal becomes much more smooth. The limit of detection of the sensor is about 100ppm, which is consistent with the predictions from the Allan deviation measurement. The response time of the sensor is measured by first purging the integrating cylinder with 50% CO\textsubscript{2} and then abruptly shutting off the gas flow. The sensor has a short response time of only 2.6 seconds, owing to the small footprint of the integrating cylinder (5 mm × 5 mm).
**Fully integrated NDIR CO₂ sensor**

An LED-photodiode (PD) optopair spectrally matched to the absorption band of CO₂ is used as the optical source and the sensing photodiode for the fully integrated sensor. The use of the relatively cheap optopair bare chips eliminates the need for mechanical chopping and the optical filters, allowing for a compact and low-cost CO₂ sensor. For the reference channel, a photodiode with a cut-off wavelength at 4 µm is used. In this way the reference signal will not change as a function of CO₂ concentration as there is no spectral overlap between the response of the photodiode and the CO₂ absorption band. The LED and the photodiode chips are integrated on the bottom silicon substrate by flip-chip bonding, after which the sensor is formed by wafer bonding of the top substrate (containing the integrating cylinder) and the bottom substrate. The schematic of the fully integrated sensor is shown in Figure 3.

![Figure 3: 3D schematic of the fully integrated NDIR CO₂ sensor. The sensor consists of two silicon substrates. On the top substrate, the integrating cylinder is defined. On the bottom substrate, the LED and photodiodes are integrated on-chip.](image)

After fabrication, the sensor chip is mounted on a printed circuit board (PCB) containing two TIAs for testing. The sensor works in the quasi-CW mode, which provides maximum average optical power from the LED. The response of the sensor to CO₂ concentration changes is shown in Figure 4. It can be seen that unlike the ‘passive’ measurement of the integrating cylinder, the normalized signal is still quite noisy, which is due to the unbalancing of the two arms. This unbalancing together with the larger optical bandwidth reaching the sensing photodiode deteriorates the limit of detection of the full sensor. The cross-sensitivity of the sensor to water vapor is investigated. Measurement results shows that there is no notable water interference effects. Further numerical simulations reveal that the transmission change due to water vapor absorption is below the limit of detection of the sensor. To study the long-term stability characteristics of the sensor, the sensor was measured for a period of five days in ambient air. It is found that the long term stability of the sensor is subject to the temperature fluctuations in the laboratory. This temperature dependence can be potentially eliminated in future iterations by...
constantly monitoring the chip temperature and using look up tables created during a pre-calibration of the sensor.

Figure 4: (a) Response of both the sensing signal and reference signal to CO₂ concentration steps. (b) Normalized transmission, obtained by dividing the sensing signal by the reference signal. The CO₂ concentration steps are listed on the right side.
1 Introduction

The goal of this work is to develop a highly integrated and low-cost non-dispersive infrared (NDIR) CO\textsubscript{2} gas sensor. In this chapter, the background and motivation of the work presented in this thesis are given, as well as a review of the current NDIR sensors from both academia and industry. CO\textsubscript{2} sensors have many applications, including mobile air conditioning systems, greenhouse farming, demand-controlled ventilation, medical diagnosis, and industrial process control. The main driver of this work is the use of CO\textsubscript{2} as refrigerant in mobile air conditioning (MAC) systems due to safety reasons. To prevent accidents due to the decline of human cognitive ability in case of on-board CO\textsubscript{2} leakage, sensors are needed for cabin CO\textsubscript{2} concentration monitoring and warning. The application of CO\textsubscript{2} sensors in smart buildings and medical diagnosis will also be elucidated. An overview of the CO\textsubscript{2} sensing technologies is given, with an emphasis on the NDIR technique, which is the fundamental working principle of the sensor developed in this thesis work. An overview of state-of-the-art CO\textsubscript{2} sensors on the market is also given. Finally, the outline of this thesis and a list of publications realized during the course of this PhD are given.

1.1 Background and motivation

Carbon dioxide(CO\textsubscript{2}) is the fourth most abundant gas present in the earth’s atmosphere, and it is a major component of the terrestrial carbon cycle. During the past 200 years, the CO\textsubscript{2} concentration in the ambient earth’s atmosphere has increased
from ∼280ppm to ∼400ppm [1], largely due to human activities since the Industrial Revolution. CO₂ is the most significant gas responsible for global warming, accounting for 81% of the overall greenhouse gas emission, followed by Methane (10%) and Nitrous oxide (7%) [2]. Both natural and anthropogenic sources can emit CO₂. The natural sources include ocean-atmosphere exchange, plant/animal respiration, and soil respiration and decomposition. These natural sources release approximately 770 billion tons of CO₂ totally into the atmosphere per year. However, these natural sources are also huge carbon sinks, absorbing roughly the same amount of CO₂ by means of gas diffusion into the ocean, terrestrial vegetation, and land processes [3, 4], thus maintaining the global CO₂ concentration at a relatively safe level [5]. On the other hand, the amount of CO₂ emitted from anthropogenic sources is much smaller than that from natural emissions. Burning fossil fuels alone gives about 33.2 billion tons of CO₂ every year, which is about 87% of the total anthropogenic emission [6]. Although much smaller, the lack of anthropogenic carbon sink exerts pressure on nature’s capability of carbon regulation, causing the CO₂ concentration to increase extensively.

CO₂ plays an important role in both nature and human society. The atmospheric CO₂ captures the radiation from the sun and keeps the temperature in a proper range thus supporting life on earth. It is also the reactant for the process of photosynthesis of both micro-organism and plants. The generated oxygen is essential for the existence and evolution of life on earth. In the context of human society, CO₂ is widely applied in the food industry either as an additive for acidity regulations or as a refrigerant for food storage and transportation due to its high volumetric cooling capacity. In the medical domain, low-concentration (∼0.8%) CO₂ is used as a stimulus to treat apnea in preterm infants and to restore O₂/CO₂ balance in blood [7].

Despite the benevolent role in the above-mentioned applications, CO₂ can be potentially detrimental. As discussed earlier, CO₂ is the most significant greenhouse gas. The elevated atmospheric CO₂ concentration during the past few hundred years has caused quite a notable impact on the environment. A recent study shows that the global average temperature has increased by approximately 1°C compared to pre-industrial levels [8]. Although seemingly a negligible increase, we have suffered severe consequences, including extreme climate events [9], increased global economic inequality [10], and reduced biodiversity [11]. Physiologically, exposure to elevated CO₂ concentrations may produce a variety of health problems. For instance, one may experience headaches and dizziness at a CO₂ concentration of 5% [12]. Even at a concentration as low as 2000ppm, prolonged exposure can cause symptoms like drowsiness and sleepiness [13]. Strict laws and regulations have been issued in order to provide a safe and comfortable occupational environment [14, 15].

To fight against global warming and mitigate the aforementioned risks, it is
critical to monitor and control the CO\textsubscript{2} concentration and emission at every possible site. Therefore, a low-cost, highly sensitive CO\textsubscript{2} sensor with minimal power consumption is required in both domestic and industrial domains. The research of CO\textsubscript{2} gas sensing is receiving increasing attention in both industry and academia, driven by its vast market. The European advanced CO\textsubscript{2} sensor market is expected to grow to 3.6 billion dollars by 2021 with a compound annual growth rate of 14.6\% \cite{16}. The CO\textsubscript{2} sensor market can be segmented into different domains. In the chemical industry, CO\textsubscript{2} is a common reactant or byproduct of many chemical processes, therefore monitoring the CO\textsubscript{2} concentration is essential for good process control or for CO\textsubscript{2} emission management \cite{17,18}. In the food industry, CO\textsubscript{2} sensors are increasingly applied in both food production and packaging sectors \cite{19,20}. As CO\textsubscript{2} is often used to modify the atmosphere in food packaging to prevent spoilage, CO\textsubscript{2} sensors can be applied to identify the leakage in a package. In greenhouse farming, CO\textsubscript{2} is one of the key ingredients for photosynthesis. To control the CO\textsubscript{2} concentration and thus maximize the crop yield, CO\textsubscript{2} sensors are indispensable elements for precision agriculture. The application of CO\textsubscript{2} sensors is not limited to the aforementioned domains. The following sections enumerate several key envisioned applications of the cost-effective and sensitive CO\textsubscript{2} sensors developed in this work.

1.1.1 Mobile air conditioning and refrigeration systems

It is estimated that there are currently over one billion vehicles on the roads and this number will hit two billion in 2035 \cite{21}. Apart from the CO\textsubscript{2} emitted directly from the combustion process, greenhouse gases such as hydrofluorocarbons (HFCs) emitted from the mobile air conditioning (MAC) systems account for 30\% of the total greenhouse gas emission in passenger cars \cite{22}. The Intergovernmental Panel on Climate Change estimated that in 2015 alone, the greenhouse gas emission from the refrigerant in MAC systems is about 270 million tons of CO\textsubscript{2}-eq \cite{23}. Before the 1990s, chlorofluorocarbons (CFCs, commercially known as R12) were used as a refrigerant in MAC systems due to its properties such as non-flammability, low toxicity, and high chemical stability. However, its role as a refrigerant was phased out under legislative regulations due to concerns of ozone layer depletion \cite{24}. Other substances such as perfluorocarbons (FCs) and hydrofluorocarbons (HFCs) with virtually no ozone depletion potential have been developed, and they are the most commonly used substitute refrigerants (commercially known as R134a) to date in MAC systems. Nevertheless, R134a has a relatively high global warming potential (GWP\textsuperscript{1}), which raises concerns about its impact on climate change. The use of R134 as a refrigerant in MAC systems is currently under prohibition discussions \cite{25}. As global warming is becoming an im-

\textsuperscript{1}defined as the warming potential of a substance relative to CO\textsubscript{2} with the same mass over 100 years.
minent threat, it’s important to develop a new efficient, environmentally friendly, and safe refrigerant for MAC systems. The European Union initiated a directive that R134a must be replaced by a refrigerant with GWP smaller than 150 in all new passenger cars since January 2017 [26]. Two refrigerants have been proposed as alternatives of R134a: R1234yf and CO₂, and the battle between them is not yet settled. The advantage of R1234yf is that it’s compatible with the existing technology in existing air conditioning systems, therefore minimizing the cost for the development of new MAC systems. As a result, the majority of the automotive community has accepted R1234yf as the refrigerant in their MAC systems starting from 2017 [27]. CO₂ as a natural refrigerant (commercially known as R744), was extensively used before the discovery of R12 and it’s now being revisited and investigated due to its excellent properties such as high volumetric cooling capacity, non-flammability, and zero ozone depletion potential [28]. As the scientific society and the general public are calling for low CO₂ emissions, CO₂ is actually the most environment-neutral refrigerant, not only in MAC systems but also in buildings and homes. It has a GWP of only 1, which is about 1400 times smaller than the currently deployed refrigerant R134a [29]. Moreover, research shows that MAC systems with CO₂ are more efficient [30], thus the energy consumption for heating and cooling can be reduced which gives a low overall CO₂ emission. The technology of R744-based MAC systems is now being developed for series production, and several German car manufacturers are planning or have committed to using CO₂ as a refrigerant in their MAC systems [31, 32]. CO₂(R744) is considered as the next generation refrigerant in MAC systems due to its superior environmental and heating performance [33–35], and extensive standards for R744-based MAC systems are being developed by institutions such as the German Association of the Automotive Industry (VDA) [36].

The transition from fluorinated gases to CO₂ is beneficial from environmental and economic aspects, yet at the same time it raises safety concerns about potential leaks from the car’s onboard CO₂ tank. As the MAC compressor usually works at high pressure and the vehicle cabin is a small-volume, relatively confined environment, a trivial leakage of CO₂ onboard can lead to a surge of CO₂ concentration in the cabin. Researchers studied the CO₂ concentration level in the cabin of a B-segment car (with cabin volume ~ 2.1 m³). The results are shown in Figure 1.1. One can see that if the charge (CO₂ volume of 1.3 liter) is completely released within 60 seconds, the mean CO₂ concentration can spike up to 8.5% when no ventilation or re-circulation is activated. Even with maximum ventilation and re-circulation, the CO₂ concentration can go up to 6.5% in less than a minute [37]. Although CO₂ is not classified as a toxic gas, short-term exposure to relatively high (a few percents) CO₂ concentration levels can produce serious health problems [12]. For instance, one may experience headache, dizziness, and increased blood pressure when exposed to 4% CO₂ for just a few minutes. A
INTRODUCTION

Figure 1.1: $\text{CO}_2$ concentration at the driver’s place in the vehicle with four passengers. With a sudden leak, all the $\text{CO}_2$ is released in 60 seconds, the $C_{\text{Feet}}$, $C_{\text{head}}$ and $C_{\text{mean}}$ denote the $\text{CO}_2$ concentration at the driver’s feet, head and the average value, respectively.

(a) With no ventilation and no re-circulation. (b) With maximum ventilation and re-circulation. Adapted from [37].

higher concentration of 7% can lead to near-unconsciousness or even complete unconsciousness in less than a minute. As a consequence, a trivial leakage of $\text{CO}_2$ on board can lead to fatal situations for the people onboard as well as for the traffic. As $\text{CO}_2$ is colorless and odorless, it is essentially not possible to detect with human senses when a sudden leak happens. Therefore, a low-cost, sensitive, and more importantly fast-responsive $\text{CO}_2$ sensor is needed to constantly monitor the onboard $\text{CO}_2$ concentration to remove the safety concerns.

1.1.2 Demand controlled ventilation

$\text{CO}_2$ is an important byproduct of the human metabolism, the average $\text{CO}_2$ concentration in human breath is about 4% and an average person breaths out approximately 1kg of $\text{CO}_2$ every day [38]. In an enclosed environment such as schools, workplaces, and residential buildings, human beings are the major source of $\text{CO}_2$ generation. If poorly ventilated, the exhaled $\text{CO}_2$ will accumulate and the overall indoor $\text{CO}_2$ concentration will increase, leading to unpleasant conditions. Figure 1.2 shows a simulation of the indoor $\text{CO}_2$ concentration in a typical bedroom (volume = 21m$^3$) during the night for different ventilation rates [39]. One can see that when poorly ventilated (ventilation rate $<$ 2$L/s$), the $\text{CO}_2$ concentration can exceed 1000ppm in the first hour, and a maximum concentration of $\sim$ 5000 ppm is reached after 8 hours when no ventilation is activated.

To provide a pleasant environment for the occupants and eliminate complaints, various countries and regions have established standards and exposure limits of $\text{CO}_2$ concentrations for indoor air quality (IAQ) control [40, 41]. For instance, the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) suggested that the indoor $\text{CO}_2$ concentration should not exceed 1000ppm [40], which is only about three times higher than the typical outdoor concentration. These standards are rooted on numerous studies of the impact of elevated
CO₂ concentration on humans. It should be noted that CO₂ is not considered as a hazardous gas for human beings, however, intermittent- and long-term exposure to elevated CO₂ concentrations can lead to direct health risks. A review article summarizes the findings in 18 studies in various fields, including physiology, cognitive psychology, and environmental health [42]. It is shown that chronic (a few hours) exposure to a CO₂ concentration of less than 5000ppm can lead to symptoms such as a decline in cognitive abilities, bone demineralization and kidney calcification. To mitigate these health risks, constant ventilation is needed and it’s essential to maintain good indoor air quality. However, this comes with the cost of increased energy consumption. In Europe, the ventilation alone consumes more than 2% of the total electricity usage, amongst the biggest consumers of indoor electricity [43]. As most of the buildings (offices, schools, residential buildings) are not always occupied during the day, the demand for ventilation also varies over time. Therefore, it is possible to achieve energy saving without compromising the occupants’ well-being by using demand-controlled ventilation (DCV). Demand-controlled ventilation refers to the smart control of the ventilation rate according to the actual need. For the reasons mentioned above, the CO₂ concentration is commonly used as an indicator for indoor air quality and for DCV systems. It has been proven that energy consumption for ventilation is significantly lower when using DCV systems. A study performed in Norwegian primary schools shows that the energy consumption of a CO₂-based DCV system is reduced by 62% when compared to a constant air volume ventilation system [44]. In such DCV systems, a low-cost, sensitive (typically sub-100ppm) and low-power consumption CO₂ sensor is the key element to interrogate the indoor CO₂ concentration and adjust the ventilation rate accordingly.

1.1.3 Medical diagnosis

Another important application of CO₂ sensors is situated in the medical domain. CO₂ is a gaseous byproduct produced by the cellular metabolism in human bodies, and its concentration is strictly controlled by the carbonate buffer system in blood. The CO₂ concentration in the breath of a healthy individual is about 4% and deviations from this value may be an indicator of deterioration in the respiration system. The monitoring of the CO₂ concentration or partial pressure in the respiratory gas is called capnography. Capnography is a non-invasive, simple, and relatively low-cost technique that is commonly used in emergency medical services, intensive care units, and anaesthesiology. The measured waveform (capnogram) of the respiratory CO₂ concentration gives valuable information on the condition of the cardiorespiratory system. In the case of airway obstruction, the waveform will differentiate from that of a healthy individual. This deviation is typically used for diagnosis of cardiorespiratory diseases such as Chronic Obstructive Pulmonary
Disease (COPD), asthma, and pneumonia [45]. Figure 1.3 shows a comparison of a typical capnogram for a healthy person and an asthma patient [46]. A normal capnogram consists of four successive phases: (1) expiratory upstroke marked by the segment PQ, in this phase the CO$_2$ concentration increases rapidly corresponding to the expiration of mixed air; (2) Alveolar plateau (segment QR), in this phase, the CO$_2$ concentration is relatively high due to the constant air-outflow from the alveoli in the lungs, the non-flatness of segment QR reflects the continuous diffusion of CO$_2$ between alveoli and the capillary blood. The CO$_2$ concentration at the highest point (R), typically referred to end-tidal CO$_2$(EtCO$_2$), is an indicator of the respiratory effort; (3) inspiratory downstroke. At the end of the plateau phase, the CO$_2$ concentration drops quickly as the inspiration begins, noted by the segment RS; (4) respiratory baseline, in this phase, the CO$_2$ concentration is equal to the ambient concentration, and it’s used as the baseline in capnography. In patients with asthma, the airway is obstructed and the capnogram will exhibit a decreased slope in phase 1 (segment PQ) and an increased slope in phase 2 (segment QR), as shown in Figure 1.3. This deformation of the capnogram is of particular interest because it shows a higher resistance in the airway and marks a sign for the need of administration devices for forced ventilation and perfusion. In such capnography devices, real-time monitoring of CO$_2$ concentrations is vital and thus choosing a CO$_2$ sensor with short response time and high sensitivity is essential to recover the correct time-dependent CO$_2$ concentration.
1.2 Overview of gas sensing technology

In this section, a review of the state-of-the-art of gas sensing technologies is given. These technologies can be categorized into optical methods and non-optical methods. As we are targeting a nearly commercial-ready gas sensing device in this work, a brief overview of the state-of-the-art of commercially available CO\textsubscript{2} sensors is also presented.

1.2.1 Non-optical gas sensors

In general, there are three types of non-optical gas sensors: gas chromatographic sensors, mass sensors, and electrochemical sensors. Gas chromatography is the workhorse of analytical chemistry due to its high sensitivity and high reliability. Mass sensors determine the concentration of the gas by measuring its mass with a transducer. In terms of CO\textsubscript{2} sensing, electrochemical sensors are the main competitors of optical sensors. In what follows, each of the three techniques is discussed with respect to its advantages and limitations.

**Gas Chromatography**

Gas chromatography (GC) is a common technique used in analytical chemistry to analyze the concentration of each substance in a mixed gas. It can also be used to analyze non-gaseous samples provided that the compounds can be easily vaporized and separated without decomposition. A typical GC device consists of a carrier gas injection and regulation system, a sample injection port, a column, and a detector. The carrier gas is an inert gas such as helium or nitrogen and it is used to dilute the sample gas and force it through the column. The column is typically
a few meters long and arranged in a coil, and it is enclosed in an oven chamber to maintain a uniform temperature. At the inner wall of the column, a coating is applied to adsorb or desorb the substances based on their polarity, and the adsorption/desorption rate is molecule specific. Different molecules travel at different velocities in the column and thus separation of the substances occurs. At the end of the column, a detector is used to monitor the outlet stream and various substances can be recognized based on their retention time in the column. The most commonly used detectors in GC are either flame ionization detectors (FID) or thermal conductivity detectors (TCD). In a flame ionization detector, the concentration of the substances is determined by measuring the current change induced by the ionization of the substances between the two electrodes. A thermal conductivity detector relies on the change of thermal conductivity introduced by the molecules eluted from the column. GC devices have advantages such as high sensitivity, high reliability, and they can be used to detect a wide variety of substances. Nevertheless, the cost of such devices is relatively high and their application is limited by their bulky nature. In recent years, various efforts to miniaturize chromatographic gas detectors have been made, taking advantage of the Micro-Electro-Mechanical System (MEMS) technology [47–49]. However, their applicability is fundamentally limited by the need for a relatively long capillary column and a gas pump for the carrier gas.

Mass sensor
Mass sensors measure the change of resonance frequency upon the adsorption of the gas analyte by the transducer. The resonance frequency depends on the total mass of the transducer and the adsorbed analyte. Depending on the type of transducers, mass sensors can be further categorized into three classes, namely micro-cantilevers sensors, quartz crystal micro-balance sensors, and surface acoustic wave sensors [50]. The micro-cantilevers sensor uses a free-hanging cantilever as a transducer, while in quartz crystal micro-balance sensors and surface acoustic wave sensors, quartz crystal or piezoelectric material is used. To capture the gas, a gas-adsorbing material is applied on top of the transducer and the adsorption is gas specific. During the measurement, a change in the resonance frequency, amplitude, or phase due to the added mass of the adsorbed gas can be used to extract its mass. These mass sensors can be very sensitive, surface acoustic wave sensors with a detection limit of from ppm to ppb have been reported [51]. However, the main drawback of those mass sensors is the limited selectivity as the gas adsorption depends on the specificity of the coating. Various approaches have been investigated to overcome this limitation, including using a sensor array with different selective coatings for each analyte [52], as well as utilizing pattern recognition algorithms to distinguish different analytes [53]. Even so, finding the appropriate coating material for a specific gas with superior adsorption properties is not an easy practice, and the quality of the coating can be deteriorated with the presence
of contaminants such as water vapor and dust.

**Electrochemical sensors**

Electrochemical sensors measure the gas concentration by measuring a change in the electrical properties such as resistance, capacitance, or electric potential induced by the adsorption of the gas. Depending on the sensing material used, they can be further divided into three sub-categories: Metal Oxide Semiconductor (MOS) sensors, polymer sensors, and Carbon Nanotube (CNT) sensors. The MOS sensor consists of a thin layer of metal oxide film and its resistance changes through the redox reaction with the target gas at elevated temperature, and this change in resistance can be correlated to the gas concentration [53]. Metal oxides such as TiO$_2$, ZnO, SnO$_2$, and WO$_3$ are commonly used materials and their thickness can vary from a few nanometers to tens of nanometers [50]. MOS sensors have been intensively studied and widely deployed due to their advantages such as simplicity, compactness, easy fabrication, and low cost. However, their response time is relatively long, ranging from several tens of seconds to several hundreds of seconds [4]. In addition, the high operating temperature (from 150°C to 400°C) leads to higher power consumption [50]. Compared with MOS sensors, polymer-based gas sensors are more power-efficient as they can work at room temperature and the need for a micro-heater is eliminated. The working principle is quite similar to that of MOS sensors. Conducting polymers such as polyaniline (PANI) and polypyrrole (PPy) have been used for the detection of ammonia and hydrogen with excellent sensor response [54]. However, the response of such polymer-based gas sensors is temperature-dependent and they typically have a short lifetime. Carbon Nanotubes (CNT) are introduced in gas sensing applications mainly due to their unique properties such as high aspect ratio and low density, which leads to better sensitivity and shorter response times [50]. Gas sensors based on monolayer CNTs have been reported for the detection of NH$_3$ and NO$_2$, with a detection limit of 44 ppb and 262 ppb, respectively [55].

The aforementioned electrochemical sensors are advantageous because of their easy fabrication, low cost and high sensitivity to a wide range of compounds. The main issue of these sensors is the interference from unknown or unwanted molecules, as it is quite challenging to find a material with high sensitivity as well as good selectivity. The limited long-term stability is another issue that needs to be tackled.

### 1.2.2 Optical sensors

Optical gas sensors measure the gas concentration by means of optical absorption, emission, or scattering. Optical gas sensors offer benefits such as long-term stability, high sensitivity, and fast response. As most gases have their specific absorption lines in the optical spectrum, the interference between various gas species can be
minimized, resulting in high gas specificity. In what follows, several common optical gas sensing techniques are reviewed, including both indirect absorption methods such as refractive index sensors and photothermal/photoacoustic sensors, and direct absorption (or scattering) techniques such as tunable diode laser absorption spectroscopy (TDLAS), non-dispersive infrared spectroscopy (NDIR), and Raman spectroscopy.

1.2.2.1 Indirect methods

Refractive index sensor
Refractive index sensors are typically interferometric devices: they use the interference between two or multiple optical paths. In these refractive index sensors, an adsorbing coating is applied to capture the target gas, leading to a refractive index change of the coating layer, and this change in refractive index is correlated to the concentration of the target gas. Interferometric structures such as micro-ring resonators and Mach-Zehnder interferometers (MZI) are commonly used [56–59]. These interferometry-based sensors are quite sensitive: a tiny change in the refractive index can lead to a dramatic change in the output and a detection limit down to 5ppm has been reported for ammonia sensing [57]. In recent years, label-free refractive index sensors have also been extensively studied [60], in which a functionalization layer is applied to provide molecular binding with the analyte, and this binding gives rise to a change in refractive index that can be measured optically. Similar to the aforementioned non-optical sensors, refractive index sensors require a gas specific coating or a specific functionalization layer to achieve high selectivity.

Photothermal and photoacoustic sensors
Photothermal spectroscopy (PTS) is a highly sensitive and selective approach that relies on the indirect measurement of the optical absorption of the analyte. The absorption introduces a change in the thermal characteristics such as temperature, density, or pressure of the analyte, and this thermal response is probed and analyzed to gain information about the sample gas [61]. In PTS, the sample is illuminated with an excitation source (pump laser), the emission wavelength of which is carefully tuned to align with the characteristic absorption lines of the sample. Upon illumination, the sample is heated and the temperature change leads to a modulation of the optical properties (e.g. refractive index) of the transducer which can be probed by another optical beam (probe laser). Similar to refractive index sensors, interferometric configurations such as microring resonators and Mach-Zehnder interferometers are commonly used to retrieve this very small refractive index change [62, 63]. To further enhance the sensitivity, the output signal can be read out using schemes such as lock-in detection. In photoacoustic spectroscopy (PAS) sensing, the working principle is quite similar to that of PTS. Instead of optically probing the temperature change due to
absorption, PAS measures the acoustic wave generated by the periodic heating and expansion of the gas. If the pump beam is chopped or modulated, the expansion produces pressure waves at the same frequency of the pump beam that can be detected using a microphone. The key advantage of this technique is that it eliminates the need for an optical detector and the generated PAS signal can be amplified by tuning the modulation frequency to the acoustic resonance frequency of the transducer. A ppb-level detection limit for trace gases has been demonstrated using a quartz tuning fork [64]. Although quite sensitive, the relatively large size and fabrication complexity limit its applicability. A few efforts have been made recently to miniaturize PAS sensors on-chip by using MEMS technology [65–67].

1.2.2.2 Direct methods

Raman spectroscopy
Raman spectroscopy is a commonly used technique to identify and quantify unknown substances in forms of gas, liquid, and solid. This technique is based on the inelastic scattering of photons (known as Raman scattering) by molecules. In this process, the incident photon may gain or lose energy due to the interaction with the vibrational or rotational modes of the molecules, causing the wavelength of the incident light to be red-shifted (Stokes) or blue-shifted (Anti-Stokes). The shift in energy provides distinctive information about the target molecules. Raman spectroscopy is considered as a complementary method to infrared spectroscopy as some substances have limited absorption in the infrared region. However, spontaneous Raman scattering is typically very weak and difficult to measure. Stimulated Raman spectroscopy (SRS) can be used to resonantly enhance the scattering process by using a second optical beam. On-chip integration of stimulated Raman spectroscopy sensors has been demonstrated using silicon nitride photonic integrated circuits [68, 69]. Furthermore, the Raman signal can be enhanced using slot waveguides or plasmonic slot waveguides by locally enhancing the optical field in the waveguide slots [70, 71]. Alternatively, metallic nanostructures such as bow-tie nano-antennas and gold nano-domes can be used [72, 73]. However, additional elements such as a light source, isolators and optical filters with an extremely high extinction ratio have to be integrated for a fully integrated Raman sensor.

Tunable Diode Laser Absorption Spectroscopy
Tunable Diode Laser Absorption Spectroscopy (TDLAS) is the most common laser-based absorption technique for identifying and quantifying certain gas species. In TDLAS, the emission wavelength of the laser is tuned over one particular characteristic absorption line of the gas, the light is absorbed by the gas and experiences an exponential decay as it propagates. The intensity change is measured by an optical detector and this change in intensity is used to determine the concentration of the gas by means of Beer-Lambert’s law, which describes the relationship be-
between the light intensity and the gas concentration. TDLAS offers advantages in terms of specificity, fast response as well as high sensitivity, and it is well suited for trace gas sensing applications. The factors determining the detection limit of the TDLAS are threefold: (1) system noise; (2) strength of the selected absorption line; (3) effective path length of the beam [74]. The system noise can be reduced by modulating the laser at a high frequency instead of using continuous-wave operation. Additionally, the detector can be cooled to reduce the noise. As most of the gases have significantly larger absorption strength in the mid-infrared (MIR) region, MIR quantum cascaded laser (QCL) can be used as the optical source. To increase the effective interaction length of the beam and the gas, optical cavities or multi-pass cells are used and they can have an effective path length of several meters. A ppb-level detection limit has been demonstrated for water vapor using a QCL operating at 7.12<sub>µm</sub>, and a multi-pass cell with an effective path length of 3.5 meters [75]. Although quite sensitive, the system is quite bulky as a very long interaction length is needed to achieve low detection limits. The advances of photonic integrated circuit (PIC) technology give the potential for miniaturization of TDLAS sensors. On-chip TDLAS sensing of methane has been demonstrated at 1.65<sub>µm</sub> using an external distributed-feedback (DFB) laser and an uncooled InGaAs detector, in which a 10cm-long waveguide is integrated on-chip, yielding a detection limit of 100ppm [76]. Furthermore, The heterogeneous integration of single-mode short-wave infrared (<sub>2</sub>µm wavelength range) lasers on silicon photonics platform have also been reported and can be used in integrated TDLAS gas sensing [77].

Non-Dispersive Infrared Spectroscopy
As the name suggests, Non-Dispersive Infrared Spectroscopy (NDIR) is a non-dispersive technique, namely no dispersive elements (prisms, diffraction gratings, etc.) are required to interrogate the content of the spectrum in order to determine the gas concentration. The working principle of NDIR sensors is quite similar with that of TDLAS sensors, they both rely on direct light absorption of the target gas. However, instead of tuning the emission wavelength of the laser across a single absorption line, NDIR sensors use a broadband light source where the emission spectrum covers a ‘group’ of absorption lines. Therefore low-cost light sources such as light emitting diodes (LEDs) or thermal emitters such as micro-light bulbs can be used. The infrared light emitted by the light source is directed through the gas chamber, which allows the light to be absorbed by the target gas, and the change in light intensity is measured by the detector. NDIR sensors typically consist of two channels: the sensing channel and the reference channel. In the sensing channel, an optical filter with passband overlapping with the absorption region of the gas of interest is used, while the optical filter in the reference channel is chosen such that its passband does not overlap with the absorption spectrum of any commonly encountered gases. The NDIR technique has been used for quantita-
tive determination of a variety of gas species such as carbon monoxide, methane, and ammonia [78]. Depending on the type of optical sources, detectors as well as the interaction length, the detection limit ranges from a few ppm to hundreds of ppms. Moreover, the NDIR technique is particularly suited for CO₂ sensing, as CO₂ sensing is notoriously difficult with non-optical methods and NDIR opens a new path due to the significant absorption strength of CO₂ in the MIR region.

NDIR CO₂ sensors rely on the very high absorption coefficient that spans over $500 \text{ cm}^{-1}$ in the mid-infrared wavelength range around $4.26 \mu\text{m}$, which is two orders of magnitude higher than that of the typical trace gases. Due to this high absorption coefficient, the requirement on the interaction length is relaxed and an optical path length of a few centimeters is sufficient to detect small changes in CO₂ concentration. Furthermore, the choice of the source and the detector is less demanding: simple broadband optical sources such as micro-light bulbs and LEDs can be used, and on the detector side, low-cost devices such as thermopile detectors or pyroelectric detectors are sufficient. Due to the aforementioned advantages, NDIR CO₂ sensors are commercially successful: it is estimated that about 83% of the total advanced CO₂ sensors are based on the NDIR technique [16].

However, NDIR sensors tend to be bulky because to achieve ppm level detection a long (typically several cm) interaction length is required. The cost of current NDIR sensors is also high as they are typically based on discrete co-assembled optical elements, which limits their application in price and size sensitive markets. Intense efforts have been made to miniaturize NDIR sensors: integrated optical sources such as micro-heaters based on CMOS and MEMS technologies have been developed [79, 80], LED-photodiode opto-pairs have also been used [81], giving the potential for a highly miniaturized NDIR gas sensor. To increase the gas-light interaction length while maintaining the compactness of the sensor, the design of the gas cell can be optimized by using either a multi-pass cell [82] or an optical cavity with various shapes [83–85]. A recent study demonstrated an ultra-long path length of 25 cm on a chip area of less than $10 \text{mm}^2$, using highly reflective mirrors formed by photonic crystals on a silicon-on-insulator (SOI) platform [86]. The advantage of this approach is that it uses CMOS-compatible fabrication technology, giving it the potential for mass production with extremely low cost. Apart from these techniques to increase the interaction length, pre-concentrator coatings can be employed to effectively 'amplify' the gas concentration in the vicinity of the optical field, thus the required optical path length can be reduced [87]. The relatively simple structure and the use of low-cost optical sources and detectors enable multi-gas detection. By choosing multiple optical filters with different passbands, simultaneous measurement of various gas species can be achieved with one single NDIR sensor [83, 88]. Although convenient and straightforward, the need for optical filters to select the desired spectrum from the broadband source hinders the miniaturization of NDIR sensors. Typical NDIR sensors use filters that are
directly coated onto a packaged detector (e.g., in a TO9-can), which is not compatible with on-chip sensors. The need for optical filters can be eliminated by using narrow-band optical sources or detectors based on meta-material emitters or absorbers, in which the emission or detection spectrum can be tuned to align with the absorption band of the gas of interest. This tuning is achieved by engineering the shape and period of the nano-structure on the emitter or absorber [79]. These meta-material based emitters can be very efficient, with an emissivity almost equal to unity. The full width at half maximum (FWHM) of the emission spectrum is typically on the order of 300nm, which is sufficient to cover the absorption band of CO$_2$ around 4.26µm without introducing cross-sensitivity by other common gases. Using such an emitter together with a meta-material based thermopile detector, a filter-free NDIR sensor is demonstrated with a multi-reflection cell (5.7 × 5.7 × 4.5mm$^3$). The cell gives an equivalent path length of around 5cm and a sensitivity of 23ppm/√Hz is achieved [89]. From the commercial point of view, the size of the commercially available NDIR gas sensors has been dramatically reduced over the last decades [90], as illustrated in Figure 1.4. This size reduction is attributed to the aforementioned reasons, i.e., the use of optical sources such as MEMS-based micro-heaters and LEDs, as well as the miniaturization of the gas cell. However, the state-of-the-art commercial NDIR sensors are still a few centimeters big, limiting their applications such as in portable devices. The development of a highly integrated, cost- and energy-efficient NDIR gas sensor can not only extend its current applications in price- and size sensitive markets, but also introduce new applications in portable devices such as smartphones.
1.3 State-of-the-art NDIR CO\textsubscript{2} sensors on the market

As mentioned previously, NDIR CO\textsubscript{2} sensors account for the largest share in the gas sensor market. The NDIR CO\textsubscript{2} sensor market is mainly dominated by a few companies such as Gas Sensing Solutions, Senseair, Sensirion, and each company has its own story of advancing the performance of the sensors and reducing the cost and size. Table 1.1 lists a few state-of-the-art commercial NDIR CO\textsubscript{2} sensors from these manufacturers together with their major specifications. It should be noted that the purpose of the 'price' column is to offer a preliminary apprehension of the costs of the sensors, as the price can vary geographically and over time and it is also dependent on other factors such as purchase quantity. The maximum detectable CO\textsubscript{2} concentration of these sensors spans from a few thousands of ppm to a few percents and it is suitable for most common applications such as air quality monitoring and greenhouse farming. The detection limit is typically around 50 ppm, which is sufficient for most of the applications. To achieve such a low detection limit, a path length of a few centimeters is required. These sensors either use a single-pass cell and curl it to certain shapes or use cavities with specific shapes (eg. an ellipse) to attain the required path length on a small area. Nevertheless the size of these sensors is still on the order of ~ 3cm. Consequently, the response time of the sensors is typically long (~ 30s or longer) as most of the sensors are based on gas diffusion. The response time can be shortened by pumping the gas into the gas cell, leading to a fast response of the sensor. For instance, the response time of the sensor K30 (from SenseAir) is only 2 seconds at 0.5/l gas flow rate, which is 10 times shorter than that of its diffusion counterpart. Another important concern is the power consumption of the sensor, which is the summation of the power dissipated by the electronic circuits and the power consumption of the optical source and detector. In general, LED and photodiode opto-pairs are more energy-efficient than their thermal counterparts; the typical power consumption of an LED is tens of mW. The power consumption is also related to the measurement rate of the sensor, as a continuous measurement is not required in some cases and the power consumption can be reduced by increasing the measurement interval. To avoid sensor blindness by contaminants such as dust particles, a diffusion membrane with microscopic pores is applied to filter out dust particles from the airflow.

\[\text{From https://www.co2meter.com/ accessed 30-May-2020}\]
<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Product</th>
<th>Detection range</th>
<th>Detection limit</th>
<th>Response time (s)</th>
<th>Warm up time (s)</th>
<th>Dimensions ((L \times W, mm^2))</th>
<th>Power consumption (mW)</th>
<th>Price (€)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gas sensing solutions</td>
<td>GC-0034</td>
<td>0-1%</td>
<td>±45ppm + 3% of reading</td>
<td>30</td>
<td>3.3</td>
<td>31×20</td>
<td>75</td>
<td>114</td>
</tr>
<tr>
<td>SenseAir</td>
<td>SE-017</td>
<td>0-0.5%</td>
<td>±30ppm + 3% of reading</td>
<td>20</td>
<td>60</td>
<td>57×51</td>
<td>110</td>
<td>84</td>
</tr>
<tr>
<td>Sensirion</td>
<td>S03158A</td>
<td>0-4%</td>
<td>±30ppm + 3% of reading</td>
<td>20</td>
<td>–</td>
<td>35×23</td>
<td>48</td>
<td>–</td>
</tr>
<tr>
<td>Sensera</td>
<td>S01128A</td>
<td>0-0.5%</td>
<td>±50ppm + 5% of reading</td>
<td>30</td>
<td>120</td>
<td>32×20</td>
<td>89</td>
<td>–</td>
</tr>
<tr>
<td>Amphenol</td>
<td>M30032</td>
<td>0-0.2%</td>
<td>±30ppm + 3% of reading</td>
<td>180</td>
<td>120</td>
<td>30×15.6</td>
<td>125</td>
<td>97</td>
</tr>
<tr>
<td>Figaro</td>
<td>F010</td>
<td>0-0.5%</td>
<td>±50ppm + 3% of reading</td>
<td>120</td>
<td>–</td>
<td>32×17</td>
<td>50</td>
<td>105</td>
</tr>
</tbody>
</table>

*Table 1.1: Overview of state-of-the-art commercial NDIR CO\(_2\) sensors from several major manufacturers.*
1.4 Thesis objective and outline

This doctoral research is mainly driven by the increasing interest and the vast market for better gas sensing solutions, as well as by the technological opportunities offered by the mature silicon technology, which allow for the potential mass production of gas sensors with extremely low cost. As discussed earlier in the introduction, there is a growing need for low-cost, sensitive optical gas sensors. The main driver for this work is the gradual ban of the F-gas-based refrigerant in MAC systems enacted by the European Union. CO$_2$ has been proposed as an environment-friendly refrigerant due to its smaller GWP. To avoid possible incidents and remove the safety concern due to on-board CO$_2$ leakage, sensors are needed to monitor the CO$_2$ concentrations in the cabin. Fast response time (a few seconds) of the sensor is of paramount importance, while detection limit down to ppm-level is less critical. Another envisioned application of the sensor developed in this work is in indoor air quality monitoring for demand-controlled ventilation systems. In this scenario a highly-sensitive CO$_2$ sensor is required, while the response time of the sensor is less important. Such a CO$_2$ sensor can also contribute to a larger sensor network for internet of things (IoT) systems. In both cases, the cost and the energy consumption of the sensor should be low. As mentioned previously in the introduction, the NDIR technique gives the potential for highly miniaturized, low-cost, and highly sensitive gas sensors with low power consumption. In this thesis we aim at such a fully integrated NDIR CO$_2$ optical sensor by using silicon technologies. The remaining part of this thesis will address the challenges for the implementation of the sensor, which includes: (1) how to achieve a relatively long interaction length on a limited chip area to reach the desired detection limit; (2) how can we integrate the optical source and detector on the chip in a low-cost way; (3) how to optimize the overall sensor performance; (4) what is the future path to industrialize the sensor developed in this work.

The scope of this research extends from fundamental studies on gas sensing to experimental demonstrations of a sensor prototype with practically relevant specifications. In Chapter 2, a brief introduction to silicon photonics and MEMS technologies is given, with an emphasis on their extension to the mid-infrared wavelength range. Infrared absorption spectroscopy, more particularly non-dispersive infrared (NDIR) spectroscopy is explained in more depth, and one will see the numerous benefits of CO$_2$ sensing in the mid-infrared wavelength range. Relevant concepts behind absorption spectroscopy such as molecular energy levels and absorption line broadening mechanisms are also discussed.

In Chapter 3, three different configurations are proposed to realize an integrated on-chip gas sensor. These sensor configurations are based on multi-slot suspended silicon waveguides, hollow metallic waveguides and a metallic integrating cylinder. A high-level comprehensive study of the three sensor structures
is given, simulations are performed to investigate the waveguide losses, power budget and achievable detection limit for the three candidates. Based on this analysis, the optimal configuration is determined. One will see that the sensor based on a metallic integrating cylinder is the most promising approach in terms of its footprint, sensitivity, and potential for future low-cost fabrication.

With the optimal configuration of the sensor fixed, Chapter 4 provides a detailed investigation of the ‘gas cell’ part of the sensor, namely the metallic integrating cylinder. Such an integrating cylinder can be considered as a 2D version of an integrating sphere, implemented on silicon. Theoretical calculations of the integrating cylinder are given, simulation results on the loss, effective path length, and sensitivity are presented, as well as the response time of the sensor. The fabrication process of the sensor is discussed in detail. The details of the dedicated experimental setup to characterize the fabricated sensor are given. In this chapter an external optical source and external detectors are used. The experimental results on CO$_2$ sensing are presented.

Chapter 5 is devoted to the integration of the optical source and detector with the integrating cylinder. A mid-infrared light-emitting diode (LED) is chosen as the light source, due to its relatively narrow emission spectrum compared to a thermal emitter, such that the need for optical filters can be eliminated without introducing notable cross-sensitivity by other gases. Moreover, LED light sources have the benefits of low cost, ease of integration, and low power consumption. Mid-infrared photodiodes are used as detectors mainly due to their high sensitivity compared with thermal detectors such as thermopile detectors and bolometers. Optical simulations regarding the coupling efficiencies from the LED to the integrating cylinder and from the integrating cylinder to the photodiodes are performed. The process flow of the integration of the LED and photodiodes is also discussed. The design of the transimpedance amplifiers (TIA) to amplify the output signal is also presented. Eventually, the fully integrated optical sensor is characterized and the results are discussed.

Overall conclusions and future perspectives are given in Chapter 6.
1.5 List of publications

1.5.1 Patent


1.5.2 Publications in international journals


1.5.3 Publications in international conferences


INTRODUCTION

References


2.1 Optical Absorption Spectroscopy

Optical absorption spectroscopy is the technique used to determine the content and concentration of the analyte by means of direct optical absorption. Gas species absorb light at particular wavelengths in the electromagnetic spectrum, resulting in a reduction in the light intensity when the beam passes through the sample. Optical absorption spectroscopy measures this intensity attenuation as a function of wavelength to extract information on the gas species and their concentration.

2.1.1 Absorption spectra and the mid-infrared

Many chemical species absorb light at specific wavelengths in the ultraviolet (UV), visible, near-infrared (near-IR), and mid-infrared (mid-IR) spectral ranges. The absorption is chemical specific and thus this phenomenon can be used for sensing purposes. The mechanisms behind the absorption in different wavelength ranges are different. The absorption of UV or visible radiation originates from the electronic transition of the molecules, in which the outer electrons are excited from the ground state to an excited state upon light absorption. This is also the mechanism giving colors to the chemicals. Near-IR (from 780nm to 2.5µm) absorption arises from the excitation of overtones or combination modes of molecular vibrations in the mid-IR range. Such transitions are forbidden by the selection rules of quantum mechanics and therefore the absorption strength in this wavelength range is typically very weak. Unlike UV or visible radiation, mid-IR (from 2.5µm to 14µm)
The absorption from UV to mid-IR has been widely used in optical absorption spectroscopy. UV/visible absorption spectroscopy has been used to measure gaseous species such as ozone, NO₂, and SO₂, as well as the monitoring of nitrate in groundwater [2,3]. The absorption in the near-IR range is very weak and therefore in principle it is not suitable for sensing. However, the advances in datacom and telecom technologies enrich the availability of high-quality and low-cost optical sources and detectors, which still makes near-IR sensing systems attractive. For instance, on-chip glucose sensing at around 1590\text{nm} has been reported [4]. Moreover, the overtone bands in the near-IR can overlap significantly at high temperature and high pressure, making the absorption comparable to that in the mid-IR wavelength range. Based on this phenomenon, ammonia sensing at 1512\text{nm} has been demonstrated [5].
In symmetric stretching, the two C=O bonds stretch in opposite directions while in antisymmetric stretching one of the C=O bond stretches while the other compresses.

The mid-IR range is of particular interest for sensing applications, as the absorption strength can be significantly larger (up to three orders of magnitude) than that in the near-IR range. Moreover, various commonly-encountered gas species exhibit specific absorption in this wavelength range, therefore the mid-IR is usually called the ‘fingerprint region’ of the electromagnetic spectrum. Figure 2.1 shows the absorption coefficients for several common gases in the mid-IR range, and more information on other gas species can be accessed via well-known databases such as HITRAN [6].

For a molecule to be infrared active, a change in the dipole moment of the molecule must occur when the molecule vibrate or rotates. This change in dipole moment is induced by the internal motion (vibration and/or rotation) of the atoms. If the incident electromagnetic radiation is at the same frequency with this oscillation the radiation can be absorbed by the molecule. For instance, CO$_2$ is a centrosymmetric molecule with two C=O bonds and only two vibrational modes are present: symmetric stretching mode and antisymmetric stretching mode, as shown in Figure 2.2. In symmetric stretching, the changes in the electric dipoles of the C=O bonds are the same but in opposite directions, leading to no net change in the dipole moment and thus this vibration is IR inactive. However, this symmetric stretching contributes to the Raman scattering of CO$_2$ at a wavenumber of 1480 cm$^{-1}$, which can be used for Raman sensing. In the antisymmetric stretching, one of the C=O bonds stretches while the other compresses, resulting in an overall change in the dipole moment which gives rise to the strong absorption band at 2349 cm$^{-1}$ (4.26 µm) as shown in Figure 2.3. This absorption band spans over 80 cm$^{-1}$ and the absorption coefficient is around 300 cm$^{-1}$, which is two orders of magnitude higher than the overtone at 2.6 µm. Therefore for CO$_2$ sensing, the wavelength at 4.26 µm is often used.

---

1. reciprocal of wavelength, in cm$^{-1}$
2. temperature at 296 K and pressure at 101.325 kPa
2.1.2 Absorption line shape and line broadening

The shape of an ideal absorption line can be a narrow δ-like function. However, several line-broadening mechanisms are present and the absorption lines exhibit a finite linewidth. In general form, the absorption line profile can be described by:

\[ k(\nu) = S g(\nu - \nu_0) \]  \hspace{1cm} (2.1)

in which the line strength \( S \) is the integrated absorption coefficient given by:

\[ S = \int_{-\infty}^{+\infty} k(\nu) d\nu \]  \hspace{1cm} (2.2)

and \( g(\nu - \nu_0) \) is the line shape function normalized to unity. The width of the line shape is usually quantified by the Full Width at Half Maximum (FWHM). There are two well-known line broadening mechanisms that can alter the absorption line shape: Doppler broadening and pressure/collision broadening. Doppler broadening is caused by the Doppler frequency shift due to the finite molecular velocity distribution at a certain temperature. This broadening effect leads to a Gaussian profile of the absorption line and it is more prominent at low pressures. At atmospheric pressure, the pressure/collision broadening is dominant due to the molecular collisions between the same or different gas species. The resulting line shape profile is Lorentzian and the FWHM is proportional to the pressure of the gas \( \text{[7]} \). In general, the line shape of the spectral lines can be described by a Voigt profile, which is the convolution of Gaussian and Lorentzian line shapes.
2.1.3 Beer-Lambert law

Optical absorption spectroscopy is based on the Beer-Lambert law, which states the relationship between the transmitted light intensity and the concentration of the analyte, for monochromatic light:

\[ I = I_0 \exp(-\alpha l) \]  

(2.3)

where \( I \) is the intensity of light transmitted through the gas cell, \( I_0 \) is the initial light intensity incident on the gas cell, \( \alpha \) is the absorption coefficient (in \( cm^{-1} \)) of the gas and \( l \) is the optical path length (in \( cm \)) of the gas cell [8]. The absorption coefficient \( \alpha \) can be calculated by multiplying the gas concentration \( c \) (in partial pressure of the atmosphere, atm) with the molar specific absorptivity of the gas \( \epsilon \) (in \( cm^{-1} \text{atm}^{-1} \)). The ratio of the transmitted light intensity and the incident light intensity \( \frac{I}{I_0} \) is termed as optical transmittance \( T \). The absorbance \( A \), defined as the fraction of light absorbed by the gas, can be approximated using the Taylor expansion of the exponential function. At low gas concentrations, i.e., when \( \alpha l << 1 \), the absorbance can be written as:

\[ A = 1 - T = \frac{I_0 - I}{I_0} \approx \alpha l \]  

(2.4)

The absorbance \( A \) is unitless but it is often described in ‘absorbance units’ (AU). The limit of detection (LOD) of the sensing system is quantified as the noise equivalent absorbance (NEA, in AU) or the minimum detectable absorption coefficient (in \( cm^{-1} \)), this enables the comparison between different sensing methods without referring to the specific target gas. For gas sensing applications, it is often useful to convert the concentration of target gas to more typical units such as ppm by volume. For this conversion the ideal gas equation \( PV = Nk_BT \) is used, where \( P \) is the pressure, \( V \) is the volume of a closed cell, \( k_B \) is the Boltzmann constant and \( N \) is the number of molecules.

2.2 Mid-infrared Silicon Photonics

Over the past years, silicon photonics has emerged as an important platform for photonic integrated circuits (PIC) [9]. A PIC consists of various optical elements for light generation, detection and manipulation. While different material platforms have been developed to realize the photonic systems-on-chip, the silicon-on-insulator (SOI) platform is emerging as one of the most popular platforms. The main reason is that the SOI platform is compatible with CMOS MEMS processes thus the infrastructures in a CMOS MEMS foundry can be re-used without the need for intensive investment. The SOI platform comprises a silicon device layer (typically 220 nm thick, but can go up to several micrometers), a thermally grown
buried oxide (SiO$_2$) layer (with a typical thickness of 2µm), and a thick silicon substrate layer. The SOI platform provides a very high refractive index contrast which offers high optical confinement and small bend radii. Therefore an abundance of functionalities can be integrated on a chip with a very small footprint.

Most of the applications of silicon photonics are in telecommunications in the near-IR wavelength range. However, the mid-IR band presents a significant growth opportunity for silicon photonics due to its potential for sensing applications. Mid-IR low-loss waveguides have been demonstrated in the 2µm – 4µm wavelength range on SOI, with performance comparable to their near-IR counterparts [10–14]. Other building blocks for PICs such as grating couplers to couple light in and out of the chip [15, 16], high-Q microring resonators [13], wavelength (de)multiplexers based on arrayed waveguide gratings or planar concave gratings [12, 17], and polarization control devices [18, 19] have also been successfully demonstrated in the IR wavelength range below 4µm on SOI or on other CMOS/MEMS-compatible platforms.

The SOI platform can be used at wavelengths shorter than 4µm: beyond 4µm, the absorption of the buried oxide layer increases dramatically and the substrate leakage loss increases exponentially, making the waveguiding not practical [20]. However, the wavelength range above 4µm is of great interest as it contains strong absorption peaks of many important gases such as CO$_2$, CO, and NH$_3$, and the absorption coefficients are much larger than the overtones in the near-IR range. To extend the operating wavelength range of silicon photonics without leaving the CMOS foundry, materials other than silicon or silicon oxide can be used. Figure 2.4 shows the transparency windows of a few materials that are compatible with CMOS/MEMS technology. One can see that sapphire and silicon nitride are transparent up to 4.5µm and 7µm, respectively. They can be used as cladding material due to their relatively low refractive indices ($n_{\text{sapphire}} < 1.75$ and $n_{\text{SiN}} \approx 2$). Germanium is transparent up to 14.5µm and it can be used as core material together with silicon cladding due to its high refractive index ($n_{\text{Ge}} \approx 4$ and $n_{\text{Si}} \approx 3.45$), provided that the (Si) Ge can be grown on silicon with high quality. With these available materials, mid-IR optical sensing has been reported on various platforms such as silicon-on-sapphire for heavy water and CO$_2$ sensing around 4µm [21, 22], silicon-on-nitride for CO$_2$ sensing at 4.26µm [23], and Germanium-on-silicon [24] for Poly(methyl-methacrylate) (PMMA) polymer sensing at 9µm.

Another approach to extend the SOI platform to the mid-IR wavelength range is to locally remove the buried oxide beneath the silicon waveguide, making the silicon device layer suspended on the buried oxide layer through which the propagation loss due to oxide absorption is greatly reduced. Different geometries have been reported for silicon suspended waveguide structures, including suspended silicon ridge waveguides laterally supported by a thin continuous Si mem-
CHAPTER 2

Figure 2.4: Optical transparency window of several common materials used in mid-IR silicon photonics. In the yellow regions the material loss is negligible while the red regions denote heavy material absorption. Adapted from [20].

brane [25, 26], suspended waveguides laterally supported by a sub-wavelength grating [27, 28], suspended photonic crystal waveguides [29], and a strip waveguide supported by a periodic SiO₂ pedestal [30]. Figure 2.5 shows a few examples of such suspended silicon waveguide structures.

Figure 2.5: Reported suspended waveguide structures for the mid-IR wavelength range. Images taken from [25, 28–30].

2.3 Free-space and guided-wave optical sensing

An optical gas sensing system can either be implemented on-chip by using waveguiding structures such as dielectric waveguides or metallic waveguides, or it can be implemented by using bulk optical components such as beam splitters and folded mirrors in free space, which can also be considered as a light-guiding system. In
the free-space configuration, the light beam undergoes direct interaction with the environment while in the guided-wave configuration the light propagates along the waveguide and interacts with its surroundings. In general form, the transmission $T$ can be written as (Beer-Lambert law):

$$T = \frac{I}{I_0} = \exp(-\alpha_b L) \exp(-\Gamma \epsilon_{gas} c_{gas} L)$$  \hspace{1cm} (2.5)

where $\alpha_b$ (in cm$^{-1}$ or m$^{-1}$) is the attenuation coefficient of the waveguide, or the ‘effective’ attenuation per unit length in the free-space configuration, $c_{gas}$ is the gas concentration, and $\epsilon_{gas}$ is the molar specific absorptivity of the gas. $\Gamma$ is the external confinement factor of the waveguide and is a measure of the overlap between the optical field and the environment. In the free-space configuration, $\Gamma$ is unity as the whole light beam can interact with the environment. For waveguide-based sensors, the external confinement factor $\Gamma$ can be calculated by:

$$\Gamma = \frac{\int_C |\vec{E}(x,y)|^2 dxdy}{\int_\infty |\vec{E}(x,y)|^2 dxdy}$$  \hspace{1cm} (2.6)

in which the integration domain indices, i.e., $C$ and $\infty$, stand for cladding cross section and the whole computational domain, respectively. The sensitivity of the sensor can be quantified by the transmission change as a function of the variation in gas concentration:

$$S = \left| \frac{\delta T}{\delta c_{gas}} \right| = T \Gamma c_{gas} L$$  \hspace{1cm} (2.7)

We see that the sensitivity is waveguide length dependent. The length of the waveguide can be optimized by taking the derivative of $S$ with respect to waveguide length and then equating it to zero:

$$L_{opt} = \frac{1}{\alpha_b + \Gamma \alpha_{gas} c_{gas}}$$  \hspace{1cm} (2.8)

At this optimal waveguide length, the transmission of the waveguide is $1/e$, as can be obtained from Eq. 2.5 and Eq. 2.8. It can be seen from Eq. 2.8 that the optimal length of the waveguide is inversely proportional to the waveguide loss $\alpha_b$ and the external confinement factor $\Gamma$. For a free-space sensing system $\Gamma$ is unity, and the loss per unit length $\alpha_b$ is very low, thus the path length can be made very long to achieve high sensitivity. In a waveguide-based sensor, the loss of the waveguide itself is dominant therefore the waveguide length needs to be optimized accordingly to enhance the sensitivity. To reduce the footprint of the sensor, the waveguide is typically routed into a spiral [4]. The external confinement factor can be increased by careful design of the waveguide [10], by using the TM mode of the waveguide [11], or by using a slot waveguide in which the optical field can be greatly enhanced in the slot region of the waveguide [11, 32].
Another approach to increase the overlap of the light and the environment is to use a hollow-core waveguide. In this case the external confinement factor can be very close to unity. Different hollow-core waveguide structures exist including hollow-core fibers, metallic hollow waveguides and antiresonant reflecting optical waveguides (ARROW). In a hollow-core fiber, a silver (Ag) layer is usually deposited on the inner wall and a dielectric layer of AgI is formed by flowing iodine solution through the fiber [33, 34]. The AgI layer can protect the Ag from oxidation. By controlling the thickness of the AgI dielectric layer, the transmission window of the fiber can be engineered and optimized for a specific wavelength range. Similar to hollow-core fibers, hollow metallic waveguides also utilize reflective metals to guide light: they can either be implemented by coiling a metallic tube [35], or integrated on a metalized substrate [36, 37]. The design of the metallic hollow-core waveguide is very versatile as many metals have strong reflection in the IR wavelength range [38] and an optional dielectric layer can be deposited to tailor the working wavelength. Moreover, hollow-core waveguides can be potentially integrated on silicon substrates, as we will see in Chapter 3. Alternatively, one can use multiple dielectric layer stacks to guide light in the waveguide [39–41]. The light guiding is achieved by confining the leaky modes in a low-index core (air or liquid) by using high-index cladding layers, in which only the wavelengths in antiresonance (hence the name ARROW) will be guided.

2.4 Conclusion

In this Chapter, we discussed the absorption spectra of analytes in the UV/visible, near-IR and mid-IR wavelength ranges and the origins behind them. The mid-IR region is particularly interesting for sensing applications (especially for CO₂ sensing) as it contains strong absorption lines which are orders of magnitudes stronger than the overtones in the near-IR region. Beer-Lambert law was introduced, which is the common working principle for all absorption-based optical sensing techniques. Mid-IR silicon photonics enables the extension of the working wavelength range of the relatively mature silicon photonics technology, making it quite promising for sensing applications. To accommodate silicon photonics for mid-IR sensing applications, either the absorbing SiO₂ layer can be locally removed, or exotic CMOS-compatible material systems can be introduced. Free-space and guided-wave optical sensing were discussed. The latter allows for highly miniaturized optical sensors built on a small chip. For on-chip optical sensors, Beer-Lambert law was revisited by introducing the waveguide loss and the concept of external confinement factor. To increase the sensor sensitivity, a low waveguide propagation loss and a high external confinement factor are desired. Different approaches to enhance the sensor sensitivity by increasing light-matter interaction were reviewed. In the following Chapter, we will explore several pos-
sible sensor configurations that can be used in this thesis work and the optimal configuration will be chosen.
References


[34] Jason M Kriesel, Nahum Gat, Bruce E Bernacki, Rebecca L Erikson, Bret D Cannon, Tanya L Myers, Carlos M Bledt, and James A Harrington. Hollow core fiber optics for mid-wave and long-wave infrared spectroscopy. In Chemical, Biological, Radiological, Nuclear, and Explosives (CBRNE) Sensing XII, volume 8018, page 80180V. International Society for Optics and Photonics, 2011.


3

NDIR CO\textsubscript{2} Sensor Configurations

3.1 Introduction

In this Chapter, three possible configurations to realize an on-chip NDIR CO\textsubscript{2} sensor are proposed, including sensors based on a multi-mode slotted silicon waveguide, sensors based on a metallic hollow waveguide, and sensors based on a metallic integrating cylinder. The multi-mode slotted waveguide approach can be implemented on the SOI platform and therefore a compact and low-cost CO\textsubscript{2} sensor can be potentially obtained using silicon photonics technology. The hollow metallic waveguide approach uses metallic waveguide instead of dielectric waveguide for wave/light guiding and for gas sensing. It can also be fabricated using CMOS/MEMS-compatible technologies, resulting in potential mass production of the sensor at low cost. The idea of the integrating cylinder arises by asking ourselves the question “can we make the sensor even smaller than that using the metallic waveguide?”. The working principle of the integrating cylinder resembles that of an integrating sphere, which is widely employed in photometric applications. In what follows, the three approaches will be explored and compared. Simulations are carried out at both system and component level to address the design considerations such as waveguide loss and source/detector coupling efficiencies. The simulation results are used to calculate the power budgets of these three approaches. Finally, based on the power budget calculation together with other factors such as manufacturing complexity and cost, the optimal approach is chosen and fixed for further exploration.
3.2 Sensor system

The simplest implementation of an on-chip NDIR CO$_2$ sensor system is a single-beam configuration as shown in Figure 3.1a. The sensor consists of a broadband source, a sensing waveguide, and a detector. The main drawback of this configuration is that any light intensity variations originating from factors other than CO$_2$ absorption, such as source power fluctuations and waveguide aging, are also measured. It is therefore impossible to distinguish between the actual signal and the false signals. A dual-beam configuration can be used to eliminate these types of noise, in which part of the light from the source is directed to the reference arm and measured. A second detector is used to monitor the variations due to e.g. source intensity fluctuations, therefore auto-calibration of the sensor can be achieved. The reference arm can be implemented using a physical reference waveguide isolated from the ambient, therefore the reference signal will not change when there is a change in the CO$_2$ concentration. Alternatively, one can use a reference wavelength (e.g. 3.9 µm for CO$_2$ sensing) which does not overlap with the absorption band of CO$_2$ or any commonly encountered molecules, thus the need for the reference waveguide can be eliminated. In the dual-beam configuration, the optical power at the sensing arm and the reference arm, not taking into account the noise, can be written as:

\[
P_{\text{sens}} = P_{in} \exp(-\alpha_{WG}L) \exp(-\epsilon c \Gamma L) \\\nP_{\text{ref}} = P_{in} \exp(-\alpha_{WG}L)
\]

(3.1)

where $P_{in}$ is the incident power, $\alpha_{WG}$ is the waveguide propagation loss, $L$ is the waveguide length, $\epsilon$ is the molar absorption coefficient, $\Gamma$ is the confinement factor in air of the guided mode, and $c$ is the CO$_2$ concentration. The optical power change due to variations of CO$_2$ concentration can be calculated by:

\[
\Delta P = P(c + \Delta c) - P(c) = P_{in} \exp(-\alpha_{WG}L) \exp(-\epsilon c \Gamma L)(1 - \exp(-\epsilon \Delta c \Gamma L)) \approx P_{in} \exp(-\alpha_{WG}L)(\epsilon c \Gamma L)
\]

(3.2)
The CO$_2$ concentration change can thus be related to the optical power change by:

$$\Delta c = \frac{\Delta P \exp(\alpha_{WG} L)}{P_{in} \epsilon L}$$  \hspace{1cm} (3.3)

For a certain detector, the minimum optical power change can be detected (termed as noise equivalent power, or NEP) is fixed. Therefore to increase the sensitivity of the sensor (i.e. to measure a lower CO$_2$ concentration change), the following approaches can be explored:

- Increase the confinement factor in air $\Gamma$ of the waveguide
- Decrease the waveguide loss $\alpha_{WG}$
- Increase the optical power $P_{in}$ in the waveguide

In the following sections, these three approaches for increasing the sensor sensitivity will be discussed for the proposed sensor configurations, namely sensor configurations based on a multi-slotted silicon waveguide, a hollow metallic waveguide and a metallic integrating cylinder. The corresponding integration modalities of the optical source and detector will also be discussed. The major sub-components of the three sensor configurations will be investigated by numerical simulations, the simulated parameters (propagation losses, coupling efficiencies, etc.) are used to calculate and compare the power budgets of the three sensor configurations.

### 3.3 Multi-slot waveguide

In an integrated optical gas sensor, the design of the waveguide is crucial to ensure high sensing performance. A simple strip waveguide implemented on the SOI platform can be used for absorption-based optical sensing \[1\]. However, the high refractive index contrast between the core material (Si) and the cladding material (SiO$_2$ or air) results in strong confinement of the optical field in the high-refractive index core region. Therefore, the evanescent field available for sensing is usually quite limited, leading to poor sensitivity of the sensor. In this context, the slot waveguide is more suitable for sensing applications, as it provides a high optical field confinement in the slot region where the analyte to be detected is concentrated \[2, 3\]. The multi-slot waveguide has been proposed to further enhance the optical confinement in the low-index slot regions and to increase sensitivity of the sensor \[4–6\]. The concept of a multi-slot waveguide is similar with a slot waveguide, except that multiple slots are formed in the waveguide in between the periodic thin silicon strips. A 3D schematic of such a multi-slot waveguide is depicted in Figure 3.2. When the dimensions of the silicon strips and the slots are much smaller than a wavelength, the multi-slot waveguide core will act as an effective medium for the optical field \[4\]. The buried oxide (BOX) layer beneath the silicon waveguide
Figure 3.2: Schematic of a multi-slot silicon waveguide, the buried oxide (BOX) layer is partly under-etched to reduce the absorption loss of the optical mode, periodic SiO\textsubscript{2} pillars are formed for mechanical support.

is partially under-etched to reduce the absorption loss of the waveguide, as SiO\textsubscript{2} has a significant absorption at wavelengths beyond 4\,\mu m. To provide mechanical support for the multi-slot waveguide, periodic SiO\textsubscript{2} pillars are formed by selective etching of the BOX layer. The advantage of this approach is that the waveguide can be implemented on the SOI platform without leaving the CMOS/MEMS foundry, resulting in potential mass production with extremely low cost.

3.3.1 Waveguide design

Waveguide simulation

In the multi-slot waveguide proposed in this thesis work, the silicon strips are 300nm wide, separated by slots that are 200nm wide. To increase the coupling efficiency from a surface-emitting LED or a thermal light source, a large waveguide core is preferred. We set the height of the silicon strips to be 5\,\mu m, due to the limited aspect ratio of deep reactive ion etching (with a typical aspect ratio of \sim 20). For a wavelength of 4.26\,\mu m, the waveguide is highly multi-modal and the large waveguide core makes it easier to couple light from a surface-emitting LED or a thermal light source, which will be elaborated in the next section. Figure 3.3(a) shows the cross-section of the waveguide structure with 10 silicon strips. The multi-slot waveguide is simulated using a full vectorial finite difference eigenmode solver (Lumerical MODE solutions). The electric field profiles of the fundamental TE and TM modes of the multi-slot waveguide are depicted in Figure
Figure 3.3: (a) Cross-section of the multi-slot waveguide. (b) Horizontal components of the electric fields of the fundamental TE and TM modes in the middle of the waveguide. (c) 2D field profiles of the first three TE and TM modes.

It can be seen that due to the boundary conditions of Maxwell’s equations, the electric fields have peaks in the low-index (air) slots region for the TE mode, and have dips for the TM mode. Figure 3.3(c) shows the modal profiles of the first three quasi-TE and quasi-TM eigenmodes of the waveguide at 4.26\(\mu m\), with their corresponding confinement factors in air (\(\Gamma\)). The confinement factors in air for the quasi-TE and quasi-TM polarizations are about 80% and 35%, respectively. The propagation loss of the multi-slot waveguide originates from several contributions and can be reduced by the following approaches:

- Material absorption. The absorption loss of SiO\(_2\) above 4\(\mu m\) plays an important role for the propagation loss of the multi-slot waveguide. Calcula-
tions show that the absorption loss of bulk SiO$_2$ at 4.26\mu m is about 18dB/cm$^1$. Therefore, the SiO$_2$ underneath the silicon waveguide is partly under-etched to reduce the absorption loss of SiO$_2$.

- Substrate leakage loss. The leakage loss from the waveguide to the substrate is negligible for a standard silicon wire waveguide (450nm × 220nm) due to the high confinement of the optical mode. For the multi-slot waveguide proposed in this work, the optical modes are less confined (which is beneficial for sensing applications) due to the presence of the sub-wavelength slots. This may increase the leakage loss if the thickness of the BOX layer is not chosen properly, especially for the poorly-confined higher order modes. Simulation shows that the leakage loss can be minimized by using a BOX thickness of above 5\mu m.

- Sidewall scattering loss. The scattering at the waveguide sidewalls is an important contribution of the waveguide loss and it stems from the Rayleigh scattering due to the sidewall roughness of the waveguide after etching [8]. The scattering loss for the multi-slot waveguide is higher than that of a standard Si wire waveguide as more sidewalls are felt by the guided modes. However, since the strength of Rayleigh scattering is inversely proportional to the fourth power of the wavelength ($\sim \frac{1}{\lambda^4}$), the scattering loss is expected to be smaller for longer wavelengths such as 4.26\mu m. Furthermore, the waveguide perturbation caused by the sidewall roughness of the multi-slot waveguide can introduce mutual coupling between different guided modes, as a result, the overall propagation loss can be reduced.

- BOX pillar loss. As shown in Figure 3.2, the suspended multi-slot waveguide needs to be supported by periodic SiO$_2$ pillars. The field profile of a certain guided mode in the supported and un-supported region are different, leading to scattering losses at the BOX pillar. To estimate this loss, the field profiles of the eigenmodes in the supported region and un-supported region are simulated in Lumerical MODE, and the scattering loss for each mode is determined by calculating the overlap integral of each mode in the supported and un-supported regions. Simulation shows that for a SiO$_2$ pillar that is 5\mu m wide (in longitudinal direction), the loss is about 0.3dB/pillar, assuming that all guided modes carry equal optical power. This is however an overestimation, since light can also be scattered to other guides modes.

$^1$Calculated from [7].
3.3.2 Optical source and detector

3.3.2.1 Optical source

As shown in Figure 3.4, two types of optical sources are envisioned for the multi-slot waveguide: either a mid-IR III-V LED is heterogeneously integrated onto the silicon waveguide circuit using die-to-wafer bonding or transfer printing, or a thermal (black body) source can be integrated by locally doping the silicon waveguide and heat it to a target temperature.

**Mid-IR III-V LED**

Mid-IR LEDs offer superior power spectral density compared to thermal sources as the emission spectrum is narrow (relative to a thermal emitter) and can be better matched to the gas absorption band. The emission wavelength of the III-V mid-IR LEDs can be tuned by engineering the composition of certain elements in the III-V semiconductor alloys, provided that the alloys can be lattice-matched to a certain substrate in order to minimize the defects. Figure 3.5(a) shows the bandgap energy and the lattice constants of various III-V semiconductors. It can be seen that for LEDs to emit at $4.26 \, \mu m$ wavelength range ($E_g \sim 0.3 \, eV$), a narrow-bandgap material system based on antimonide semiconductors such as GaSb, AlSb, InSb and their alloys has to be used. Figure 3.5(b) shows the layer stack of one possible implementation of such a mid-IR LED emitting at $4.26 \, \mu m$, with the layer stack and the thicknesses extrapolated from [9]. The epi-layers consist of a homogeneous p-i-n diode based on AlInSb, a GaSb buffer layer, and a GaAs substrate layer. The LED can be integrated with the photonic waveguide circuits using transfer printing [10]. In this regard, a release layer of AlGaAs can be grown in between the GaAs substrate and the GaSb buffer layer.

To estimate the coupling efficiency from the transfer-printed LED to the silicon waveguide, a 3D ray tracing approach (in OpticStudio) is used as the thickness of the LED ($\sim 6 \, \mu m$) is much larger than a wavelength (in the material). The simulation setup is as the following:

- Simulation is performed using the non-sequential mode of OpticStudio, as
Figure 3.5: (a) Bandgap energy and lattice constants of the III-V semiconductor alloys. (b) Layer stack of the transfer-printable mid-IR LED proposed in this work, extrapolated from [9].

- The refractive index (RI) of the active region (Al$_x$In$_{1-x}$Sb) is calculated from the Ravindra relation [11]: $n = 4.084 - 0.62E_g = 3.9$. We assume one single RI for the LED because the change in the RI due to doping is very small and the layers are thin compared to the wavelength. The RI of silicon is set to be 3.42.

- The absorption coefficient of the LED is set to be 3000 cm$^{-1}$, which is a reasonable assumption for AlInSb alloys with low Al composition [12].

- A gold mirror is placed on the top surface of the LED, as the top surface can be fully metallized to realize the top contact.

- A rectangular volume source emitting at 4.26 µm is placed at the center of the LED, with isotropic emission in all directions. A rectangular power detector is placed at the cross-section of the silicon waveguide to record the power coupled to the waveguide.

In the simulation, a total number of 100,000 rays are emitted from the LED, and the optical power is equally distributed in each ray. The coupling efficiency from the LED to the silicon waveguide is simulated to be about 10% (in one direction).

**Thermal source**

The thermal (black body) source approach is conceptually more novel and it has the advantage of lower cost and easier mass manufacturability. The thermal source can be implemented by locally doping part of the silicon waveguide (or alternatively a cavity is etched and filled with a material that has a good emissivity in

$\text{Al}_x\text{In}_{1-x}\text{Sb}$ at $x = 6.1\%$ for light emission at 4.3 µm [7].
the 4\(\mu\)m region) and then heat it to a target temperature by electrical current. It should be placed on a freestanding membrane to increase the thermal isolation and hence limit the power consumption. The working temperature of the thermal source should ensure sufficient light in the required wavelength range while still be acceptable as operating temperature for the electrical connections.

To model the incoherent radiation from the thermal source, a dipole source is used in the finite-difference time-domain simulation (in Lumerical FDTD solver). A series of simulations are carried out to calculate the coupling from the thermal source to all the guided modes of the silicon waveguide, in each simulation the dipole source is placed at one position of the waveguide cross section. At each position, the transmission from the dipole source to the waveguide is averaged over 3 dipole orientations (x, y and z). As the thermal source is a homogeneous radiator, the coupling efficiency is averaged over the dipole positions (i.e. number of dipoles). The simulated coupling efficiency from the thermal source to the waveguide is about 30\% (in one direction). The difference in the coupling efficiencies between the thermal source (30\%) and the transfer-printable LED (10\%) is because the thermal source can be integrated 'in' the silicon waveguide, therefore the emitted light can be captured by the waveguide due to the high refractive index contrast between silicon and air. While in the LED case, the relatively high refractive index contrast between the III-V materials (3.9) and silicon (3.4) makes it more difficult for light to be coupled into the waveguide, leading to a lower coupling efficiency.

### 3.3.2.2 Optical detector

At the detector side, thermal detectors such as thermopiles and microbolometers are used. These devices can be operated over a broad temperature range. Moreover, they can be fabricated using silicon-based MEMS/CMOS technologies and can be monolithically integrated on the silicon waveguide circuit. In a thermopile detector, the thermoelectric effect is used as temperature readout. Two dissimilar materials are connected to form a thermocouple, and an electrical potential is generated when a temperature gradient is applied between the hot side and the cold side. The output impedance of the thermometer can be increased by using many thermocouples in series. The microbolometer employs an electrical resistance thermometer to measure the temperature of the radiation absorber, which can be achieved by partially doping the silicon waveguide. Waveguide-integrated microbolometers have been demonstrated on the SOI platform [14]. While these devices are typically less sensitive than semiconductor based devices, they perform better at higher temperatures, which is key for many applications (e.g. automotive). For the power budget calculation, we simply assume the coupling efficiency from the silicon waveguide to the thermal detector to be 100\%, as the microbolometer can be potentially integrated with the silicon waveguide.
Figure 3.6: Schematic of the metallic hollow waveguide, the waveguide is formed by bonding two metalized silicon substrates. The waveguide is etched on one substrate and the other one is just a planar substrate.

3.3.2.3 Transition loss

The use of a high index contrast (un-slotted) silicon waveguide allows for an efficient coupling of the radiation from both the III-V LED and the thermal source into the waveguide. However, the un-slotted silicon waveguide has to be transitioned to the multi-slot waveguide for optimal sensing, which leads to optical losses. To estimate this transition loss, the eigenmodes of the slotted waveguide region and the un-slotted waveguide region are calculated. Then the overlap integrals of each guided mode in the two regions are calculated to determine the transition loss for each mode, and the transition loss is then averaged over all guided modes assuming all guided modes carrying equal power. The simulated transition loss from the un-slotted waveguide to the slotted waveguide is $\sim 44\%$. One should note that this simulated transition loss is higher than the actual loss, as a certain guided mode in the un-slotted waveguide can couple to other guided modes in the slotted waveguide when the transition happens, which can not be considered as a transition loss.

3.4 Metallic hollow waveguide

The metallic hollow waveguide approach uses a metallic waveguide for sensing. Figure 3.6 shows the schematic of such a metallic hollow waveguide. In the hollow waveguide, light is guided by the reflective mirrors deposited on the sidewalls. Therefore all the light can be used for gas sensing ($\Gamma = 1$). Hollow waveguide structures such as hollow core fibers have been extensively used for sensing applications, as discussed in Chapter 2. In this thesis, we aim for a hollow metallic waveguide that can be integrated on a silicon chip using MEMS compatible technologies.
3.4.1 Waveguide design

The theory of the hollow metallic waveguide has been well-developed from the viewpoint of both wave and ray optics [14–16]. In general, the propagation loss of the hollow metallic waveguide depends on two factors: the dimensions of the waveguide and the reflectance of the mirrors on the sidewalls, and this dependency is given by [16]:

\[
\alpha = \left(\frac{u_{mn}}{2\pi}\right)^2 \frac{\lambda^2}{a^3} \frac{n}{n^2 + k^2}
\]  

(3.4)

where \(\alpha\) is the attenuation coefficient; \(u_{mn}\) is a mode-dependent parameter; \(a\) is the dimension of the cross section (circle or square) of the waveguide core; \(n\) and \(k\) are the refractive index and the extinction coefficient of the metal film, respectively, which determine the reflectance of the metallic mirror. One can see that the waveguide loss \(\alpha\) scales with \(1/a^3\), therefore to reduce the propagation loss of the waveguide, large waveguide dimensions are needed. It has been demonstrated that the propagation loss can be lower than 1dB/m at 1.55\(\mu m\) for a hollow waveguide with a core size of 1mm [17]. For the on-chip hollow core waveguide proposed in this work, we limit the height (etch depth of the Si substrate) of the core to be below 400\(\mu m\) to provide a high rigidity of the chip, as the thickness of the silicon wafer used in this work is about 700\(\mu m\). For such large core dimensions, the waveguide is highly multi-modal, and the waveguide can be easily simulated using less computation-intensive techniques such as ray tracing. Figure 3.7(a) shows the simulated propagation loss for a hollow waveguide with various widths. The height of the waveguide is fixed at 300\(\mu m\), and the simulation setup is depicted in the inset. In this simulation, an isotropic point source \((S)\) emitting at 4.26\(\mu m\) is placed at the center of the cross section of the waveguide, and two rectangular power detectors \((P_1, P_2)\) are placed in the waveguide with a separation distance of \(L\). The inner walls of the hollow waveguide are covered by a thin layer (thickness = 1\(\mu m\)) of gold, with the optical constants at 4.26\(\mu m\) given by Table 3.1. The waveguide loss is calculated by \(\text{Loss[dB]} = -10 \log(\frac{P_2}{P_1})/L\). For a square waveguide that is 300\(\mu m \times 300\mu m\), the propagation loss is about 2.3dB/cm. One should note that the simulated propagation loss is also related to the cone angle of the rays emitted from the source, as shown in Figure 3.7(b). The increased propagation loss for larger cone angles corresponds to the higher losses for higher order modes.

In the proposed hollow waveguide, we use gold as a mirror. The choice of gold over other metals originates from the fact that gold has a very high reflectance in the 4\(\mu m\) wavelength range, therefore the propagation loss can be minimized. The reflectance of a metal film depends on the wavelength of the incident light, the polarization state of the incident light, and the angle of incidence. At normal incidence, the reflectance of a metal film can be calculated from the optical con-
Figure 3.7: Metallic hollow core waveguide simulation. (a) Propagation loss of the hollow waveguide with various widths, the height of the waveguide is 300\(\mu\)m, an isotropic point source (at 4.26\(\mu\)m) is used and gold coatings are applied on the sidewalls as reflective mirrors. (b) Propagation loss for various cone angles of the source for a 300\(\mu\)m \(\times\) 300\(\mu\)m square waveguide, a cone angle of 90\(^\circ\) corresponds to half a sphere.

Figure 3.8: Reflectance of gold at 4.26\(\mu\)m against angle of incidence.

At oblique incidence, polarization becomes important and the average reflectance decreases as the incident angle increases, Figure 3.8 shows the reflectance of gold at 4.26\(\mu\)m as a function of incidence angle, the dip in the reflectance of the \(p\)-polarization corresponds to the Brewster’s angle.

Table 3.1 lists the optical constants (at 4.26\(\mu\)m) of several common metals [19, 20] and the calculated reflectance using Eq. (3.5). It can be seen that gold has the highest reflectance (\(\sim\) 99.5\%) among gold, silver, aluminium and copper. Another advantage of using gold as a reflector is that gold is chemically inert and stable,
leading to a better long-term stability of the sensor.

<table>
<thead>
<tr>
<th>Metal</th>
<th>n</th>
<th>k</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Au</td>
<td>1.0723</td>
<td>29.143</td>
<td>0.995</td>
</tr>
<tr>
<td>Ag</td>
<td>1.9658</td>
<td>31.121</td>
<td>0.992</td>
</tr>
<tr>
<td>Al</td>
<td>7.3795</td>
<td>40.921</td>
<td>0.983</td>
</tr>
<tr>
<td>Cu</td>
<td>1.4247</td>
<td>30.636</td>
<td>0.994</td>
</tr>
</tbody>
</table>

*Table 3.1: Optical constants and calculated reflectance of gold, silver, aluminum, and copper at 4.26 µm.*

3.4.2 Optical source and detectors

For the hollow waveguide, off-the-shelf mid-IR LED and photodiode bare chips can be used as source and detectors, and these bare chips can be very cheap when purchased in a large volume. The LED and photodiode chips can be integrated with the hollow waveguide with the following processes: (1) a trench is formed on a planar silicon substrate by selective silicon etching using potassium hydroxide (KOH) or Tetramethyl ammonium hydroxide (TMAH); (2) gold is deposited on this planar substrate as well as on the other substrate containing the hollow waveguide structure; (3) the LED/photodiode chip is then bonded in the trench (with emission/detection sides facing up) by flip-chip bonding; (4) the substrate containing the hollow core waveguide structures is flip-chip bonded with the planar substrate (with LED and photodiode on-chip). A detailed explanation of the integration procedures will be given in Chapter 5. Figure 3.9 shows the schematic of the hollow waveguide structure with an integrated LED (or photodiode).

The coupling efficiency from the LED to the hollow waveguide ($\eta_{LED}$) and from the waveguide to the photodiode ($\eta_{PD}$) are simulated with 3D ray tracing in OpticStudio. The simulation gives a coupling efficiency $\eta_{LED} = 8\%$ and $\eta_{PD} = 30\%$, the details of this simulation will be elaborated in Chapter 5.

3.5 Metallic integrating cylinder

The integrating cylinder can be considered as a '2D' version of an integrating sphere. In an integrating sphere, the incident light from the entrance port is uniformly scattered/diffused by the white diffusely reflective coating on the interior of the sphere. Light rays incident on any point of the inner surface are reflected multiple times before reaching the exit ports, leading to a long optical path length and hence the integrating sphere can be used for sensing applications [21, 23]. The fabrication process of the waveguide structures will be elaborated in Chapter 4.
implementation of the integrating cylinder proposed in this work is similar to that of an on-chip hollow waveguide: instead of using a long channel-like waveguide, a cylindrical cavity with rough sidewalls is used for light guiding. A schematic of the on-chip integrating cylinder is depicted in Figure 3.10.

3.5.1 Integrating cylinder analysis and simulation

The integrating cylinder consists of a hollow gold-coated cylindrical cavity along with hollow metallic waveguides, one at the input and two at the output, realized by Deep Reactive Ion Etching (DRIE) and wafer bonding. The theoretical analysis and numerical simulation of the integrating cylinder will be presented in detail in Chapter 4. Here we present some main results from the ray tracing simulation of the integrating cylinder:

- The total loss (in dB) from the input waveguide to one of the output waveguides and the equivalent path length of the integrating cylinder scale with the radius of the cylinder (nearly) linearly, and it is larger for smaller input/output waveguide widths.

- The sensitivity of the integrating cylinder, defined as the transmission difference with/without the presence of an absorbing medium (e.g., CO₂), is maximized for a certain cylinder radius and waveguide width combination.

- For a cylinder radius of 2mm and a waveguide width of 200µm, the sensitivity is maximized. The corresponding total loss and equivalent path length are 7.4dB and 3.5cm, respectively.
3.5.2 Optical source and detectors

For the integrating cylinder, the same off-the-shelf LED and photodiode bare chips can be used as source and detectors. The same processing procedures can be applied to integrate the LED/photodiodes on-chip, therefore, the coupling efficiencies for the source and detector are the same (also see Chapter 5).

3.6 Comparison

In the above sections, the three approaches to implement the on-chip NDIR CO$_2$ sensor have been introduced, and the waveguide losses and the source/detector coupling efficiencies have been investigated by numerical simulations. In what follows, the three approaches will be compared in terms of power budget considerations, fabrication complexity, and the envisioned cost of the sensor. Finally the optimal approach will be chosen for further exploration in the following chapters.

3.6.1 Power budget considerations

In this section, we will calculate the power budget for the three sensor configurations, using the parameters extracted from optical simulations. We assume a 50ppm detection limit for the sensor, which is similar to most of the state-of-the-art commercial sensors as introduced in section 1.3. A few remarks for the power budget calculations:
1. In the calculation we do not consider the additional noise produced by the electronic read-out circuitry, but limit the power budget to the optical link.

2. The absorption coefficient of 50ppm CO$_2$ is averaged over the absorption band (4.2 – 4.35µm) according to the Hitran database [24].

3. For the multi-slot waveguide approach, both a transfer-printable mid-IR LED and a waveguide-integrated thermal source will be considered as the optical source. While for the hollow waveguide and integrating cylinder approaches, a LED bare chip which can be flip-chip bonded on the silicon substrate is considered.

4. The noise equivalent power (or NEP, see Chapter 5) of the thermopile detector used in the multi-slot waveguide approach is calculated according to a state-of-the-art thermopile detector (MLX 90632, from Melexis). The NEP of the photodiode detector is calculated using the normalized detectivity $D^*$ of a high-performance photodiode (PD42BS, from Mid-IR Diode Optopair Group) with a sensitive area $S$ using the following formula:

$$\text{NEP} = \sqrt{S/D^*}$$

(3.6)

5. As the propagation loss of the multi-slot waveguide is dominated by the fabrication process (sidewall roughness), an exact value of the propagation loss is not available at the moment. We assume a 2 dB/cm propagation loss for the multi-slot waveguide, which is obtained from [25].

6. For the integrating cylinder, the propagation loss is calculated by dividing the total loss w.r.t the effective path length (further elaborated in Chapter 4).

7. The optimal waveguide length (1/e transmission) is determined by Eq. 2.8.

8. Other parameters such as propagation loss of the hollow waveguide, confinement factor in air, and source/detector coupling efficiencies are all obtained from the numerical simulations as discussed in the previous sections.

The power budget calculation yields the required source power such that a change in the transmitted power due to 50ppm CO$_2$ concentration change is three times larger than the noise produced in the detector. Table 3.2 shows the power budget calculation for the three proposed sensor configurations.

It can be seen that in terms of power budget, the hollow waveguide and the integrating cylinder are the most attractive approaches, the required source power for those two options are nearly identical, about 3.3µW. This level of output power from a mid-IR LED is achievable, as the typical power for a LED bare chip
### NDIR CO₂ Sensor Configurations

**Table 3.2: Power budget breakdown for the three sensor configurations.**

<table>
<thead>
<tr>
<th></th>
<th>Multi-slot WG (LED/thermal)</th>
<th>Hollow WG</th>
<th>Integrating cylinder</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noise equivalent power of detector</td>
<td>0.25</td>
<td>0.01</td>
<td>0.01</td>
<td>nW/√Hz</td>
</tr>
<tr>
<td>Integration time</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>s</td>
</tr>
<tr>
<td>Min detectable power change (3NEP)</td>
<td>0.75</td>
<td>0.03</td>
<td>0.03</td>
<td>nW</td>
</tr>
<tr>
<td>Required resolution</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>ppm</td>
</tr>
<tr>
<td>Absorption coefficient for 50ppm CO₂ averaged over 4.2 – 4.35μm</td>
<td>1.2e-3</td>
<td>1.2e-3</td>
<td>1.2e-3</td>
<td>cm⁻¹</td>
</tr>
<tr>
<td>WG propagation loss</td>
<td>2.0</td>
<td>2.3</td>
<td>2.2</td>
<td>dB/cm</td>
</tr>
<tr>
<td>Optimal WG length (1/e transmission)</td>
<td>2.2</td>
<td>1.9</td>
<td>3.5</td>
<td>cm</td>
</tr>
<tr>
<td>WG transition loss</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>dB</td>
</tr>
<tr>
<td>Total loss</td>
<td>10.35</td>
<td>4.35</td>
<td>7.4</td>
<td>dB</td>
</tr>
<tr>
<td>Transmission of sensing path</td>
<td>0.1</td>
<td>0.37</td>
<td>0.18</td>
<td>-</td>
</tr>
<tr>
<td>Confinement factor in air</td>
<td>0.6</td>
<td>1.0</td>
<td>1.0</td>
<td>-</td>
</tr>
<tr>
<td>Effective path length</td>
<td>1.3</td>
<td>1.9</td>
<td>3.5</td>
<td>cm</td>
</tr>
<tr>
<td>Absorbance of CO₂ over L_eff (50ppm)</td>
<td>1.6e-3</td>
<td>2.3e-3</td>
<td>4.2e-3</td>
<td>-</td>
</tr>
<tr>
<td>Source coupling efficiency</td>
<td>0.1/0.3</td>
<td>0.08</td>
<td>0.08</td>
<td>-</td>
</tr>
<tr>
<td>Detector coupling efficiency</td>
<td>1.0</td>
<td>0.43</td>
<td>0.43</td>
<td>-</td>
</tr>
<tr>
<td>Required power in 4.2 – 4.35μm range</td>
<td>46.8/15.6</td>
<td>1.0</td>
<td>1.1</td>
<td>μW</td>
</tr>
<tr>
<td>Estimated fraction of source power in 4.2 – 4.35μm range</td>
<td>0.3/0.04</td>
<td>0.3</td>
<td>0.3</td>
<td>-</td>
</tr>
<tr>
<td>Required source output power</td>
<td>156/390</td>
<td>3.3</td>
<td>3.7</td>
<td>μW</td>
</tr>
</tbody>
</table>

*Table 3.2: Power budget breakdown for the three sensor configurations.*

is about 10μW when the LED works in quasi-CW mode (50% duty cycle). And the output peak power can be 20× larger when pulsed mode (0.1% duty cycle)
is used, which gives a larger margin for the additional noise introduced by the electronic read-out circuits. On the other hand, the multi-slot waveguide approach requires a much larger (50× or 100×) source power to achieve the same limit of detection.

### 3.6.2 Fabrication complexity and footprint

As we are targeting a miniaturized gas sensor that can be manufactured in a large volume with low cost, the fabrication complexity of the sensor is of paramount importance. Wafer-level fabrication of the sensor using CMOS/MEMS compatible technologies will finally yield low cost sensors. In this respect, we rule out the multi-slot waveguide approach as the slots and beams of the multi-slot waveguide require very high aspect ratio etching which is difficult to yield. Moreover, stiction between the silicon beams of the multi-slot waveguide has been observed during the removal of the SiO$_2$ layer [25], which limits the fabrication yield. The hollow waveguide-based sensor and the integrating cylinder-based sensor rival each other in terms of fabrication complexity. The fabrication process involves optical lithography, deep reactive ion etching of silicon, metal deposition, and wafer bonding, which are all wafer-level processes that have been well-developed and widely applied in the semiconductor industry. However, the footprint of the sensor implemented with the integrating cylinder is smaller than that of the hollow waveguide, as the latter requires large waveguide bends to fit cm-long waveguide on a small chip area. Therefore, the integrating cylinder approach is now chosen as the optimal approach to realize an on-chip CO$_2$ sensor and will be explored further in the following chapters.

### 3.7 Conclusion

In this chapter, the system-level sensor configuration was firstly introduced, a dual-path configuration is chosen in which the reference path is used to mitigate the measurement errors due to common noise sources such as source power fluctuations and waveguide aging. Then we gave an overview of the three sensor configurations proposed in this thesis work: sensors based on a multi-slot waveguide, a metallic hollow waveguide, and a metallic integrating cylinder. We explored the design considerations and some key aspects of the three approaches such as waveguide loss and source/detector coupling efficiencies via numerical simulations. The simulation results were used for power budget calculations to determine the required optical power at the source side. The power budget calculations reveal that the hollow waveguide approach and the integrating cylinder approach require similar output power from the mid-IR LED, while the required source power for the multi-slot waveguide approach is much higher because of the assumed use of a
Si thermopile detector or bolometer. In addition, the complex fabrication of the multi-slot waveguide makes it not practical to realize a low-cost sensor with high yield. On the other hand, both the hollow waveguide and the integrating cylinder can be fabricated on wafer level, giving them the potential for mass production at low cost. In terms of the footprint of the sensor, the sensor based on integrating cylinder is smaller than the sensor based on hollow waveguide. Based on the above-mentioned reasons, the integrating cylinder approach is now chosen as the optimal approach to realize a compact and low-cost CO$_2$ sensor, and it will be further discussed in the following chapters.
References


In Chapter 3, the optimal approach to realize an on-chip CO$_2$ sensor has been chosen, namely the sensor based on an integrating cylinder. The goal of this chapter is to provide a comprehensive analysis of the integrating cylinder. First, a theoretical analysis is given, followed by detailed numerical simulations regarding the total transmission loss, equivalent optical path length, and the response time. The design parameters are then optimized by maximizing the sensitivity of the integrating cylinder with 3D ray tracing simulations. The fabrication process of the integrating cylinder will be discussed in detail, together with the proof-of-concept experimental results of CO$_2$ sensing. This chapter is based on our published article [1], with proper extension and supplement.

4.1 Introduction

The concept of the integrating cylinder is inspired by the integrating sphere. In an integrating sphere, the incident light is diffusely reflected by the coatings on the inner walls of the sphere, leading to a uniform intensity distribution within the sphere. Therefore the power collected by the detector does not depend on the relative positions of the source and the detector provided there is no substantial direct light path between the source and detector. As the incident light experiences multiple reflections before it reaches the detector, the optical path length is effectively
enhanced. The average path length travelled by a photon within the integrating sphere is given by:

\[ L_{\text{eff}} = \frac{2}{3} \cdot M \cdot D \]  

(4.1)

with \( D \) being the diameter of the sphere, which is typically on the order of a few centimeters. \( M \) is the sphere multiplier, which accounts for the enhancement of the optical path length due to multiple reflections inside the sphere \[2\]. The multiplier \( M \) can be calculated by the following equation:

\[ M = \frac{\rho}{1 - \rho(1 - f)} \]  

(4.2)

where \( \rho \) is the reflectance of the sphere’s internal surface, and \( f \) is the port fraction, which is the ratio of the area of the port openings to the total internal surface area. It can be seen that to increase the optical path length for a given sphere with a certain diameter, the reflectance \( \rho \) of the inner surface has to be high and the port fraction \( f \) has to be small. It has been demonstrated that the multiplier \( M \) can be as high as 50 for a standard integrating sphere (diameter = 5cm) with 98% inner surface reflectance, and the effective path length can exceed two meters \[3\]. Due to this very long optical path length, the integrating sphere is ideal for gas sensing applications, especially for gas species with low absorption coefficients such as CO and NO\(_2\). Optical sensing of various gas species have been demonstrated using an integrating sphere, with the optical path length ranging from 40cm to 5m \[4–7\].

The concept of the integrating sphere has been taken into the silicon photonics world to realize a 2D integrating cell with a long path length on a small chip area \[8\]. The integrating cell is implemented on the SOI platform, where high-reflectivity mirrors based on photonic crystals are used to confine the guided mode in the silicon slab. In the 2D integrating cell, the propagation of the slab mode follows a random path and experiences multiple reflections before escaping from the cell, just as in the 3D integrating sphere. An optical path length of 25cm is achieved on a chip area of \( \sim 10\, \text{mm}^2 \), due to the high reflectivity (> 99%) of the photonic crystal mirrors. Such a 2D integrating cell can be used for evanescent gas sensing, however, the effective interaction length between the slab mode and the ambient can be significantly smaller due to the high confinement of the slab mode. Moreover, the fabrication of the high-quality photonic crystal mirrors requires E-beam lithography, which hinders the application of the integrating cell in the low-cost gas sensing market.

In this chapter, we will take the concept of the 2D integrating cell a step further, by using a hollow metallic cylinder as the wave/light guide. In the integrating cylinder, the light is confined in the cylinder by the highly reflective mirrors coated on the inner surfaces. The hollow cylinder enables a total (\( \Gamma = 1 \)) interaction between the guided light and the absorbing medium (e.g., CO\(_2\) gas), hence the effective interaction length is not compromised. Furthermore, the integrating cylinder can be
Figure 4.1: (a) 3D schematic of the integrating cylinder with access waveguides. The integrating cylinder is formed by wafer bonding of two silicon substrates. (b) 2D analysis of the integrating cylinder, with $R$ being the radius of the circle and $d$ being the width of all access waveguides. $P$ is the input power, $P_{WG}$, $P_{rad}$ and $P_{prop}$ are the power coupled to the output waveguides, the power absorbed/radiated at the cavity sidewall, and the power lost during propagation in the cavity, respectively.

fabricated using wafer-level processes such as deep reactive ion etching and wafer bonding, which gives the advantage of potential high-volume production with low cost.

4.2 Theoretical analysis

Figure 4.1(a) shows a 3D schematic of the integrating cylinder, which consists of a hollow metallic cylindrical cavity with one input waveguide and two output waveguides. To confine the light in the cavity, a thin layer of gold is coated on the inner surfaces to form the high-reflectivity mirrors. The incident light is launched into the integrating cylinder from the input waveguide and it is extracted from the output waveguides. The two output waveguides can act as the sensing path and the reference path with the addition of two external optical filters. Similar to an integrating sphere, the incident light experiences multiple reflections in the cavity, leading to a long effective path length. For the analysis of the integrating cylinder, we assume the incident light is randomly reflected on the sidewall, which can be achieved by deliberately applying a random roughness on the sidewalls. Since the incident light is just specularly reflected on the parallel top and bottom surfaces (coated with gold), we can reduce the analysis of the integrating cylinder from 3D to 2D, as shown in Figure 4.1(b). With the assumption that the light incident on the sidewall is randomly reflected, the average path length of the light beam between
two consecutive reflections on the circle, is given by [9]:

\[ L_{\text{avg}} = \frac{4}{\pi} \cdot R \]  

(4.3)

where \( R \) is the radius of the circle. When incident light with power \( P \) is launched into the integrating cylinder from the input waveguide, it can be "lost" through three mechanisms: (1) extraction loss from the access waveguides (\( P_{WG} \)); (2) radiation/absorption loss upon incidence on the sidewalls (\( P_{rad} \)); (3) propagation loss in the cavity slab (\( P_{prop} \)). At steady state, the energy in the cavity is uniformly distributed. When we define the optical power incident on the sidewall as \( P_{cell} \), the three loss mechanisms can be described by [8]:

\[ P_{WG} = \beta P_{cell} \]  

(4.4)

\[ P_{rad} = \gamma P_{cell} \]  

(4.5)

\[ P_{prop} = \eta P_{cell} \]  

(4.6)

with \( \beta \) being the extraction coefficient for each of the waveguides, and it’s given by the ratio of the waveguide width (\( d \)) and the circumference of the circle:

\[ \beta = \frac{d}{2\pi R} \]  

(4.7)

\( \gamma \) is the fraction of power radiated/absorbed at the sidewall, and it is related to the reflectivity \( \rho \) of the sidewall mirror by:

\[ \gamma = 1 - \rho \]  

(4.8)

\( \eta \) is the loss factor of the light propagating in the cavity slab between two consecutive reflections on the sidewalls. It is either due to the waveguide loss \( \alpha_{prop} \), or due to the \( \text{CO}_2 \) absorption, characterized by the absorption coefficient \( \alpha_{\text{CO}_2} \). \( \eta \) can be written as:

\[ \eta = 1 - \exp(-L_{\text{avg}}(\alpha_{\text{prop}} + \alpha_{\text{CO}_2})) \approx L_{\text{avg}}(\alpha_{\text{prop}} + \alpha_{\text{CO}_2}) \]  

(4.9)

The summation of all losses equals the input power because of energy conservation:

\[ P = 3P_{WG} + P_{prop} + P_{rad} \]  

(4.10)

The fraction of power coupled to the waveguide \( P_{WG} \) can thus be related to the input power \( P \) using equations (4.4)–(4.7) and equation (4.10):

\[ P_{WG} = \frac{\beta}{3\beta + \gamma + \eta} P \]  

(4.11)
ON-CHIP CO₂ SENSING BASED ON INTEGRATING CYLINDER

The transmission from the input waveguide to one of the output waveguides can therefore be calculated by:

\[
T = \frac{P_{WG}}{P} = \frac{\beta}{3\beta + 1 - \rho + \frac{4}{\pi}R(\alpha_{\text{prop}} + \alpha_{\text{CO}_2})}
\]

(4.12)

The term \(\frac{4}{\pi}R\) is the average distance between two consecutive reflections, as shown in Equation 4.3.

The sensitivity \(S\) of the integrating cylinder is defined as the transmission change induced by CO₂ absorption, and it is can be obtained by taking the partial derivative of the transmission \(T\) w.r.t \(\alpha_{\text{CO}_2}\):

\[
S = -\frac{\partial T}{\partial \alpha_{\text{CO}_2}} = \frac{2\beta R}{(3\beta + 1 - \rho + \frac{4}{\pi}R(\alpha_{\text{prop}} + \alpha_{\text{CO}_2}))^2}
\]

(4.13)

The sensitivity \(S\) can be maximized as a function of \(R\) (assuming a fixed \(\beta\)) when

\[
3\beta + 1 - \rho = \frac{4}{\pi}R(\alpha_{\text{prop}} + \alpha_{\text{CO}_2})
\]

(4.14)

In Eq. 4.14, the term on the left-hand side can be considered as the “extraction” from the integrating cylinder, which consists of extraction by the access waveguides and the radiation/absorption loss from the sidewalls as the reflectivity \(\rho\) is less than 1. The term on the right-hand side is the absorption in the integrating cylinder, which consists of waveguide loss and CO₂ absorption loss. It can be seen from equation 4.13 that the integrating cylinder is most sensitive when the “extraction” equals “absorption”, i.e., when critical coupling occurs inside the cavity. The optimal radius can be calculated by solving equation 4.14, which gives:

\[
R_{opt} = \frac{3\beta + 1 - \rho}{\frac{4}{\pi}(\alpha_{\text{prop}} + \alpha_{\text{CO}_2})}
\]

(4.15)

Since \(1 - \rho << 3\beta\) in our design (\(\beta\) is of the order 0.03, while the reflectivity \(\rho\) is about 0.995 [10]), by substituting equation 4.15 into equation 4.12 we find the total transmission loss from the input to the output at optimal radius:

\[
\text{Loss(dB)} = -10 \log T = -10 \log \left(\frac{\beta}{3\beta + 1 - \rho}\right) \approx 10 \log 6 = 7.8 dB
\]

(4.16)

One can see that the total transmission loss is constant (7.8 dB) at the optimal cylinder radius. This is because at critical coupling, “absorption” equals “extraction”, adding 3dB to the loss, and in the “extraction” part, power is equally coupled into 3 access waveguides, which gives another 4.8dB loss (33%), so the
total transmission loss is 7.8 dB at the optimal radius. In order to calculate the equivalent optical path length, i.e., the effective path length of the beam from the input waveguide to the output waveguide, we write the transmission from the input to the output with/without CO\textsubscript{2} as:

$$
\frac{T_{\text{prop}+\text{CO}_2}}{T_{\text{prop}}} = \frac{1}{3} \exp(-\alpha_{\text{prop}} + \alpha_{\text{CO}_2}) L_{eq}
$$

$$
T_{\text{prop}} = \frac{1}{3} \exp(-\alpha_{\text{prop}}) L_{eq}
$$

(4.17)

The factor $\frac{1}{3}$ is because the extracted power is equally coupled into 3 access waveguides. We take the ratio on both sides of Equation (4.17), and since $\alpha_{\text{CO}_2} L_{eq} \ll 1$, in first order approximation:

$$
\frac{T_{\text{prop}+\text{CO}_2}}{T_{\text{prop}}} = \exp(-\alpha_{\text{CO}_2} L_{eq}) 
\approx 1 - \alpha_{\text{CO}_2} L_{eq}
$$

(4.18)

On the other hand, $\frac{T_{\text{prop}+\text{CO}_2}}{T_{\text{prop}}}$ can be also calculated from Equation (4.12),

$$
\frac{T_{\text{prop}+\text{CO}_2}}{T_{\text{prop}}} = \frac{3\beta + 1 - \rho + \frac{4}{\pi} R \alpha_{\text{prop}}}{3\beta + 1 - \rho + \frac{4}{\pi} R (\alpha_{\text{prop}} + \alpha_{\text{CO}_2})}
$$

(4.19)

The equivalent path length $L_{eq}$ can thus be calculated by equating Equations (4.18) and (4.19):

$$
L_{eq} = \frac{\frac{4}{\pi} R}{3\beta + 1 - \rho + \frac{4}{\pi} R (\alpha_{\text{prop}} + \alpha_{\text{CO}_2})}
$$

(4.20)

We notice that the numerator is the average path length between two consecutive reflections given by Equation (4.3), and the denominator is the summation of the coefficients of all losses. We substitute Equation (4.15) into Equation (4.20), and we can calculate the equivalent path length at optimal radius:

$$
L_{eq,\text{opt}} = \frac{1}{2(\alpha_{\text{prop}} + \alpha_{\text{CO}_2})}
$$

(4.21)

### 4.3 Simulations

As the dimensions of the integrating cylinder ($\sim 300 \mu m$ in height) is much larger than a wavelength ($\sim 4 \mu m$), we can use ray tracing to simulate the integrating cylinder. A commercial ray tracing software OpticStudio capable of doing 3D ray tracing is used. In the simulation, we use the non-sequential system module of OpticStudio since it allows multiple reflections to be accounted at a certain interface.
4.3.1 Total loss

In OpticStudio, a cylindrical cavity with three access waveguides (one input and two outputs) is constructed. We fix the height of the cylinder and the waveguides at 300µm, which is also the etch depth of the silicon wafer when the device is being fabricated. The lengths of the access waveguides are 1mm and they have equal widths \( d \). A gold coating is applied on the inner surfaces of the cavity and the waveguides to model the reflective mirrors. An isotropic point source emitting un-polarized monochromatic light at 4.26µm is placed at the center of the input waveguide, and two rectangular detectors are placed at the input waveguide \((D_1)\) and one of the output waveguides \((D_2)\). Figure 4.2(a) shows a 2D schematic of the simulation setup. A total number of 100,000 rays carrying a total power of \( P_0 \) is launched from the point source into the input waveguide, with the energy equally distributed in each ray. To model the random reflections of the incident light on the sidewalls, we implemented a Lambertian scatterer on the sidewall of the cavity. The probability of the directions of the reflected light follows a Lambertian distribution, and the reflectance on the sidewalls is determined by the incident angle of the light. In the simulation, the power at the two detectors are recorded as \( P_1 \) and \( P_2 \), and the propagation loss from the input waveguide to the output waveguide can be calculated by:

\[
\text{Total loss [dB]} = -10 \cdot \log\left(\frac{P_2}{P_1}\right)
\]

Figure 4.2(b) shows the total transmission loss from the input waveguide to the output waveguide for various cylinder radii \( R \) and access waveguide widths \( d \). It can be seen that the total loss increases as the cylinder radius increases or as the access waveguide width decreases. This dependency stems from the fact that light is trapped in the cylinder for a longer time and experiences more reflections, and the reflection losses from each reflection accumulate to give an overall larger transmission loss.

4.3.2 Equivalent path length

The equivalent path length is the average path length the incident light experiences from the input waveguide to the output waveguide. To simulate the equivalent path length of the integrating cylinder, we applied a volume absorbing medium with an absorption coefficient \( \alpha_{CO_2} \) in the cavity. Two simulations with/without the absorbing medium need to be executed to determine the equivalent path length \( L_{eq} \), which can be obtained from Eq. 4.18:

\[
L_{eq}(cm) = -\ln \frac{T_{prop+CO_2}}{T_{prop}} / \alpha_{CO_2}
\]
4.3.3 Design optimization

To optimize the design parameters of the integrating cylinder, namely to find a cylinder radius and waveguide width combination for which the sensitivity $S$ (defined as the transmission change for a CO$_2$ concentration change) is maximized, we plot the sensitivity $S$ for various cylinder radii and access waveguide widths in Figure 4.4. In this simulation, we set the absorption coefficient of the absorbing medium $\alpha_{CO_2} = 0.03$ cm$^{-1}$, which is equivalent to the absorption of 1000 ppm CO$_2$ averaged over 4.2–4.35 $\mu$m. It can be seen that for a certain waveguide width $d$, the sensitivity $S$ first increases and then decreases as the cylinder radius $R$ is increased. We define the optimal radius $R_{opt}$ as the cylinder radius when the sensitivity $S$ is maximized. The optimal radii for different access waveguide widths is summarized in Table 4.1.

We can see from Table 4.1 that the total transmission loss as well as the equivalent path length are almost constant at the optimal cylinder radius, which agrees with theory. For an integrating cylinder with a radius $R = 2$ mm and an access

with $T_{prop+CO_2}$ and $T_{prop}$ being the transmission ($P_2/P_1$) from the input waveguide to the output waveguide with/without the absorbing medium, respectively.

The simulated equivalent path length for various cylinder radius and waveguide width combinations is shown in Figure 4.3. One can see that the equivalent path length follows the same trends as the total loss, as the same argument applies when the cylinder radius increases or the width of the access waveguide decreases: the incident light stays in the cavity for a longer time and experiences more reflections, leading to a longer path length.
Figure 4.3: Equivalent path length from the input waveguide to the output waveguide for various cylinder radius and waveguide width combinations.

Figure 4.4: Sensitivity (in absorbance unit, or AU) due to the absorption of 1000 ppm CO$_2$ for various cylinder radius and access waveguide width combinations.
waveguide width $d = 200\mu m$, the equivalent path length is about 3.5 cm, and the total transmission loss from the input waveguide to the output waveguide is about 7.4 dB.

### 4.3.4 Sidewall roughness and source emission pattern

In the simulations presented in previous sections, a lambertian scatterer is applied to model the roughness on the sidewalls. The lambertian angular distribution of the reflected light on the sidewalls is an ideal approximation and the simulation results agree well with the theory. In reality, the sidewall scattering is implemented by deliberately roughening the sidewalls with a random function (as we will see in the next sections), therefore the angular distribution of the directions of the reflected light can be different from a lambertian distribution. To study the angular distribution of the directions of the reflected light over a rough surface, the simulation as shown in Figure 4.5 is performed:

- In the simulation, a parallel beam is incident on a flat surface, the surface is either roughened with a random function, or a lambertian reflector is applied on the surface.
- The amplitude of the roughness on the rough surface is determined by its root-mean-square (rms) value, the correlation length of the roughness is fixed at $30\mu m$.
- A polar detector is used to record the angular distribution of the directions of the reflected light.

The angular distributions of the directions of the reflected light from a rough surface and from a lambertian surface are shown in Figure 4.6. It can be seen that the incident light is reflected into certain discrete directions at a rough surface, while the directions of reflection from a lambertian surface follows a lambertian distribution. Moreover, in the roughened surface case, the angular distributions for smaller ($5\mu m$ and $10\mu m$) roughness amplitudes are 'more lambertian' than larger ($15\mu m$) amplitudes. Although the reflection pattern is not perfectly lambertian, the multiple reflections in the cylinder averages it out a bit. The angular distribu-

<table>
<thead>
<tr>
<th>$d$ [\mu m]</th>
<th>$R_{\text{opt}}$ [mm]</th>
<th>TL [dB]</th>
<th>$L_{\text{eq}}$ [cm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>1.4</td>
<td>7.5</td>
<td>3.4</td>
</tr>
<tr>
<td>200</td>
<td>2.0</td>
<td>7.4</td>
<td>3.5</td>
</tr>
<tr>
<td>300</td>
<td>2.6</td>
<td>7.6</td>
<td>3.8</td>
</tr>
<tr>
<td>400</td>
<td>3.0</td>
<td>7.6</td>
<td>3.9</td>
</tr>
</tbody>
</table>

Table 4.1: Optimal cylinder radius and access waveguide width combinations, and the corresponding equivalent path length and total transmission loss.
Figure 4.5: Simulation setup to study the angular distributions of the reflected light over (a) a roughened surface; (b) a lambertian surface.

Figure 4.6: (a)-(c) Angular distributions of the directions of the reflected light over a rough surface with various roughness amplitudes. (d) Angular distribution of the directions of the reflected over a lambertian surface.
Figure 4.7: Three types of integrating cylinders. (a) A specular cylinder. (b) A roughened cylinder. (c) A lambertian cylinder.

tion of the directions of the reflected light will ultimately influence the propagation loss and the sensitivity of the sensor, as the reflectance at the gold surface is also angle dependent as can be seen in Figure 3.8. To investigate the impact of the angular distribution of the reflected light on the sensitivity of the sensor, three configurations are compared: a specular integrating cylinder, a roughened integrating cylinder, and a lambertian integrating cylinder, as shown in Figure 4.7. In a specular cylinder, no roughness is applied on the sidewalls, therefore the incident light on the sidewalls is reflected specularly. In a roughened cylinder, the sidewalls of the cylinder is deliberately roughened with a random function, yet the incident light is still specularly reflected locally. In a lambertian cylinder, a lambertian scatterer is applied on the sidewalls and the directions of the reflected light follows a lambertian distribution. Another important factor determining the sensitivity is the emission pattern of the light source. Ideally the sensitivity of the sensor should not depend on the emission pattern of the light source, as the emission pattern can change over time which introduces error in the measurement. For instance, in a fully integrated sensor with an on-chip mid-IR LED, the emission pattern from an LED can change with temperature due to the temperature dependence of the refractive indices of the semiconductor materials. In the simulations presented in previous sections, a point source emitting uniformly in a half sphere (i.e. 180° cone) is used. To investigate how the emission pattern of the light source can influence the sensitivity, three types of sources are used for all three cylinder configurations: a uniform point source emitting in a 180° cone, a uniform point source emitting in a 120° cone, and a lambertian point source. The simulated sensitivity of the three types of integrating cylinder for the three types of source emission patterns are shown in Figure 4.8. One can see that the sensitivity of the specular cylinder strongly depends on the emission pattern of the light source, while for the roughened cylinder, the dependence is much smaller. In the lambertian cylinder, although the sensitivity is smaller, there is negligible dependence on the emission pattern of the light source. The reason behind this phenomenon is that the lambertian scatterer on the sidewalls reflects the incident light in all directions, as can be
seen from Figure 4.6(d). Therefore the optical energy in the cylinder is uniformly distributed, leading to the independence on the source emission patterns. Yet in the specular or roughened cylinder, the incident light is only reflected into one single direction or limited number of directions (as shown in Figure 4.6(a-c)), making the sensitivity less immune to emission pattern variations of the light source. One can also see from Figure 4.8(c) and (d) that in the roughened cylinder case, the emission pattern dependence of the sensitivity is a bit smaller when we change the roughness amplitude from $5\mu m$ to $10\mu m$, due to a 'more lambertian' reflection pattern for a $10\mu m$ roughness amplitude (Figure 4.6(a) and (b)). As discussed earlier, the emission pattern variations of the light source can introduce errors in the measurement. In this regard, the sidewalls of the integrating cylinder should be roughened in a way that the roughness profile produces a 'more lambertian' reflection pattern.
4.3.5 Response time simulation

The response time of the integrating cylinder is determined by the gas diffusion process, where CO₂ gas has to diffuse from the access waveguides into the cylinder. In many applications, the response time of a sensing system is defined as the time needed to reach 90% of the total response, which can be approximated by the diffusion time at which the minimum concentration in the cylinder reaches 90% of the set concentration. The CO₂ diffusion process inside the integrating cylinder is modeled using a multiphysics solver COMSOL, by numerically solving the diffusion equation:

\[
\frac{\partial c}{\partial t} + \nabla \cdot J = R
\]  

(4.24)

in which \( c \) is the concentration of the target gas, \( R \) is the net source, and \( J \) is the total diffusive flux, which can be approximated by Fick’s first law \( J = -D \nabla c \), with \( D \) being the mass diffusivity of the gas (CO₂ in this case). In this simulation, an integrating cylinder with a radius of \( R = 2\, mm \) and access waveguide width of \( d = 200\, \mu m \) is used. The initial CO₂ concentration inside the cylinder is set to be 0.0176 mol/m³ (~ 400 ppm), and three CO₂ sources with a concentration of 0.044 mol/m³ (~ 1000 ppm) are placed at the opening of the 3 access waveguides. The diffusivity of CO₂ in air is about 0.16 cm²/s (1 atm, 300K). In the simulation, no convection is activated and the molecular movement is only due to gas diffusion. Figure 4.9(a) shows the CO₂ concentration distribution inside the integrating cylinder at \( t = 2.5s \). We can see that the lowest CO₂ concentration appears at the right edge of the cylinder, which is logical as this is the furthest point from the CO₂ sources. The time-dependent CO₂ concentration at this point is depicted in Figure 4.9(b), it can be seen that the time needed for the CO₂ concentration at this point to reach 90% of the source level is approximately 2.6 seconds.
4.4 Device fabrication

The fabrication of the integrating cylinder starts with two blank silicon substrates. The cylindrical cavity with the access waveguides is etched by Deep Reactive Ion Etching (DRIE) on the bottom substrate. Then a thin layer of gold is deposited on both surfaces by sputtering to form the reflective mirrors. Finally, the two substrates are bonded using gold-to-gold direct bonding. The fabrication process flow of the integrating cylinder is depicted in Figure 4.10.

4.4.1 Deep reactive ion etching and sidewall scattering

Deep reactive ion etching (DRIE) is a highly anisotropic etching technique that is widely applied in the semiconductor industry. It is typically used to achieve deep trenches or holes with steep (nearly vertical) sidewalls on silicon wafers. Two major technologies exist to realize DRIE: the cryogenic process [13] and the Bosch process [14]. In the cryogenic process, the substrate is cooled down to a very low temperature ($< -100^\circ$), at which the chemical reaction resulting in isotropic etching is significantly slowed down, while the incoming ions can still bombard and etch the surfaces that are facing the incoming ions. Therefore a very steep sidewall can be obtained with this process. The Bosch DRIE process is also known as the time-multiplexed or pulsed etching process. It utilizes the concurrence of two alternating phases: (1) nearly-isotropic etching phase, which uses ion plasma to etch the substance as in standard reactive ion etching; (2) passivation phase, in which a chemically inert layer is deposited on the substrate. The passivation layer on the sidewalls of the trench/hole protects it from further etching by the ions in the etching phase, while the passivation layer at the bottom of the trench/hole can still be bombarded and sputtered off by the directional ions, thus the bottom of the trench is exposed for further etching. For the DRIE of silicon, source gases such as sulfur hexafluoride (SF$_6$) and octafluorocyclobutane (C$_4$F$_8$) are often used for etching and passivation, respectively [15].

For the fabrication of the integrating cylinder, the Bosch DRIE process is used for...

---

**Figure 4.10: Fabrication process flow of the on-chip integrating cylinder.**
Figure 4.11: (a) Silicon chip after DRIE, the chips consists of two cylinders. (b) Microscope image of the roughness on the sidewalls. (c) Amplitude distribution of the sidewall roughness.

silicon etching. As there is no DRIE tool in our research group, the DRIE of the silicon wafers is outsourced to Vmicro in France. The silicon substrate consisting of two integrating cylinders after dicing is shown in Figure 4.11(a). The thickness of the silicon substrate is 750µm and we choose an etching depth of 300µm to provide enough rigidity. The roughness of the sidewalls after etching is about 160nm, which is much smaller (< λ/10) than a wavelength (4.26µm), therefore the incident light can be considered specularly reflected in the access waveguides.

On the sidewalls of the integrating cylinder, we deliberately roughen the sidewalls to scatter the incident light randomly in the cylindrical cavity, the roughness is implemented during lithography mask design. A microscope image of the roughness on the sidewall can be seen in Figure 4.11(b). Figure 4.11(c) shows the amplitude distribution of the roughness function, the amplitude of the roughness has a Gaussian distribution, with a root-mean-square (rms) value of ~ 5µm, the correlation length of the roughness is about 50µm.

4.4.2 Gold deposition

Thermal evaporation and sputtering deposition are two common techniques that can be used to deposit high-quality metal films on a substrate. In this work we use sputtering deposition as it provides better step coverage for the vertical sidewalls of the integrating cylinder and the access waveguides. To determine the minimum thickness of the gold coating on the sidewalls, a 3D ray tracing simulation on the propagation loss of the hollow waveguide (300µm × 300µm) is performed. The
The simulation setup is similar to the one in section 3.4.1. An isotropic point source is used in the simulation and the propagation loss is calculated against the thickness of the gold coating, shown in Figure 4.12(a). It can be seen that in order to provide a high reflectivity, the thickness of the gold film on the sidewalls of the hollow waveguide should be at least 80 nm. To determine the minimum thickness of gold that needs to be sputtered, we did several gold sputtering tests (using UniveX 450, from Leybold) with a hollow waveguide that is $300 \mu m \times 300 \mu m$. In each sputtering test, the thicknesses of gold at 5 locations of the waveguide are measured, as shown in Figure 4.12(b). We see that the thicknesses of gold on the sidewalls are substantially smaller than the thickness on the top surface. The minimal thickness of the sputtered gold needs to be $\sim 700 \text{nm}$ for the sidewalls to be covered with thick enough gold (80 nm). Therefore, for the gold deposition of the integrating cylinder, we choose the thickness of Ti/Au (Ti for better adhesion) to be 50 nm/1 µm to make sure that the gold film can function as a high-reflectivity mirror.

### 4.4.3 Wafer bonding

Various wafer bonding techniques have been explored in both academia and industry, including thermo-compression bonding [16], solder bonding [17], eutectic bonding [18], and anodic bonding [19]. A comprehensive review of wafer bonding techniques can be found elsewhere [20, 21]. For the fabrication of the integrating cylinder, we use gold-gold thermo-compression bonding, in which heat and pressure are applied on the two substrates and bonding is achieved by the diffusion of gold atoms. The principal advantage of gold thermo-compression bonding is that it requires relatively low temperatures and pressures to achieve high bonding strength, as gold is highly resistant to oxidation and corrosion [22]. Since the bond-
Figure 4.13: (a) Bonding recipe. (b) Fabricated chip with a EURO 10 cent coin, the chip consists of two integrating cylinders. (c) Scanning electron microscope (SEM) image of the gold-gold interface after bonding.

ing strength strongly depends on the quality of the surfaces, the two substrates are cleaned and activated with an Argon plasma treatment before bonding (pressure = 53 mTorr, power = 100 W, duration = 60 s). During bonding (SB6, from SUSS Microtech), the chamber is first pumped down to vacuum (1e-3 mBar) to clean the chamber, then the two substrates are brought into contact and they are heated to 300 °C. A high pressure of ∼ 1MPa is applied on the substrates to overcome the surface asperities. The chamber pressure, substrates temperature and the tool pressure profiles are shown in Figure 4.13(a), the duration of the bonding is 30 min. A picture of the fabricated chip consisting of two integrating cylinders can be seen in Figure 4.13(b) and the SEM image of the bonding interface is shown in Figure 4.13(c).

4.5 Experimental settings

In this section, the characterization of the fabricated integrating cylinder will be discussed. First, the details of the experimental setup are presented, together with the characterization procedure. Then an Allan deviation measurement is performed to determine the system noise and the detection limit. Finally, the proof-
of-concept CO\textsubscript{2} sensing results with the integrated cylinder are presented, together with the response time measurement of the integrating cylinder.

### 4.5.1 Experimental setup

To experimentally evaluate the fabricated integrating cylinder, the experimental setup as shown in Figure 4.14 is built. The setup consists of two paths, one for the sensing arm and one for the reference arm. The light emitted from the optical source is butt-coupled into the input waveguide of the integrating cylinder via a multimode fiber (CIR500/550, Art photonics). The distance from the fiber to the input waveguide is \(\sim 50\mu m\). We use a multimode fiber with a large core size (\(\sim 500\mu m\)) to collect more power from the optical source. At the sensing arm, light from the output waveguide of the integrating cylinder is collimated and re-focused onto a photodiode by two identical lenses (C037TME-E, Thorlabs), with a band pass optical filter sandwiched in between. Since the lens-filter-lens system sits very close to the integrating cylinder, which will be filled with CO\textsubscript{2} gas for the sensing measurement, any gas diffusion into the gaps between the filter and the

![Figure 4.14: Schematic of the experimental setup. L: lens, MMF: multimode fiber, PD: photodiode, TIA: trans-impedance amplifier.](image)
lenses will lead to measurement errors. To prevent this error, the gaps between the filter and the lenses are sealed with rubber O-rings and isolated from the ambient. The signal from the photodiode is amplified by a trans-impedance amplifier (TIA) and then fed into a lock-in amplifier (SR830, Stanford Research System) for data acquisition and readout. The light is modulated by an optical chopper (MC1F10 blade with MC2000 controller, Thorlabs) placed between the source and the multimode fiber. We set the modulation frequency at 635Hz to avoid any interference or noise pickup from the overtones of the power grid. The duty cycle of the modulation is 50%. The reference arm only differs from the sensing arm in the optical filter, the two arms are otherwise identical. The two lock-in amplifiers are triggered by the optical chopper and are synchronized during the measurement.

4.5.2 Source and optical filters

We use a stabilized broadband optical source (SLS202, Thorlabs) in the setup. The source is a Tungsten halogen light bulb with internal electronic feedback loop to stabilize the output intensity. The working temperature of the source is 1900K, which gives a black body emission spectrum from 450nm to 5500nm, with the
peak wavelength at 1500nm. An additional collimation package (SLS202C, Thorlabs) is used to further collimate and clean the beam. The total output power of the collimated beam is $\sim 15mW$, with a stability $< 0.05\%$. A lens (LA5315-E, Thorlabs) is mounted at the end of the collimation package to focus the beam onto the multimode fiber. The emission spectrum of the source in the $3 \sim 5\mu m$ range is shown in Figure 4.15(a), with the whole spectrum shown in the inset. The dip around $4.26\mu m$ in the spectrum is due to the absorption of CO$_2$ in the atmosphere ($\sim$400ppm).

At the sensing arm, we use a band pass filter with pass band centered at $\lambda_0 = 4.25\mu m$, hence the light absorption by CO$_2$ at $\sim 4.26\mu m$ can be interrogated. The pass band of the reference filter is centered at $\lambda_0 = 3.75\mu m$, with no overlap with the CO$_2$ absorption band. Therefore the signal at the reference arm will not change due to CO$_2$ absorption. The transmission spectra of the sensing and reference filters are shown in Figure 4.15(b), with the absorption spectrum of CO$_2$ superimposed in the figure.

### 4.5.3 Photodiodes and transimpedance amplifiers

At the detector side, two identical commercial InAsSb photodiodes (P13023-013CA, Hamamatsu) are used for the sensing and reference arms. We use the photovoltaic (unbiased) mode of the photodiode, which is the preferred operation mode for low frequency measurement and low light level detection. The photodiode is soldered on a custom PCB board which is mounted on the cage system together with the lenses and the optical filter. The photodiode works at room temperature (un-cooled), and its major specifications are summarized in Table 4.2.

<table>
<thead>
<tr>
<th>Peak sensitivity wavelength $\lambda_p$</th>
<th>Cut-off wavelength $\lambda_c$</th>
<th>Photo-sensitivity $S$</th>
<th>Detectivity $D^*$ $\lambda = \lambda_p$</th>
<th>Noise equivalent power $\text{NEP} = \lambda_p$</th>
<th>Rise time</th>
</tr>
</thead>
<tbody>
<tr>
<td>$4.1\mu m$</td>
<td>$5.3\mu m$</td>
<td>$4.5 \text{ mA/}W$</td>
<td>$1 \times 10^9 \text{ cm} \cdot \sqrt{Hz/}W$</td>
<td>$7 \times 10^{-11} \text{ W/}\sqrt{Hz}$</td>
<td>$15 \text{ ns}$</td>
</tr>
</tbody>
</table>

*Table 4.2: Specifications of the photodiodes used in the experimental setup.*

We designed (with the help of Jasper Jans) a two-stage transimpedance amplifier (TIA) to convert the photo-current from the photodiode into a 'photo-voltage'. The circuit of the TIA is schematically shown in Figure 4.16(a). The TIA is designed to amplify a photo-current of $10\text{nA}$ to a voltage of $100\text{mV}$, with a transimpedance gain of $10^7\Omega$. We set the bandwidth of the TIA to be only $2\text{kHz}$, considering the relatively low modulation frequency (635Hz) of the optical source. An electronic circuit simulator LTspice (Analog Devices) is used for the simulation of the TIA.
The simulated transient response and the frequency response are shown in Figure 4.16 (b-c).

4.5.4 Gas generation

$\text{CO}_2$ gas is supplied from certified gas cylinders (Praxair inc.), and nitrogen is used as a carrier and diluting gas to generate the sample gas with various concentrations of $\text{CO}_2$. The concentration of $\text{CO}_2$ in the diluted gas mixture can be calculated by:

$$C_{\text{CO}_2}(\text{ppm}) = \frac{F_{\text{CO}_2} \cdot 10^6}{F_{\text{CO}_2} + F_{\text{N}_2}} \times C_{\text{cylinder}} \quad (4.25)$$

where $F_{\text{CO}_2}$ and $F_{\text{N}_2}$ are the flow rates of $\text{CO}_2$ and nitrogen, respectively. The flow rates are controlled by two mass flow controllers (M13212646C, Bronkhorst). $C_{\text{cylinder}}$ is the $\text{CO}_2$ concentration in the gas cylinder. For the generation of the sample gas with $\text{CO}_2$ concentration larger than 1000ppm, a gas cylinder with 50% $\text{CO}_2$ is used. And a gas cylinder with 1000ppm $\text{CO}_2$ is used to generate the sample.
gas with CO₂ concentration lower than 1000 ppm. A gas tube made of Polyvinylidene Fluoride (PVDF) is used to deliver the gas mixture to the integrating cylinder. The diluted gas mixture is fed into the integrating cylinder by bringing the gas tube in close proximity to the integrating cylinder (not shown in the setup).

4.6 Proof-of-concept experimental results

As discussed in section 2.1.3, the signal (i.e. absorbance of CO₂) can be related to the transmission change by:

\[ S = 1 - T = 1 - \frac{I_A}{I_{0A}} \]  

in which the subscript 0 denotes the signal without the presence of CO₂ and A denotes the active (sensing) channel. The limit of detection (LOD) of the sensing system can be described by the minimum detectable transmission change \( \Delta I_{A0} \). Although quite stable, the source power fluctuations and particularly the source drift can contribute to the transmission change, leading to an error in the measurement. To eliminate this error caused by source fluctuations and drift, the signal at the reference channel \( I_R \) is constantly measured and used as a substitute for \( I_{0A} \). The imbalance between the two channels can be corrected by the factor \( \frac{I_R}{I_{0A}} \), which can be obtained by flushing the integrating cylinder with pure N₂ at the beginning of the measurement. The normalized absorbance signal can thus be written as:

\[ S_{\text{norm}} = 1 - \frac{I_{R0}}{I_{A0}} \frac{I_A}{I_R} \]  

With this formula, the common-mode noise and source fluctuations at both arms can be eliminated. We use the quantity \( S_{\text{norm}} \) as a figure of merit to evaluate the performance of the sensing system based on the integrating cylinder in all following measurement results.

4.6.1 Allan deviation measurement

The measurement resolution of any practical measurement system is ultimately limited by the noise in the system and the drift of the system, and this limitation can be characterized by means of the Allan deviation measurement [24]. The Allan deviation measurement offers a quantitative description of the variability of the measured quantity averaged over different time intervals, which enables the quantification and differentiation of the various noise processes present in the measurement system. The Allan deviation is a two-sample deviation by averaging the differences between two adjacent values of a time-series measured signal. The
Allan deviation is mathematically given by \cite{25}:

\[ \sigma_y(\tau) = \sqrt{\frac{1}{2\tau^2} \left( \left( x_{n+2} - 2x_{n+1} + x_n \right)^2 \right)} \] (4.28)

where \( \tau \) is the averaging time period, and \( x \) is the measurement quantity. Using this equation, we can calculate the Allan deviation of the normalized signal and determine the optimal integration time of the lock-in amplifiers.

To determine the Allan deviation of the system, we measured the sensing signal and the reference signal over a period of \(~3\) hours. During the measurement the integrating cylinder is constantly purged with pure nitrogen at a flow rate of 2L/min. Figure 4.17(a) shows the signals at the two arms as a function of time, the time interval (i.e. the integration time of the lock-in amplifier) between two consecutive data points is 100ms. One can see that both the sensing signal and
the reference signal are relatively stable, except that some (common) sharp peaks (∼ 0.2 mV) are present occasionally. This is probably due to the vibrations of the setup (especially the fiber tip) caused by opening/closing the doors when people are entering/exiting the laboratory. The ratio of the sensing signal and the reference signal is plotted in Figure 4.17(b), which represents the normalized absorbance signal $S_{\text{norm}}$ at zero CO$_2$. We see the sharp peak variations occurring at the same time with the sensing/reference signals, which means the common mode noise cannot be completely eliminated due to the imperfect matching between the two arms. Furthermore, large scale variations over long time period in the normalized signal is visible, which can be attributed to the source power fluctuations/drift (the temperature stability of the source is ±15 K), or the temperature change in the laboratory.

Using these data, the Allan deviations of the sensing/reference signals as well as the normalized signal are calculated and plotted in Figure 4.18. One can see that as expected, the Allan deviation first decreases as the integration time increases. The minimum deviation of the normalized signal is about $2 \times 10^{-4}$ AU, which is also the detection limit of the system ($1\sigma$). The Allan deviation increases at longer integration time intervals, which is due to the drift of the source and/or the changes in the detector performance. The optimal integration time (i.e. when Allan deviation of the normalized signal is at its minimum) is approximately 2
seconds. In all following measurements, we use an integration time of 1 second, which is restricted by the limited resolution of the time constant of the lock-in amplifier. From Figure 4.18, it is also possible to compare the detection limit when only the sensing arm is used and when both arms are used. The minimum deviation of the sensing signal is around $6 \times 10^{-3} mV$ (at 10s integration time), and its mean is about $10 mV$ (from Figure 4.17), resulting a detection limit of $6 \times 10^{-3} mV/10 mV = 6 \times 10^{-4}$ AU. With the addition of the reference arm, the detection limit is reduced to $2 \times 10^{-4}/1 = 2 \times 10^{-4}$, which is 3 times smaller than that just using the sensing arm.

### 4.6.2 CO₂ sensing measurement

The response of the integrating cylinder to a change of CO₂ concentration is measured by feeding the integrating cylinder with a series of sample gases containing various concentrations of CO₂. At the beginning of the measurement, the integrating cylinder is purged with pure N₂ to determine the baseline of the normalized signal ($I_{R0}/I_{A0}$). During the measurement, the sensing signal $I_A$ and the reference signal $I_R$ are recorded for a period of 5 min at each concentration step. Figure 4.19(a) shows the two signals as the CO₂ concentration is varied. The letters from A to L correspond to different CO₂ concentrations which are listed on the right side of the figure. It can be seen that the reference signal stays relatively stable while the sensing signal decreases when the CO₂ concentration increases. Furthermore, we can see that the fringes on the two signals are strongly correlated, which is due to the common mode noise on both arms. These fringes can be eliminated by normalizing the sensing signal w.r.t. the reference signal $I_A/I_R$, which is shown in Figure 4.19(b). The normalized absorbance $S_{norm}$ for each CO₂ concentration can be calculated using Eq.4.27, with the correction factor $I_{R0}/I_{A0}$ determined at zero CO₂ concentration. Figure 4.20 shows the calculated normalized absorbance at different CO₂ concentration levels. The absorbance is also simulated by propagating the spectrum of the source (shown in Figure 4.15(a)) through the active filter shown in Figure 4.15(b)) and then through different concentrations of CO₂ (with absorption spectrum shown in Figure 4.15(b)). A path length of 3.5 cm (as obtained from Table 4.1) is used in this simulation. It can be seen that the agreement between the measurement and the simulation is reasonably good. The response of the integrating cylinder at 100 ppm CO₂ is shown in the inset of Figure 4.20, where 10 measurements are superimposed over one figure. One can see there is a clear and repeatable step response. The normalized absorbance of 100 ppm CO₂ is about $9 \times 10^{-4}$ AU, which is approximately three times of the signal deviation ($3 \cdot 2.5 \times 10^{-4} = 7.5 \times 10^{-4}$ AU) as obtained from the Allan deviation plot.
4.6.3 Response Time

The response time of a gas sensor is of great importance for applications such as leak detection in MAC systems, and real-time CO\textsubscript{2} concentration measurement can be possible with a fast response time. In industry, the response time is usually defined as the time needed for a sensor to reach 90% \((T_{90})\) of its total response when a step CO\textsubscript{2} concentration change is applied to the sensor. The response time of the integrating cylinder is measured using the same setup as shown in Figure 4.14. The integrating cylinder is first purged with a sample gas containing a high CO\textsubscript{2} concentration (50%), and the evolution of the sensing signal is recorded when the reading becomes stable. Then the gas flow is abruptly shut off and the CO\textsubscript{2} concentration in the integrating cylinder reaches ambient level (~400 ppm) due
Figure 4.20: Normalized absorbance of the integrating cylinder at different CO\textsubscript{2} concentrations. The response at CO\textsubscript{2} = 100 ppm is in the inset, with 10 measurements superimposed in one figure.

to gas diffusion. Figure 4.21 shows the sensing signal as a function of time. In this measurement we use an integration time of 100ms to resolve the step of the signal. We can see that the integrating cylinder has a response time (T\textsubscript{90}) of only 2.8 seconds, which is consistent with the simulation. The fast response time is primarily due to the small footprint of the integrating cylinder.

### 4.7 Conclusions

In this chapter, a detailed analysis and characterization of the integrating cylinder were presented. The integrating cylinder can be considered as a 2D version of the integrating sphere, implemented on a silicon chip. The multiple reflections of the incident light in the cylinder and the high-reflectivity mirror coated on the inner surfaces allow a long path length to be folded on a small chip area. The mathematical description of the integrating cylinder was given, showing that the sensitivity, defined as the transmission change due to CO\textsubscript{2} absorption, can be maximized for a certain cylinder radius \( R \) and access width \( d \) combination. 3D ray tracing simulation was performed to confirm the theoretical analysis. Simulations show that for an integrating cylinder with optimal design parameters (eg. radius \( R = 2\text{mm} \) and access waveguide width \( d = 200\mu\text{m} \)), the equivalent path length and the total loss are 3.5cm and 7.4dB, respectively. The gas diffusion in the integrating cylinder was also modeled with finite difference element method, yielding a response
time of 2.6 seconds of the integrating cylinder. The fabrication process flow of the integrating cylinder was introduced, involving DRIE, gold deposition, and wafer bonding. These processes are all wafer-level techniques that are widely used in the semiconductor industry, making it possible for the mass production of the integrating cylinder with extremely low cost. The fabricated integrating cylinder was characterized using a dedicated gas sensing setup. An Allan deviation measurement was performed to determine the noise and stability of the sensing system. It was shown that by using the reference arm to eliminate the common mode noise, a minimum transmission change of $\sim 2 \times 10^{-4}$ AU can be obtained using an lock-in integration time of 2 seconds. Further analysis revealed that this is $\sim 3 \times$ better than just using the sensing arm. A CO$_2$ sensing measurement using the integrating cylinder was performed, showing a detection limit of $\sim 100$ ppm, which is consistent with the detection limit predicted by the Allan deviation plot. The detection limit is primarily limited by the drift of the source and the imperfect matching between the two arms. The response time of the integrating cylinder was measured by abruptly shutting off the high-concentration CO$_2$ gas flow. The response time of the integrating cylinder is approximately 3 seconds, which is attributed to the small footprint of the integrating cylinder.
References


[13] Shinichi Tachi, Kazunori Tsujimoto, and Sadayuki Okudaira. Low-


5

Fully Integrated NDIR CO$_2$ Sensor

5.1 Introduction

In Chapter 4, the integrating cylinder has been investigated both numerically and experimentally. A fully integrated NDIR CO$_2$ sensor based on the integrating cylinder will be explored in this chapter. For the fully integrated sensor, we use a mid-IR LED and photodiode (PD) optopair as the optical source and detector. The LED-PD optopair provides the advantages of a large overlap between the emission spectrum of the LED and the absorption spectrum of CO$_2$. The responsivity of the LED and photodiode is higher than its thermal counterparts (e.g. a micro-light bulb and a thermopile detector), thus a lower limit of detection can be potentially achieved. In what follows, the path towards a fully integrated NDIR CO$_2$ sensor will be elaborated. First, 3D ray tracing simulations of the coupling efficiencies from the LED to the input waveguide of the integrating cylinder and from the output waveguide to the photodiodes are discussed. The fabrication process of the fully integrated sensor will be discussed in detail, including the integration process of the LED/PD on the bottom silicon substrate, as well as the integration process of the bottom substrate (containing the LED and PDs) and the top substrate (containing the integrating cylinder). Finally, the fabricated sensor will be evaluated experimentally. Experimental results on CO$_2$ sensing, response time, cross-sensitivity to water vapor, and long term stability will be presented.
5.2 LED and photodiode optopairs

An LED-photodiode optopair consists of an LED and a photodiode that are spectrally matched. Traditionally, NDIR sensors are based on thermal light sources and broadband detectors, together with narrow band optical filters with the pass band overlapping with the absorption band of the analyte of interest. The development of mid-IR LEDs and photodiodes has advanced optical sensing dramatically, due to their compactness and better power efficiency. The use of an optopair eliminates the need for mechanical modulators and optical filters, which allows to reduce the cost of the sensor system. LED/PD combinations have been widely used for sensing applications such as in CH$_4$ sensing at 3.4µm [1], CO sensing at 4.6µm [2], NO$_2$ sensing at 6.3µm [3, 4], and CO$_2$ sensing at 4.25µm [5–7]. These sensors consist of packaged LED/PD chips and co-assembled discrete gas cells, which provides a rugged construction and robust operation. However, it also compromises the cost and compactness of the sensor.

5.2.1 Mid-IR III-V LEDs

The advancement of the mid-IR LEDs is primarily driven by the growing need for low-cost optical sensors in the mid-IR spectral range, which contains strong absorption bands of many gas species as shown in Figure 2.1. Compared to thermal optical sources, mid-IR LEDs offer higher power spectral density, and they are superior to laser-based systems in terms of cost and complexity. As already discussed in Chapter 2, the emission wavelength of the III-V mid-IR LEDs can be matched to the absorption band of the analyte by engineering the composition of the III-V semiconductor layer stack. Mid-IR LEDs emitting at 4µm wavelength range typically use antimony-based material systems (e.g. InAsSbP, InGaAsSb and InAsSb). The material of the active layer can be ternary (e.g. InAsSb [8]), quaternary (e.g. InAsSbP [9]), or quintenary (e.g. AlGaInAsSb [6]) semiconductor alloys. The use of multi-element alloys provides additional degrees of freedom to control the electro-physical properties, which offers a new approach to control the optical confinement and carrier leakage [6, 9].

However, the narrow band gap ($0.29 eV$ for 4.3µm) of these semiconductors imposes limitations on the efficiency of the LED. It has been demonstrated that the Auger recombination increases when the band gap of the III-V semiconductor is decreased [10, 11], which limits the internal efficiency of the mid-IR LEDs. Moreover, current crowding effects are more pronounced for LEDs working at longer wavelengths. To minimize the overlap between the top metal contact and the top

$^1$Auger recombination is a non-radiative process involving three carriers. When an electron and a hole recombine, instead of emitting a photon, the energy is transferred to a third particle: either an electron is raised higher into the conduction band or a hole is pushed deeper into the valence band, its energy is ultimately lost to thermal vibrations.
surface (hence maximizing the extraction efficiency), a top-emitting mid-IR LED is usually fabricated as a planar mesa with lateral current injection geometry. The lateral current injection creates current crowding near the metal contact edge, thus the effective emitting area is decreased. This effect is more pronounced in LEDs emitting at longer wavelengths due to the low junction resistance of narrow band gap materials [12, 13]. Moreover, the local heating and the temperature gradient generated by the current crowding effect increases the Auger recombination and thus the internal efficiency of the LED is decreased [14, 15]. For the reasons mentioned above, the internal efficiency of a mid-IR LED is only a few percent and continuous wave (CW) operation is often forbidden. In addition to the low internal efficiencies, mid-IR LEDs also suffer from low extraction efficiencies. The high refractive indices of the narrow band gap III-V materials ($n > 3.5$) leads to a small critical angle and therefore the numerical aperture for light to escape is limited. The extraction efficiency from one side of the chip can be approximated by 
$$
\eta_{\text{ext}} = \frac{1}{n(n+1)^2} \quad [16],
$$
which is only about 1.4% for a III-V material with $n = 3.5$. As both the internal efficiency and the extraction efficiency are relatively low, the output power from a typical mid-IR LED is only on the order of 10$\mu$W.

5.2.1.1 LED used in this thesis work

In this thesis work, we use the LED bare chips provided by IBSG [17]. A picture of the LED bare chips and the layer structure are shown in Figure 5.1. The LED chips have a top ring contact and a planar bottom contact, with which the chips can be integrated on the silicon substrate with flip-chip bonding. The main specifications of the LED are summarized in Table 5.1, in which QCW stands for Quasi-CW mode of the LED (repetition rate = 0.5 kHz, pulse duration = 1 ms, duty cycle = 50%).

<table>
<thead>
<tr>
<th>L, W, H [µm]</th>
<th>Emitting area [µm]</th>
<th>$\lambda_{\text{peak}}$ [µm]</th>
<th>FWHM [µm]</th>
<th>$P_{\text{QCW}}$ [µW]</th>
</tr>
</thead>
<tbody>
<tr>
<td>400,400,250</td>
<td>300x300</td>
<td>4.15</td>
<td>0.8</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 5.1: Major specifications of the LED used in this work.

5.2.2 Mid-IR photodiode

5.2.2.1 Performance metrics

Responsivity and quantum efficiency

The responsivity of the photodiode is a quantity to measure the effectiveness of the conversion of light power into electrical current, and it is defined as the ratio of the photocurrent $I_P$ to the optical power $P$ of the incident light at a given wavelength:

$$
R_\lambda = \frac{I_P}{P}[A/W] \quad (5.1)
$$
The responsivity varies with the wavelength of the incident light as well as the applied bias voltage and temperature. Responsivity increases slightly with the applied reverse bias voltage due to improved charge collection efficiency in the photodiode, while the temperature dependence of the responsivity is due to the change in band-gap with temperature variations. The quantum efficiency $\eta$ of the photodiode is defined as the fraction of the incident photons that contribute to electrons in the photocurrent, it is related to responsivity by:

$$\eta = 1.24 \frac{R_{\lambda}}{\lambda}$$

with $R_{\lambda}$ being the responsivity in $A/W$ and $\lambda$ being the wavelength in $\mu m$.

**Noise equivalent power**

The noise equivalent power (NEP) is defined as the input optical power that results in a signal-to-noise ratio (SNR) of 1 in a 1 Hz bandwidth. It is a measure of the minimal detectable power of the photodiode, i.e., the weakest optical signal that can be detected. The NEP is defined as:

$$NEP = \frac{I_{tn}}{R_{\lambda}} [W/\sqrt{Hz}]$$

where $I_{tn}$ is the total noise current of the photodiode. The minimum detectable optical power change in a certain bandwidth ($BW$) can therefore be calculated by:

$$P_{min} = NEP \cdot \sqrt{BW}$$

**Specific detectivity**
To compare the sensitivities of photodiodes with different areas, the specific detectivity $D^*$ is often used. $D^*$ is defined as the photo sensitivity per unit area of the photodiode, and it can be derived from the NEP by:

$$D^* = \frac{\sqrt{A}}{\text{NEP} \left[ \frac{\text{cm} \sqrt{\text{Hz}}}{\text{W}} \right]}$$

(5.5)

where $A$ is the active area of the photodiode.

### 5.2.2.2 Photodiode equivalent circuit

From an electronics point of view, the photodiode can be modeled by the equivalent circuit as shown in Figure 5.2. The photocurrent $I_{ph}$ is the current generated by the incident light. Apart from this photocurrent, a dark current $I_d$ is also generated by the thermal excitation of carriers in the absence of light. The magnitude of the dark current depends not only on the dimensions of the active area (a larger active area gives a larger dark current), but also on the working temperature of the photodiode: a higher temperature leads to a larger dark current. The dark current is virtually zero when the photodiode is not biased. Moreover, the dark current in infrared photodiodes are typically larger than their visible and near-infrared counterparts, as the carrier excitation requires less energy for smaller band gaps. For instance, the dark current of the mid-IR photodiode used in this thesis work is 400 $\mu$A (bias voltage = -0.5 V, $T = 20^\circ$) [18], while the typical dark current for a photodiode with peak responsivity at 1.5 $\mu$m is only 0.1 nA (bias voltage = -5 V, $T = 25^\circ$) [19]. This large dark current of mid-IR photodiodes imposes a disturbing shot noise current and also increases the noise floor.

Shunt resistance $R_{sh}$ is another important parameter of the photodiode, and it is defined as the resistance of the p-n junction when the photodiode is zero-biased. The shunt resistance can be determined by calculating the first derivative of the bias voltage w.r.t the current for a small bias voltage:

$$R_{sh} = \frac{\delta V}{\delta I} |_{V \rightarrow 0}$$

(5.6)

The actual values of shunt resistance range from several $\Omega$ to a few $G\Omega$, depending on the operating wavelength range. Due to the small band gap of the semiconductor materials in the mid-IR photodiodes, the shunt resistance is typically on the order of tens of $\Omega$, which is many orders of magnitude smaller than the photodiodes working at visible or near-infrared wavelength range [20]. This relatively small shunt resistance produces a large Johnson noise (as per Eq. 5.9), thus limiting its sensitivity, and it also imposes challenges on the design of the amplifiers [20].

Apart from the resistive characteristics of the p-n junction, it also has a parasitic junction capacitance $C_j$: a capacitor is formed by the p- and n-region of the photodiode, separated by the depletion region. The presence of the junction capacitance
limits the response speed of the photodiode by the resistance-capacitance (RC) time constant, and the bandwidth of the photodiode is given by:

\[ f_{BW} = \frac{1}{2\pi R_{tot}C_j} \]  

(5.7)

where \( R_{tot} \) is the summation of the series resistance and the load resistance. The series resistance \( R_s \) comes from the contact resistance between the metal electrodes and the photodiode contact layers. It can also come from electrical connections such as wire bonds.

### 5.2.2.3 Noise characteristics

In a photodiode, the limit of light detection is determined by its noise characteristics. The noise of a photodiode takes three forms: \(1/f\) noise, shot noise and Johnson noise.

#### \(1/f\) noise

The mechanisms that produce \(1/f\) noise are poorly understood. The noise power of \(1/f\) noise is inversely proportional to the modulation frequency \(f\), and its magnitude is typically much smaller than shot noise except at very low modulation frequencies (lower than a few hundred Hz). To reduce the \(1/f\) noise, the photodiode should work at a reasonably high frequency, usually as high as a few KHz, which should reduce the contribution of \(1/f\) noise to a small amount.

#### Shot noise

Shot noise is associated with the particle nature of light and of electrical current, and it is related to the fluctuations in both the photocurrent and the dark current, resulting from the statistical uncertainty in photon arrival rate and electron generation rate. The magnitude of the shot noise is expressed as the root mean square
(rms) noise current:

\[ I_{sn} = \sqrt{2q(I_P + I_D)BW} \] (5.8)

where \( q \) is the electron charge, \( I_P \) and \( I_D \) are the photo current and dark current, respectively, and \( BW \) is the noise measurement bandwidth.

**Johnson noise**

Johnson noise, also known as thermal noise, originates from the thermal generation of charge carriers. All resistive materials have Johnson noise associated with them, and the magnitude of the Johnson noise current is given by:

\[ I_{jn} = \sqrt{\frac{4k_B T BW}{R_{sh}}} \] (5.9)

in which \( k_B \) is the Boltzmann constant, \( T \) is the absolute temperature, and \( R_{sh} \) is the shunt resistance of the photodiode. The total noise current (in Eq. 5.3) generated in a photodiode is then given by:

\[ I_{tn} = \sqrt{I_{sn}^2 + I_{jn}^2} \] (5.10)

### 5.2.2.4 Photodiode operation mode

As shown in Figure 5.3, a photodiode usually operates either in the photoconductive mode, or in the photovoltaic mode. In photoconductive mode, the photodiode is reverse-biased, which increases the width of the depletion region and therefore the junction capacitance is decreased. The application of a reverse bias can greatly improve the response speed (bandwidth) and linearity of the photodiode, thus the photoconductive mode is often used for high speed applications. Shot noise is the dominating noise source when the photodiode operates in photoconductive mode, especially for mid-IR photodiodes. When a reverse bias voltage is applied, the relatively large dark current generates a high shot noise as per Eq. 5.8 making low-level light detection unsuitable. In photovoltaic mode, a zero bias voltage is applied across the photodiode, which minimizes the dark and the noise currents. Therefore it’s preferred for ultra low-level light detection and low frequency operation. Johnson noise is the dominant noise in photovoltaic operation mode. For mid-IR photodiodes, the Johnson noise is more pronounced as the shunt resistance is typically only a few tens of Ohms, due to the small band gaps of the semiconductor materials. To determine the optimal operation mode (photoconductive vs. photovoltaic) of the photodiode, the magnitude of the shot noise, Johnson noise, and the total noise current of the photodiode at both operation modes are calculated. In this calculation, we use the following parameters of the photodiode (also listed in the next section): shunt resistance of 30\( \Omega \), dark current of 2050\( \mu A \), and a responsivity of 2.5\( A/W \). We assume an optical power of 10\( nW \) for the incident light and set the bandwidth of the system to be 1Hz, and we assume the dark current at the photovoltaic mode is zero. The calculated noise contributions and the
total noise current when the photodiode works in the photoconductive and photovoltaic modes are listed in Table 5.2. It can be seen that the total noise current in the photoconductive mode is about 50% higher than that of the photovoltaic mode, therefore we choose the photovoltaic mode as the operation mode of the photodiode.

<table>
<thead>
<tr>
<th></th>
<th>photoconductive mode</th>
<th>photovoltaic mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>shot noise [A]</td>
<td>2.56e-11</td>
<td>8.95e-14</td>
</tr>
<tr>
<td>Johnson noise [A]</td>
<td>2.34e-11</td>
<td>2.34e-11</td>
</tr>
<tr>
<td>total noise [A]</td>
<td>3.5e-11</td>
<td>2.34e-11</td>
</tr>
</tbody>
</table>

Table 5.2: Noise contributions and the total noise current when the photodiode works at photoconductive and photovoltaic modes.

### 5.2.2.5 Photodiodes used in this thesis work

In this thesis work, two different photodiodes are used. For the sensing channel of the CO$_2$ sensor, mid-IR photodiode bare chips (PD42) provided by the Mid-IR Diode Optopair Group [23] are used. These photodiodes have a spectral response centered at 4.15µm. For the reference channel of the sensor, the spectral response of the photodiodes should not overlap with the absorption band of CO$_2$ at 4.2µm. In this regard we choose the photodiode bare chips (PD36) provided by IBSG [18]. The photodiodes have a spectral response with a cut-off wavelength at ~ 4µm. The dimensions (L, W, H) of the photodiodes at the sensing channel and the reference channel are both 500µm, 500µm, 200µm. Figure 5.4 shows the microscope images of the sensing and the reference photodiodes, the sensing photodiode PD42 has a grid top contact and the reference photodiode PD36 has a point top contact. Both the photodiodes have a planar bottom contact. The major specifications of the two photodiodes are summarized in Table 5.3.

The normalized spectral responses of the sensing photodiode and the reference
Fully Integrated On-Chip NDIR CO₂ Sensor

Figure 5.4: Microscope images of the sensing photodiode PD42 and the reference photodiode PD36.

Figure 5.5: Normalized emission spectrum of the LED [17] and spectral responses of the photodiodes [18, 23], with the absorption spectrum of CO₂ in the 3µm – 5µm wavelength range [24].

Photodiode are shown in Figure 5.5 with the emission spectrum of the LED and the absorption spectrum of CO₂ at 4.2µm interposed on top. It can be seen that the spectral response of the sensing photodiode (PD42) overlaps with the emission spectrum of the LED and the CO₂ absorption band, while there is no overlap between the spectral response of the reference photodiode (PD36) and the CO₂ absorption band.
Table 5.3: Specifications of the photodiodes used in this thesis. PD42BS: photodiode at the sensing channel. PD36-03: photodiode at the reference channel.
5.3 Estimation of coupling efficiencies

To estimate the coupling efficiency from the LED to the input waveguide of the integrating cylinder, and the coupling efficiency from the output waveguide to the photodiodes, similar 3D ray tracing simulations (in OpticStudio) as in section 3.2 are performed. The simulation setup is as the following:

- Simulation is performed using the non-sequential mode of OpticStudio, as it allows for multiple reflections of the light rays at a certain interface.

- The LED/PD sits in a metallized trench as shown in Figure 5.6, the trench can be achieved by the selective etching of silicon with KOH. We use access waveguides to couple light from the LED and to the photodiode, as it allows for the decoupling of the LED/photodiode dimensions and the fraction of power extracted from the cavity, therefore the sensitivity of the sensor can be optimized (as discussed in section 4.4).

- As the thickness of the substrate is much larger than that of the active layer, the active regions and the substrates of the LED/PDs are given a refractive index of 3.5, that of n-InAs \( n = 2 - 3 \times 10^3 \text{cm}^{-3} \). For the same reason, an absorption coefficient of 10/cm is applied to both the substrates and the active regions.

- A coating with the refractive index of gold (200nm) is applied on the trench as well as on the inner walls of the input/output waveguides.

- For the simulation of the LED coupling efficiency, a rectangular volume source emitting at 4.26\( \mu \text{m} \) is placed at the center of the LED, with isotropic emission in all directions. A rectangular power detector is placed at the cross section of the input waveguide to record the power coupled to the waveguide.

- For the simulation of the photodiode coupling efficiency, a point source with isotropic emission is placed at the center of the output waveguide (1mm away from the PD), a rectangular power detector is placed at the cross-section of the PD to record the power coupled to the PD.

In the simulation, a total number of 100,000 rays are emitted from the LED/point source with the optical power equally distributed in each ray. The coupling efficiency from the LED to the input waveguide is simulated to be about 8% (in one direction), and the coupling efficiency from the output waveguide to the PD is about 30%.
5.4 Full sensor fabrication

The schematic of the fully integrated NDIR CO\textsubscript{2} sensor is shown in Figure 5.7. The sensor is formed by wafer bonding of two silicon substrates. The top substrate contains the integrating cylinder, of which the fabrication process has been discussed thoroughly in Chapter 4. On the bottom substrate, the LED and photodiodes are integrated on-chip. The fabrication process of the bottom substrate starts with a blank silicon substrate. The main steps of the fabrication flow are summarized below as well as in Figure 5.8:

- A first optical lithography and silicon etching step is performed, with the
purpose to define the outer boundary of the bottom substrate, and to define the alignment markers for the following processes.

• A thin layer of SiO$_2$/SiN$_x$ (300 nm/400 nm) is deposited on both sides of the substrate. The SiN$_x$ layer acts as the hard mask for silicon etching. Three rectangles are then opened for selective silicon etching with a second lithography and dry etching step.

• Three trenches are formed on the substrate by selective etching of silicon with KOH, as shown in Figure 5.8 (b). The etching depth is $\sim 250 \mu$m, and the widths of the trenches at the bottom surface are designed to be $\sim 50 \mu$m wider than the widths of the LED/PD.

• A thin layer of SiO$_2$ ($\sim 500$ nm) is deposited on the substrate to isolate the two contact pads of the LED/PD (see later), as the silicon substrate (p-doped) we use is weakly conductive.

• Gold deposition and liftoff. In this step, the gold reflectors on the bottom substrate are formed (with Ti as adhesive layer), as well as the two contact pads (anode and cathode) for the LED/PD. As shown in Figure 5.8 (c), both contact pads need to be isolated from the large-area gold reflector on the substrate. Therefore, the inclined sidewalls of the trenches need to be protected from gold deposition. In this context, we used a resist photopolymer film (FX900 series, from DuPont) as the protective layer. The large thickness of the resist film ($\sim 20 \mu$m thick) allows it to hang over the trenches. Figure 5.9 (a) shows the patterned resist film after lithography, and the contact pads after gold deposition and lift-off are shown in Figure 5.9 (b).

• LED/PD bonding. In this step, the LED/PD bare chips are bonded (with top sides facing up) in the trenches by flip-chip bonding. Due to the high roughness of the gold layer on the bottom of the trench and the bottom contacts of the LED/PD (formed by electroplating of gold), we use a conductive silver adhesive (478SS, from Electron Microscopy Sciences) as bonding agent. After bonding, the adhesive is then cured at $160^\circ$ for 30 minutes.

• Wire bonding of the LED/PD. In this step, the top contact pads of the LED/PD are wire-bonded to the anode pads on the silicon substrate, Figure 5.9 (c) shows a microscope image of the LED bare chip with wire bond.

• Finally, the sensor is formed by wafer bonding of the top substrate and the bottom substrate. The bonding is performed by gold-to-gold thermocompression bonding (at $160^\circ$, 1MPa, for 30 minutes).
Figure 5.8: Fabrication process flow of the bottom substrate containing the LED/PD chips. 
(a) Blank silicon substrate. (b) Three trenches are formed on the substrate via selective etching of silicon by KOH. (c) Gold deposition and lift-off. In this step, two contact pads (anode and cathode) are formed for the LED/PD, the deposited gold on the bottom substrate acts as a reflector for the integrating cylinder. (d) LED/PD bonding and wire bonding. The LED/PD chips are bonded in the trench using flip-chip bonding, after which the top contacts of the LED/PD are wire-bonded to the anode pads.

5.5 Trans-impedance amplifier design

When designing the trans-impedance amplifiers (TIA) for the mid-IR photodiodes, one has to take into account the low shunt resistance as it can introduce significant noise currents. Due to the relatively large overall loss of the system, the optical power reaching the photodiodes is very low (on the order of 10nW). For reasons as described earlier, the photodiode should work in photovoltaic mode (un-biased) to detect such a weak signal. In photovoltaic mode, the Johnson noise is dominant, the spectral noise density \( i_{n} \) of the photodiode is given by:

\[
\sqrt{\frac{i_{n,PD}^2}{R_{sh}}} = \sqrt{\frac{4kT}{R_{sh}}} [A/\sqrt{Hz}] \tag{5.11}
\]
The TIA for the photodiode consists of a current-voltage converter, in which the anode and cathode of the photodiode are connected to the inverting and non-inverting inputs of the op-amp, respectively. The photo current $I_P$ from the photodiode is then amplified by a feedback resistor $R_f$, which also has a Johnson noise given by:

$$\sqrt{i_{n,R_f}^2} = \sqrt{\frac{4kT}{R_f}} \left[ A/\sqrt{Hz} \right]$$ (5.12)

In the TIA circuit, the op-amps themselves exhibit both a voltage noise $\sqrt{e_{n,OP}^2}$ and a current noise $\sqrt{i_{n,OP}^2}$ at the input side. The available low-noise op-amps can be divided into two categories [20]:

- op-amps with entrance circuit based on bipolar transistors, the voltage noise for these op-amps is typically very low ($< 1 \text{nV}/\sqrt{Hz}$), but the current noise is significant (up to a few $\text{pA}/\sqrt{Hz}$);
- op-amps with entrance circuit based on field-effect transistors, the current noise for these op-amps is low (a few $\text{fA}/\sqrt{Hz}$), but the voltage noise is high (up to a few $\text{nV}/\sqrt{Hz}$);

To determine which type of op-amps can be used in the TIA circuit, the total noise current at the input of the op-amp can be calculated by:

$$i_{n,\Sigma} = \sqrt{i_{n,PD}^2 + i_{n,OP}^2 + i_{n,R_f}^2} = \sqrt{\frac{4kT}{R_{sh}} + i_{n,OP}^2 + \left( \frac{e_{n,OP}}{R_{sh}} \right)^2 + \frac{4kT}{R_f}}$$ (5.13)

in which the first and the last terms under the square root are the noise currents generated by the shunt resistor and the feedback resistor, respectively. As the feedback resistor $R_f$ is usually much larger (up to $M\Omega$) than the shunt resistance $R_{sh}$ of the mid-IR photodiodes (typically on the order of tens of $\Omega$), the last term in
Eq. 5.15 can be ignored. The second and the third terms are the noise contributions from the op-amp. It can be seen that the voltage noise of the op-amp is scaled by a factor of $1/R_{sh}$. For small $R_{sh}$, the noise contribution from the voltage noise of the op-amps is dominant. It has been demonstrated by numerical calculations that for photodiodes with $R_{sh} < 3k\Omega$, op-amps with low voltage noise is the preferred choice to minimize the total noise current [20]. Therefore, in the TIA design for both the sensing photodiode and the reference photodiode we use the op-amp LT1028 (from Analog Devices), which has a low input noise density of only $e_{n,OP} = 0.85nV/\sqrt{Hz}$.

The circuit of the TIA for the sensing photodiode (PD42) is schematically shown in Figure 5.10(a). The LED works at the Quasi-CW (QCW) mode, with a driving current of 200mA (pulse width = 1ms, duty cycle = 50%). The TIA is designed to amplify a photocurrent of 20nA to a voltage of 20mV, with a transimpedance gain of $10^6\Omega$, the bandwidth of the TIA is 2kHz. For the TIA of the reference photodiode (PD36), we set the gain to be $5 \cdot 10^6\Omega$, considering the smaller overlap between the spectral response of the reference photodiode and the emission spectrum of the LED (see Figure 5.5). The two TIAs are otherwise identical. An electronic circuit simulator LTspice (Analog Devices) is used for the simulation of both TIAs. The simulated transient response and the frequency response are shown in Figure 5.10(b-c). The two TIA circuit are then soldered on a printed circuit board (PCB), after which the fabricated sensor chip is mounted on the PCB and the contact pads of the LED/PDs are wire-bonded to the contact pads on the PCB. The PCB consisting of two TIAs is shown in Figure 5.11 with the sensor mounted and wire-bonded on it.

5.6 Experimental results

As in Chapter 4, we use the normalized absorbance $S_{norm}$ as the figure of merit to evaluate the performance of the full sensor. In the full sensor, we use an integrating cylinder with a radius of 2.6mm, and the input/output waveguide widths are 400µm. Such an integrating cylinder yields an equivalent path length of $\sim 3.4cm$ and a total loss of $\sim 7.2dB$ as shown in section 4.3. The LED is driven by a pulsed current source (LDP-3811, ILX Lightwave), and the output signals from the TIAs are fed into two lock-in amplifiers (SR380, Stanford Research Systems). The two lock-in amplifiers are triggered by the current source, and they are synchronized with identical settings during the measurement.

5.6.1 Allan deviation

As in Chapter 4, to determine the optimal integration time of the lock-in amplifiers and the limit of detection of the sensor system, an Allan deviation (with definition
given in Chapter 4) measurement is performed. The sensing signal and the reference signal are measured over a period of $\sim 3$ hours (with a time interval of 100 ms). During the measurement, the integrating cylinder is constantly flushed with pure N$_2$. Figure 5.12 shows the signals at the two arms as a function of time. It can be seen that the amplitudes of the sensing and the reference signals are about 25mV and 15mV, respectively. To compare the signal strengths with the theoretical predictions, we can calculate the strength of the sensing signal (as the strength of the reference signal is more difficult to calculate due to the uncertainties in the spectral overlap) using the specifications of the LED (optical power $P_{LED}$ from Table 5.1) and the sensing photodiode (peak responsivity $R_{PD}$ from Table 5.3), the simulated coupling efficiencies ($\eta_{LED}$ and $\eta_{PD}$, from section 5.3), as well as the simulated loss (or transmission $T_{IC}$) of the integrating cylinder (7.8dB for the cylinder used in this experiment, from Figure 4.2) and the transimpedance amplifier gain ($G = 10^6 \Omega$). The strength of the sensing signal ($I_{sens}$) can be calculated...
Figure 5.11: PCB consisting of the TIAs for the sensing photodiode and the reference photodiode. The fabricated sensor chip is mounted on the PCB and the contacts pads of the LED/PDs are connected to the board with wire bonds.

by:

\[ I_{\text{sens}} = P_{\text{LED}} \cdot \eta_{\text{LED}} \cdot T_{\text{IC}} \cdot \eta_{\text{PD}} \cdot R_{\text{PD}} \cdot G \]
\[ = 10 \mu W \cdot 0.08 \cdot 0.15 \cdot 0.3 \cdot 2.5 \text{A/W} \cdot 10^6 \Omega \]
\[ = 90 \text{mV} \] (5.14)

As we used the peak responsivity \( R_{PD} \) instead of the 'average' responsivity (data not available) of the photodiode in the calculation, the predicted signal strength at the sensing arm is about \( 90 \text{mV} / 25 \text{mV} = 3.6 \) times larger than the measured value. Other factors such as deviations in the LED output power, the feedback resistance, and the coupling efficiencies can also contribute to the difference between the prediction and the measurement. One can see in the measurement that although the gain of the reference TIA is \( 5 \times \) larger than that of the sensing TIA, the reference signal is still smaller than the sensing signal. This is because the spectral overlap between the response of the reference PD (PD36) and the emission spectrum of the LED is much smaller than that of the sensing PD (PD42) and the LED, as can be seen in Figure 5.5. Moreover, the noise (or deviation) of the reference signal is larger than the noise of the sensing signal, which is due to the difference in the trans-impedance gain. The noise present on the signals can be classified into two types: (1) correlated noise, the correlated noise is the common mode noise present on both arms, and it can originate from factors such as power fluctuations of the LED; (2) un-correlated noise, one example of this type of noise is the noise due to the performance variations of the two PDs (e.g., due to temperature fluctuations), the noises produced by the two PDs can be uncorrelated as
Figure 5.12: The sensing signal and the reference signal measured over a period of 3 hours, the measurement interval is 100 ms.

The two PDs are essentially different and they are placed far apart (~ 5 mm). The un-correlated noise can be minimized by signal averaging. To eliminate the correlated noise present on both arms, the following data processing procedure can be applied: (1) the reference signal is scaled by a factor of 1/5, which is ratio of the gain of the TIAs at the sensing arm and the reference arm; (2) an offset is added on the scaled reference signal such that both signals have the same mean values; (3) the sensing signal is then normalized w.r.t the corrected reference signal. The Allan deviations of the sensing signal, the (uncorrected) reference signal, and the normalized signal are calculated and plotted in Figure 5.13. One can see that the minimum deviation of the normalized signal is about $1.2 \times 10^{-3}$ AU, which is also the detection limit of the system ($1\sigma$). The optimal integration time of the lock-in amplifiers (i.e. when Allan deviation of the normalized signal is at its minimum) is approximately 30 seconds.

### 5.6.2 CO$_2$ sensing measurement

The response of the fully integrated sensor to CO$_2$ is measured by purging the sensor with a series of sample gases containing various concentrations of CO$_2$. In each concentration step, the sensor is first flushed with CO$_2$ and then with pure N$_2$. 
Figure 5.13: (a) Allan deviations of the sensing signal, the reference signal and the normalized signal.

The evolution of the two signals when the \( \text{CO}_2 \) concentration is varied is shown in Figure 5.14 (a). The letters from A to I correspond to different \( \text{CO}_2 \) concentrations which are listed on the right side of the figure, the signals are measured with an integration time of 30 seconds. It can be seen that for the sensing signal, a step response to \( \text{CO}_2 \) concentration changes is present, while the reference signal stays relatively stable, which is due to the lack of spectral overlap between the spectral response of the reference photodiode and the \( \text{CO}_2 \) absorption spectrum. To eliminate the common mode noise present on both signals, the sensing signal is normalized w.r.t the reference signal, and the results are shown in 5.14 (b). It can be seen that unlike the case in the ‘passive’ measurement of the integrating cylinder presented in Chapter 4, where the deviation of the normalized signal is relatively small (as shown in Figure 4.19(b)), the normalized signal for the full sensor is still quite noisy. This is because in the fully integrated sensor case, two different photodiodes and TIAs are used, thus the sensing and the reference arms are not balanced. Moreover, the sensing signal and the reference signal have opposite temperature coefficients, which will become clear in section 5.6.3. This opposite temperature coefficients make the normalized signal less immune to the low-frequency noise due to temperature fluctuations.

The normalized absorbance \( S_{\text{norm}} \) for each \( \text{CO}_2 \) concentration step, calculated using Eq(4.27) is shown in Figure 5.15. The limit of detection of the full sensor
Figure 5.14: (a) Response of both the sensing signal and reference signal to CO$_2$ concentration steps, the signals are measured with an integration time of 30s. (b) Normalized transmission, obtained by dividing the sensing signal with the reference signal. The CO$_2$ concentration steps are listed on the right side.

is about 250 ppm (for 1σ), which is significantly larger than the limit of detection of the ‘passive’ integrating cylinder as presented in Chapter 4. The reasons for this deterioration are three-fold: (1) the bandwidth of the LED emission (FWHM = 800 nm) is larger than the bandwidth of the optical filter (FWHM = 500 nm) used in passive measurement of the integrating cylinder, which leads to a larger ‘dilution’ of the CO$_2$ absorbance over the bandwidth; (2) for the passive measurement of the integrating cylinder, an ultra-stable optical source is used, while there is no stabilization circuit for the LED of the full sensor. In principle the power fluctuations of the LED can be eliminated by normalizing the sensing signal w.r.t. the reference signal, however, the two arms are not perfectly balanced due to the differences in the sensing arm and the reference arm (PDs and TIAs). (3) the shunt resistance of the photodiodes in the full sensor are much smaller than that of the photodiodes used in the passive measurement (30Ω vs. 300kΩ), which gives a
Figure 5.15: Normalized absorbance of the full sensor at different CO$_2$ concentrations.

much larger noise at the input of the TIA circuit. In principle a longer signal averaging time can minimize this noise contribution, however, a long averaging time makes the photodiodes vulnerable to temperature fluctuations as the temperature coefficients of the two photodiodes are different. The temperature dependence of the photodiodes will be discussed in section 5.6.5.

The normalized absorbance can also be calculated in the same manner as in section 4.6.2, which can be mathematically described by the following formula:

$$S_{\text{norm}}(c) = 1 - \frac{I_A(c)}{I_{0A}(c=0)} = 1 - \frac{\int_{\lambda_1}^{\lambda_2} P(\lambda) \exp[-\alpha(\lambda) \cdot L \cdot c] R(\lambda) \, d\lambda}{\int_{\lambda_1}^{\lambda_2} P(\lambda) \, d\lambda}$$

(5.15)

where $I_A$ and $I_{0A}$ are the sensing signals with/without the presence of CO$_2$, respectively. The difference of the upper and lower bounds ($\lambda_1$ and $\lambda_2$) of the integral represents the wavelength range of interest. $P(\lambda)$ is the emission spectrum of the LED, $R(\lambda)$ is the responsivity function of the sensing photodiode. The term $\exp[-\alpha(\lambda) \cdot L \cdot c]$ represents the absorption of CO$_2$, with $\alpha(\lambda)$ being the absorption coefficient, and $L$ and $c$ being the equivalent path length and CO$_2$ concentration, respectively. It can be seen that the agreement between the measurement and the simulation is reasonably good.

5.6.3 Response time

The response time of the fully integrated sensor is measured in the same manner as in section 4.6.3, i.e., the sensor is first flushed with a sample gas containing 50%
CO₂, then the gas flow is abruptly shut off, after which the CO₂ concentration in the sensor reaches ambient level (≈400 ppm) due to gas diffusion. The evolution of the sensing signal is recorded when the reading becomes stable. The sensing signal as a function of time is shown in Figure 5.16; in this measurement we use an integration time of 100ms to resolve the step of the signal. It can be seen that the response time (T₉₀) of the fully integrated sensor is approximately 2.5 seconds. This relatively short response time is sufficient for the leak detection of high CO₂ concentrations (a few percents), where a fast response time is required to give a timely warning signal. However, one should note that a much longer integration time (e.g., 30s) is needed to detect ambient-level CO₂ concentrations (a few hundreds of ppm). This long integration time also makes the response time of the sensor much longer, although it is usually not a problem for applications such as air quality monitoring.

5.6.4 Water interference

The accuracy of NDIR gas sensors can deteriorate with the presence of water molecules, which largely depends on the relative humidity (RH) of the environment. Water vapor has a broad absorption in the infrared wavelength range, and this absorption can be significant at some specific wavelengths [26]. Depending on the spectral overlap between the absorption band of the target gas and the water molecules, NDIR gas sensors exhibit different water interference characteristics. The water interference for an NDIR CO₂ sensor is typically very small as
there is virtually no overlap between the water absorption spectrum and the CO\textsubscript{2} absorption band at 4.2\textmu m. Figure 5.17 shows the absorption spectra of water vapor and CO\textsubscript{2}, as well as the emission/responsivity spectra of the LED/PDs in the 3.0\textmu m – 5.5\textmu m wavelength range. It can be seen that the overlaps between the water absorption and the emission spectrum of the LED/spectral responses of the two PDs are negligible. To experimentally evaluate the response of the sensor under different RH levels, we introduce a third gas containing pure N\textsubscript{2} with 100% RH (with flow rate of $F_{N_2,H_2O}$), obtained by flushing pure N\textsubscript{2} through a bottle with liquid water. Then sample gases with different RH levels can be generated by adjusting the flow rates of the three gases:

$$RH = \frac{F_{N_2,H_2O}}{F_{N_2,H_2O} + F_{N_2} + F_{CO_2}} \times 100\%$$  \hspace{1cm} (5.16)

During the measurement, the CO\textsubscript{2} concentration is kept constant at 2500 ppm while the RH level is swept from 0 to 99%. Figure 5.18(a) shows the normalized signal against time. It can be seen that no notable water interference can be observed for the sensor, which means that the cross-sensitivity introduced by the water vapor is below the detection limit of the sensor. To determine the magnitude of the cross-sensitivity introduced by water vapor, numerical simulations are carried out. The simulation procedure is the same as in section 5.6.2: propagating the emission spectrum of the LED through the absorption spectra of water vapor (with various RH levels), after which the resulting spectrum is ”sampled” by the photodiode by calculating its spectral overlap with the spectral response of the sensing photodiode. The absorbance of the water vapor at various RH levels are plotted in

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure5.17}
\caption{Water vapor and CO\textsubscript{2} absorption spectra in 3.0\textmu m – 5.5\textmu m wavelength range, together with the emission/responsivity spectra of the LED/PDs. The amplitude of the water vapor absorption spectrum is magnified by a factor of 100.}
\end{figure}
Figure 5.18: (a) Water interference measurement. During the measurement, the RH level is swept from 0 to 99% while the CO\textsubscript{2} concentration is kept at 2500ppm. (b) Simulation of the water vapor absorption at various RH levels, the absorption of water is added on top of the absorption of 2500ppm CO\textsubscript{2}.

Figure 5.18(b). The baseline corresponds to the absorbance of 2500ppm CO\textsubscript{2}. It can be seen that the absorbance induced by water vapor is within 5e-4, which is much smaller than the detection limit of the sensor (1\sigma = 1.2e-3).

5.6.5 Long term stability

The stability of the sensor (also known as drift) deals with the degree to which the sensor’s characteristics remain constant over time. Changes in stability can be attributed to factors such as temperature fluctuations and component aging. To study the long-term stability of the sensor, we measured the response of the sensor in ambient for a period of \(\sim\) five days. Figure 5.19(a) shows both the sensing
and the reference signal of the sensor against time. It can be seen that during the measurement, both signals exhibit relatively large scale fluctuations (up to 15%), and the sensing signal and the reference signal change in opposite directions. It is believed that the changes in both signals are due to the temperature fluctuations in the laboratory. To test this hypothesis, the ambient temperature during the course of the measurement is also monitored with a thermopile detector (MLX90632, Melexis) and the temperature profile is plotted in Figure 5.19(b). One can see that there is a strong correlation between the ambient temperature and the sensing/reference signals. The opposite correlation between the two signals and the ambient temperature can be explained by the temperature-dependence of the LED and photodiode characteristics as shown in Figure 5.20. When temperature increases, the light intensity of the LED emission decreases and the spectrum slightly red-shifts. Meanwhile the response of the sensing photodiode also decreases and the spectrum also red-shifts, these decreases in both LED emission and photodiode response contribute to the decrease of the sensing signal when the temperature is increased. For the reference photodiode, the spectral response also red-shifts and therefore its spectral overlap with the LED emission spectrum increases, leading to an increase of the reference signal. At this stage of the fully integrated sensor, a strong temperature dependence of the sensor performance was observed, possible approaches to solve this issue will be discussed in Chapter 6.
Fully Integrated On-Chip NDIR CO$_2$ Sensor

5.7 Conclusions

In this chapter, the characterization of the fully integrated NDIR CO$_2$ sensor was presented. A mid-IR LED is used as the optical source, and two mid-IR photodiodes are used as detectors for the sensing and the reference arms. The fully integrated sensor is formed by wafer bonding of two silicon substrates. The top substrate contains the integrating cylinder and the LED and photodiodes are integrated on the bottom substrate. The integration procedure involves optical lithography, silicon etching, gold deposition/lift-off and flip-chip bonding. 3D ray tracing simulation was performed to estimate the coupling efficiencies from the LED to the input waveguide of the integrating cylinder and from the output waveguide to the photodiodes. A trans-impedance amplifier (TIA) was designed and fabricated in-house for signal amplification and readout. The design challenges of the TIAs for the photodiodes with low shunt resistance were addressed. The fabricated full sensor was characterized using a dedicated gas sensing setup. CO$_2$ sensing measurement of the fully integrated sensor was performed, showing a detection limit of $\sim 250$ppm (1$\sigma$). Compared to the passive measurement of the integrating cylinder presented in Chapter 4, the detection limit is deteriorated, the reasons behind this deterioration were discussed. The cross-sensitivity of the sensor to water vapor was studied both experimentally and numerically. No notable water interference was observed in the experimental characterizations. Numerical simulations reveal that the transmission change induced by water vapor absorption is much smaller than the detection limit of the sensor. A qualitative analysis on the long term stability of the sensor was performed by measuring the response of the sensor for a period of five days. The long term stability of the sensor is subject to the temperature fluctuations in the laboratory. A strong correlation between the sensing/reference signals and the ambient temperature was observed, due to the temperature dependence of the LED/PD characteristics.
References


Conclusions and Perspectives

6.1 Conclusions

The primary goal of this work was to pursue and develop a miniaturized, low-cost optical CO$_2$ sensor operating in the mid-IR wavelength range. The applications of such a CO$_2$ sensor are many-fold. The main drive for this research is leak detection in mobile air conditioning systems. In this case a detection limit of $\sim 1\%$ is sufficient, but the response time of the sensor has to be short to give a timely warning when the leak happens. Other applications include air quality monitoring, greenhouse farming, industrial process control, etc. The requirement on detection limit/resolution is more stringent (sub-100ppm) for these applications, while the response of the sensor can be slow. Different techniques exist today for CO$_2$ sensing, of which the NDIR technique stands out due to its advantages such as high sensitivity, low cross-sensitivity, and long-term stability. However, the current NDIR CO$_2$ sensors are bulky as a cm-scale path length is required to achieve a high sensitivity and they are also expensive as they are based on co-assembled discrete optical elements. To address these issues, three on-chip waveguide structures are proposed in this research: the multi-slot silicon waveguide, the hollow metal waveguide, and the hollow metal integrating cylinder. The multi-slot waveguide approach is conceptually novel, however, the fabrication of the waveguides is difficult and is currently not suitable for low-cost and high-volume production. The hollow metal waveguide and the integrating cylinder approaches use gold-coated waveguides for gas sensing, and they can both be implemented on a silicon chip.
with wafer-level fabrication processes such as deep reactive ion etching and wafer bonding. The integrating cylinder approach is superior to the hollow waveguide approach in terms of footprint of the sensor. Therefore the integrating cylinder approach was further explored in this work.

We started with the theoretical analysis of the integrating cylinder, and numerical simulations based on ray tracing were performed to determine the optimal parameters of the integrating cylinder. The equivalent path length and the total loss were also extracted from the simulations and they agree well with the predictions from the theoretical calculations. For the experimental demonstration of the integrating cylinder, we chose an integrating cylinder with a radius of 2mm and access widths of 200 µm, such a cavity yields an equivalent path length of 3.5cm and a total propagation loss of 7.8dB. The fabrication process of the integrating cylinder was discussed thoroughly, involving deep reactive ion etching, gold deposition and wafer bonding. The proof-of-concept demonstration of the integrating cylinder using an external optical source and detectors indicates a detection limit of 100 ppm and a response time of 3 seconds. Therefore we concluded that the integrating cylinder is a viable approach to achieve a sensitive and fast CO$_2$ sensor.

A fully integrated CO$_2$ sensor based on the integrating cylinder was then developed and investigated. A mid-IR LED and two mid-IR photodiode bare chips are used as the optical source and detectors for the sensor respectively. The coupling efficiencies from the LED to the hollow core waveguide ($\eta_{LED}$) and from the hollow core waveguide to photodetector ($\eta_{PD}$) were estimated by ray tracing simulations, resulting a coupling efficiency of $\eta_{LED} = 8\%$ and $\eta_{PD} = 30\%$. The fabrication process to integrate the LED and photodiodes on-chip was discussed in detail. The use of bare chips allows the integration of the LED/photodiodes with the integrating cylinder using CMOS/MEMS compatible technologies, leading to a fully integrated, miniaturized, low-cost NDIR CO$_2$ sensor. CO$_2$ sensing measurements were performed on the full sensor, the fully integrated sensor has a detection limit of $\sim 250$ppm ($1\sigma$). Comparing to the passive measurement of the integrating cylinder using external source and detectors, the detection limit/resolution of the full sensor was compromised due to the lower power spectral density, the imbalance between the sensing and the reference channels, and the different impacts of the ambient temperature on the two photodiodes. NDIR gas sensors are often subject to the interference of other gas species (especially water vapor), in this context the cross-sensitivity of water vapor was studied. The poor spectral overlap between the water absorption spectrum and the emission spectrum of the LED ensures that there are negligible water interference issues. Experimental results confirmed that there was no notable interference from water vapors, further numerical simulations showed that interference induced by water vapor is below the detection limit of our sensor. To investigate the long-term stability and drift of the sensor, the response of the sensor and the ambient temperature were recorded for a few days. The
long-term stability of the sensor was found to be strongly related to the ambient temperature. Opposite correlations between the sensing/reference signal and the ambient temperature were observed, due to the temperature-dependent characteristics of the LED and the two photodiodes. However, this temperature-dependent drift of the sensor can be suppressed by monitoring the chip temperature and use a look up table to compensate for the different sensor and reference PD response.

To conclude, a fully integrated NDIR CO$_2$ sensor based on an integrating cylinder was developed in this thesis work. The sensor shows a detection limit of 250ppm and a response time of 30 seconds (limited by signal averaging time). Comparing to the state-of-the-art CO$_2$ sensors on the market as listed in Table 1.1, our sensor has a higher detection limit (750ppm (3σ) vs. 50ppm) and a similar response time. Possible approaches to decrease the detection limit of our sensor will be discussed in the next section. While our sensor has a poorer detection limit, it is $\sim$ 10 times smaller than the commercial competitors (5mm vs. 5cm), and the expected cost of our sensor is much lower due to the wafer-level fabrication processes of our sensor. At this stage, the CO$_2$ sensor developed in this work is not yet ready for applications such as air quality monitoring, as it requires a detection limit of sub-100ppm. However, the relatively low cost and small footprint of the sensor are competitive for applications such as in automotive/industrial air conditioning systems and industrial process control, etc.

6.2 Perspectives

6.2.1 Improving the signal to noise ratio

For the prototype demonstration of the fully integrated sensor, the wire-bonding versions of the LED and photodiode bare chips are used. The top point contact and bottom planar contact of the bare chips allow easy access to the LED and photodiodes for driving and for signal readout, and the integration process of the chips is straightforward. However, the input/output waveguide of the integrating cylinder has to remain open for wire bonding of the bare chips. It is beneficial in terms of the response time of the sensor, as the three opened access waveguides allow a fast gas diffusion in and out of the integrating cylinder, leading to a short response time of the sensor. Nevertheless, both the coupling efficiencies from the LED to the input waveguide, and from the output waveguide to the photodiode, are reduced. Therefore, the wire bonding versions of the LED/photodiode bare chips are sub-optimal for the sensor in terms of signal to noise ratio. A better solution would be using the flip-chip version of the LED/photodiode chips with top emission/detection, which has both contacts at the bottom of the chips. Therefore, the three access waveguides can be closed and gas access ports can be defined elsewhere on the integrating cylinder, leading to an increase in the coupling efficiencies and the
signal to noise ratio. Further ray tracing simulations show that when both the input waveguide and the output waveguide are closed, the coupling efficiencies from the LED to the input waveguide ($\eta_{\text{LED}}$) and from the output waveguide to the photodiode ($\eta_{\text{PD}}$) can be increased by 30% ($\eta_{\text{LED}}$ increased from 8% to 10.4%) and 28% ($\eta_{\text{PD}}$ increased from 30% to 38%), respectively, which increases the signal to noise ratio by a factor of 1.67. During the search for the LED/photodiode chips, several suppliers were contacted, however, the flip-chip version of the bare chips are either not available from those suppliers, or a large order quantity is required for customization. Therefore the flip-chip version of the bare chips was not pursued in this work due to time and budget limitations. In future iterations of the sensor, the use of flip-chip LED/photodiode chips can be explored.

6.2.2 Temperature compensation

The measurement and analysis in section 5.6.5 give a qualitative study of the long term stability of the sensor. At this stage of the sensor development, no temperature control or compensation is implemented in the circuitry, leading to a large dependence of the sensor’s response to temperature fluctuations. The influence of temperature can be eliminated or minimized by two approaches: temperature control or temperature compensation. In the temperature control case, a thermoelectric temperature control element can be packaged together with the sensor, therefore the temperature of the sensor chip is fixed during the measurement. Yet, the addition of the temperature control elements increases the complexity and cost of the sensor, and the power consumption of the sensor will also increase as the active cooling is power hungry. In the temperature compensation case, the temperature of the sensor chip is constantly monitored by a temperature sensor, and this temperature information can be used for implementing temperature compensation. In future iterations of the CO$_2$ sensor developed in this PhD, such a temperature compensation circuit is envisioned. The compensation procedure of the temperature fluctuation is as follows: (1) after fabrication, the sensor is pre-calibrated under various CO$_2$ concentrations at different temperatures; (2) the temperature-dependent CO$_2$ response of the sensor is stored in memory, and the calibrated data acts as a 'lookup table' for field measurements; (3) in field measurements, the temperature of the sensor chip is monitored, and the response of the sensor is compared with the lookup table at that particular temperature and the target CO$_2$ concentration can be extracted.

6.2.3 The shape of the integrating cavity

In the previous chapters, a NDIR CO$_2$ sensor based on an integrating cylinder was primarily explored. The key benefit of an integrating cylinder is that the incident light is reflected multiple times inside the cavity such that a long optical path length
Conclusions and Perspectives

6.2.4 Gas membrane

The working environment of the gas sensor can be harsh, and the sensor can be contaminated by, e.g., dust particles if unprotected. The contamination may 'blind' the sensor and thus measurement error occurs. For commercial gas sensors, porous gas membranes are typically used to cover the diffusion window of the sensor. The gas membranes have microscopic pores with diameters of a few tens of micrometers, and they are effective dust filters that can keep out the dust without compromising the diffusion of the gas. Figure 6.1 shows two examples of the gas membranes used in commercial gas sensors.

![Gas membranes used in commercial gas sensors. The two gas sensors are CozIR-LP3 CO₂ sensor from Gas Sensing Solutions, and PAC3500 CO sensor from Draeger, respectively.](image)

6.2.5 LED driver, TIA, and lock-in amplifier

For the proof-of-concept demonstration of the full sensor, an external current source is used to drive the LED and the TIA circuit is soldered on a PCB with discrete electrical components. In future iterations, the LED driver circuit and the TIA circuit can be integrated onto an electrical integrated circuit (EIC) chip, which can be packaged together with the gas sensor, leading to a small overall footprint.
of the sensor. Moreover, the functionality of the lock-in amplifiers can also be real-
ized on-chip, by integrating the lock-in circuit on an application-specific integrated
circuit (ASIC) chip or using a micro-controller.