Wafer-level processing of ultralow-loss Si$_3$N$_4$

Marcello Girardi
Wafer-level processing of ultralow-loss Si$_3$N$_4$

MARCELLO GIRARDI

© Marcello Girardi, 2023

ISBN 978-91-7905-920-0
Doktorsavhandlingar vid Chalmers tekniska högskola,
Ny serie nr 5386.
ISSN 0346-718X

Photonics Laboratory
Department of Microtechnology and Nanoscience (MC2)
Chalmers University of Technology
SE-412 96 Gothenburg,
Sweden
Phone: +46 (0)31-772 10 00

Front cover illustration: Picture of a wafer fabricated in this work. The photonic integrated circuits are highlighted in false colors. Each color corresponds to one design.

Printed in Sweden by
Chalmers Digitaltryck
Chalmers Tekniska Högskola,
Wafer-level processing of ultralow-loss Si$_3$N$_4$

MARCELLO GIRARDI
Photonics Laboratory
Department of Microtechnology and Nanoscience (MC2)
Chalmers University of Technology

Abstract

Photonic integrated circuits (PICs) are devices fabricated on a planar wafer that allow light generation, processing, and detection. Photonic integration brings important advantages for scaling up the complexity and functionality of photonic systems and facilitates their mass deployment in areas where large volumes and compact solutions are needed, e.g., optical interconnects. Among the material platforms available, silicon nitride (Si$_3$N$_4$) displays excellent optical properties such as broadband transparency, moderately high refractive index, and relatively strong nonlinearities. Indeed, Si$_3$N$_4$ integrated waveguides display ultralow-loss (few decibels per meter), which enables efficient light processing and nonlinear optics. Moreover, Si$_3$N$_4$ is compatible with standard complementary metal oxide semiconductor (CMOS) processing techniques, which facilitates the manufacture scalability required by mass deployment of PICs. However, the selection of a single photonic platform sets limitations to the device functionalities due to the intrinsic properties of the material and the fundamental limitation of optical waveguiding. Multilayer integration of different platforms can overcome the limitations encountered in single-platform PICs.

This thesis presents the development of advanced techniques for the wafer-level manufacturing of ultralow-loss Si$_3$N$_4$ devices and approaches to enable their interface with active components like modulators and chip-scale comb sources (microcombs). The investigation covers the tailoring of a waveguide to the functionality required, the wafer-scale manufacturing of Si$_3$N$_4$, and how to overcome the limitations of a single platform on a wafer. These studies enable high-yield fabrication of microcombs, the integration of two Si$_3$N$_4$ platforms on the same wafer, and a strategy to efficiently couple to an integrated LiNbO$_3$ layer to expand the chip functionality and scale up the complexity of the PIC.

Keywords

silicon nitride, ultralow loss, waveguide, photonic integrated circuit, multilayer integration, microcombs
List of Publications

This thesis is based on the following publications:

[A] Xiaonan Hu, Marcello Girardi, Zhichao Ye, Pascual Muñoz, Anders Larsson and Victor Torres-Company, $Si_3N_4$ photonic integration platform at 1 $\mu$m for optical interconnects
Optics Express, Vol. 28 Issue 9 p. 13019-13031.


[C] ´Oskar B. Helgason, Marcello Girardi, Zhichao Ye, Fuchuan Lei, Jochen Schröder and Victor Torres-Company, Surpassing the nonlinear conversion efficiency of soliton microcombs
Accepted in Nature Photonics.

Submitted, under review in Optics express.

[E] Marcello Girardi, ´Oskar B. Helgason, Carmen H. López Ortega, Israel Rebolledo-Salgado and Victor Torres-Company, Superefficient microcombs at the wafer level
Submitted.
Related publications and conference contributions by the author, not included in the thesis

Journal papers:


Conference presentations and papers:


[P] Louw Roel van der Zon, Marcello Girardi, Victor Torres-Company, Pascual Muñoz, Daniel Pastor, *Compact 8-channel Loop-back AWG based Integrated Comb Processor* IEEE International Conference on Group IV Photonics GFP.


Patent:

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFM</td>
<td>atomic force microscopy</td>
</tr>
<tr>
<td>ASIC</td>
<td>application-specific integrated circuit</td>
</tr>
<tr>
<td>AWG</td>
<td>arrayed waveguide grating</td>
</tr>
<tr>
<td>BEOL</td>
<td>back end of the line of the CMOS manufacturing process</td>
</tr>
<tr>
<td>BER</td>
<td>bit error ratio</td>
</tr>
<tr>
<td>CMOS</td>
<td>complementary metal–oxide semiconductor</td>
</tr>
<tr>
<td>CMP</td>
<td>chemical mechanical polishing</td>
</tr>
<tr>
<td>CW</td>
<td>continuous wave</td>
</tr>
<tr>
<td>DC</td>
<td>data center</td>
</tr>
<tr>
<td>EBL</td>
<td>electron beam lithography</td>
</tr>
<tr>
<td>FEC</td>
<td>forward error correction</td>
</tr>
<tr>
<td>FEOL</td>
<td>front end of the line of the CMOS manufacturing process</td>
</tr>
<tr>
<td>FPP</td>
<td>faceplate-pluggable</td>
</tr>
<tr>
<td>FSR</td>
<td>free spectral range</td>
</tr>
<tr>
<td>GVD</td>
<td>group velocity dispersion</td>
</tr>
<tr>
<td>IC</td>
<td>integrated circuit</td>
</tr>
<tr>
<td>IM/DD</td>
<td>intensity modulation / direct detection</td>
</tr>
<tr>
<td>IP</td>
<td>intellectual property</td>
</tr>
<tr>
<td>LER</td>
<td>line edge roughness</td>
</tr>
<tr>
<td>MMF</td>
<td>multi-mode fiber</td>
</tr>
<tr>
<td>MMI</td>
<td>multi-mode interference</td>
</tr>
<tr>
<td>MZI</td>
<td>Mach-Zehnder interferometer</td>
</tr>
<tr>
<td>NIR</td>
<td>near infrared</td>
</tr>
<tr>
<td>OOK</td>
<td>on-off keying</td>
</tr>
<tr>
<td>PAM</td>
<td>Pulse amplitude modulation</td>
</tr>
<tr>
<td>PIC</td>
<td>photonic integrated circuit</td>
</tr>
<tr>
<td>QAM</td>
<td>quadrature amplitude modulation</td>
</tr>
<tr>
<td>SEM</td>
<td>scanning electron microscope</td>
</tr>
<tr>
<td>SMF</td>
<td>single-mode fiber</td>
</tr>
<tr>
<td>SNR</td>
<td>signal-to-noise ratio</td>
</tr>
<tr>
<td>SOA</td>
<td>semiconductor optical amplifier</td>
</tr>
<tr>
<td>WDM</td>
<td>wavelength-division multiplexing</td>
</tr>
</tbody>
</table>
Acknowledgment

The PhD journey has been stimulating and challenging and it was made possible by many people that supported me. I would especially like to thank my partner Annaleena whose love, patience, and support allowed me to achieve this goal.

I would like to express my deepest gratitude to Prof. Victor Torres-Company for his supervision. His excellent guidance and vast knowledge pushed me to learn more and reach this point. Thank you for the great teaching and opportunities you have given me. My co-supervisor Prof. Anders Larsson deserves many thanks for his kind and calm encouragement and for sharing his knowledge and anecdotes about his career. Many thanks go also to my examiner, Prof. Åsa Haghund, Prof. Peter Andrekson, Prof. Magnus Karlsson and Dr. Jochen Schröder for helping with scientific and non-scientific questions. The Photonic Laboratory is a great work and study environment thanks to your efforts.

Special thanks go to Dr. Zhichao Ye, for teaching me so many things in the cleanroom and outside and for the nice chats sharing the office. I thank Dr. Xiaonan Hu for sharing his fabrication skills with me. Many thanks go to the whole staff of the cleanrooms of the myFab infrastructure. Particularly I would like to thank Dr. Niclas Lindvall for his relentless support on EBL processing and the enjoyable chats about Python and automation, Dr. Martin Karlsson for his constant efforts in maintaining the LPCVD furnace up and running, Dr. Alexei Kalaboukhov for his insight and help with AFM measurements, and Dr. Ruggero Verre for his support on many machines and his vibrantly positive attitude. Thank you also to Corrado Capriata, for introducing me to the art of CMP and teaching me how things are done in Electrum Lab.

I am also deeply thankful to Israel Rebolledo Salgado, for his beautiful friendship and for sharing his knowledge about signal processing and comb spectroscopy. Dr. Óskar Helgason deserves my gratitude for teaching me how microcombs actually work and for the endless chats about the most diverse topics. My appreciation goes to Dr. Krishna Twayana, for sharing this PhD journey and for his kindness at work and off work. Thanks to Yi Sun, for creating a nice office environment, Yan Gao for sharing his knowledge on lithium niobate, and Dr. Fuchuan Lei for sharing his knowledge. I want to extend my thanks to all the past and present colleagues in the Ultrafast Photonic group for creating an amazing environment, fostering excellent work, and enjoyable chats.
I would like to thank the colleagues I had the pleasure of collaborating with in the Photonics Lab. Particularly, Alexander Caut for sharing his designs and simulations, Dr. Mehdi Jahed for the nice discussions at work and outside while sharing some good food, and Dr. Ping Zhao for proposing new fabrication challenges. I would also like to thank Dr. Alexander Grabowski, Ekaterina Deriushkina, and Rasmus Larsson for filling break time with enjoyable topics, and Dr. Kovendhan Vijayan for his help, especially outside the work environment.

Many thanks to Prof. Kresten Yvind, Ahmed S. Alam, and Alif Laila Muthali from DTU, Prof. Pascual Muñoz and Roel van der Zon form UPV, Prof Andrew M. Weiner, Dr Cong Wang, and Dr Kaiyi Wu from Purdue University for the fruitful collaborations.

Life during the PhD has been more enjoyable thanks to my friends whom I shared fun times and experiences with. Many thanks to Hansu, Anni, Isra, Wendy, Gabi, Felix, Ali, Maryam, Kris, and Irenka.

Lastly, I would like to thank my family for their unconditional support, which allowed me to reach this goal and many others.
Contents

Abstract i

List of Publications iii

Abbreviations vii

Acknowledgement ix

1 Introduction 1
   1.1 History of telecommunication and photonics . . . . . . . . . . . . 1
   1.2 Photonic integration . . . . . . . . . . . . . . . . . . . . . . 2
      1.2.1 Si photonics . . . . . . . . . . . . . . . . . . . . . . . . 2
      1.2.2 Si₃N₄ photonics . . . . . . . . . . . . . . . . . . . . . . 3
      1.2.3 Beyond the single platform . . . . . . . . . . . . . . . 4
      1.2.4 Foundries . . . . . . . . . . . . . . . . . . . . . . . . . 6
   1.3 This thesis . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

2 Mode engineering in real waveguides 9
   2.1 Waveguide characteristics . . . . . . . . . . . . . . . . . . . . 9
   2.2 Material refractive index and thickness . . . . . . . . . . . . . . 12
      2.2.1 Refractive index measurement . . . . . . . . . . . . . . . 12
      2.2.2 Material uniformity . . . . . . . . . . . . . . . . . . . . 15
   2.3 Scattering loss . . . . . . . . . . . . . . . . . . . . . . . . . . 17
   2.4 Engineering the mode . . . . . . . . . . . . . . . . . . . . . . 20

3 Subtractive process and wafer-level manufacturing 23
   3.1 Overview of the fabrication process . . . . . . . . . . . . . . . . 23
   3.2 Crack barriers . . . . . . . . . . . . . . . . . . . . . . . . . . 25
   3.3 Wafer scale electron beam lithography . . . . . . . . . . . . . . 27
      3.3.1 Writing strategy of large patterns . . . . . . . . . . . . . . 29
      3.3.2 Time and placement optimization . . . . . . . . . . . . . . 30
   3.4 A foundry approach in academia . . . . . . . . . . . . . . . . . 31
4 Multilayer Si$_3$N$_4$ integration .................................................. 35
  4.1 A coupler for multilayer platforms .................................. 35
  4.2 Multilayer fabrication ..................................................... 37
    4.2.1 Chemical mechanical polishing ................................. 38
    4.2.2 Multilayer integrated devices ................................. 41

5 Study case: application to datacom ........................................... 45
  5.1 Short reach 1 \( \mu \)m interconnect ................................. 48
  5.2 Long reach 1.55 \( \mu \)m interconnect ............................... 49

6 Future outlook ........................................................................ 53

7 Summary of Papers .................................................................. 55

Bibliography .............................................................................. 59

Papers A-E .............................................................................. 71
Chapter 1

Introduction

1.1 History of telecommunication and photonics

The ability to communicate has been the basis of humankind evolution and a crucial incentive for technological development. In history, long-haul communication required physically moving a messenger from one place to another. One had to literally run a marathon to deliver the message "We have won" across the $\sim 40$ km between Marathon and Athens after the battle. Even at a world-record marathon pace, the communication link would not be good enough to stream a movie. Long-haul communications relied on horse riders, pigeons and other physical means until 1837 when S. Morse demonstrated his telegraph, based on switching an electrical signal transmitted through a conductive wire. This technology evolved forty years later into the telephone by A. G. Bell, where the voice was carried in real time over long distances. Telephones had great success and became widespread, however, an operator was required to connect the caller to the right receiver by physically connecting cables on a switchboard. The operators were replaced in the 1960s thanks to the development of the transistor in 1947\textsuperscript{1} and their integration in 1957\textsuperscript{2}. Integrated transistors and other components could now be fabricated in a planar device, creating the first integrated circuit (IC). The semiconductor manufacturing industry was born and semiconductors like silicon, germanium, gallium arsenide and indium phosphide were processed together with silicon dioxide, silicon nitride, copper, gold, aluminum and other materials to create increasingly complex ICs \cite{1}. Sixty years of advancements in IC technologies led to what is now known as the CMOS industry, which relies on advanced fabrication techniques and nanometer accuracy to provide high-performance processors in consumer devices and communication switches. Now, the switchboard operator is replaced by an application-specific integrated circuit (ASIC) that can operate a system with a capacity of $> 50 \times 10^{12}$ bit/s (50 Tb/s).

The development of communication cables followed a similar evolution path of the electronic at the endpoints. Early on, the transmission relied on

\textsuperscript{1}by J. Bardeen, W. H. Brattain, and W. B. Shockley, who won the Nobel prize in 1956.

\textsuperscript{2}by J. Kilby, who won the Nobel prize in 2000, and R. Noyce.
conductive cables where data was encoded by modulating a carrier frequency. Increasing the carrier frequency meant increasing the link bandwidth [2], however, electrical cables have a frequency-dependent loss [3], so increasing the channel capacity increased the transmission loss. The solution came with the invention of the laser in 1960 by C.H. Towens, N.G. Basov and A.M. Prokhorov (Nobel in 1964), and the advent of low-loss optical fibers, theorized by C. K. Kao in 1966 (Nobel in 2009) and now commercially available with 0.2 dB/km at 1.5 µm. This new technology provided an improvement of a factor of 700 in terms of losses and a 300-fold reduction in weight compared to a coaxial cable [4]. Now GaAs semiconductor lasers emitting light at 820nm could be modulated to send data on a link tens of kilometers long before the need for regeneration [5]. The development of erbium-doped fiber amplifiers (EDFA) in the 1980s [6] allowed to amplify the an optical signal at ~1550 nm without converting it to the electrical domain Long-haul cables crossing the oceans became easier to implement and more popular, leading to the connected world that we now live in.

Until the mid-2000s, the telecommunication market relied solely on discrete optical elements. However, the integration of optical systems was proposed early on to the community. Indeed, shortly after the development of low-loss optical fibers, Miller published in 1969 [7] what is considered the first proposal of an integrated waveguide, exploiting the same planar process used in the at the time fast-developing IC industry. However, the dream of integrating all the components present on a bulky optical table without requirements for alignment quickly faded due to the challenges in the fabrication process [8]. Nevertheless, single-material platform integrated optical circuits started to emerge based on the material platforms commercially developed for electronic applications [1]. These developments evolved in what is today called Photonics, a field strongly driven by the communication market [9] where integration of optical components can push further the limits of bandwidth and power consumption. Indeed, photonic integrated circuits (PICs) can be used to modulate and combine the signals from multiple lasers sources and transmit in on the same optical fiber [10] multiplying the capacity of a single optical link. Recent advances in integration exploiting mature photonic platforms [9], heterogeneous integration of gain materials [11] and co-packaging of PICs and ICs [12] led to the demonstration of transceivers capable of 1.6Tb/s [13], which enable the current growth of the telecommunication industry, pushed by our increasing demand for online content and services.

1.2 Photonic integration

1.2.1 Si photonics

Early on InP, GaAs and LiNbO₃ were considered for photonic integration due to their optical properties [9]. However, silicon became the most popular photonic platform in the research environment due to its low cost and well-established processes in the CMOS industry. The development of Si photonics rapidly evolved after the high-quality crystalline silicon-on-insulator (SOI)
wafers developed by Soitec in 1991 with the Smartcut process [14]. This created a standard platform for the photonic community to build upon [15]. The qualities of Si as a platform can be summarized in four main aspects. Firstly, the high refractive index contrast with SiO$_2$ provides a strong confinement that allows to guide the light in tight bends and attain compact circuits. Secondly, the material is transparent in the near-infrared (1.1-3.8 $\mu$m) [16], hence is compatible with the gain window of InP lasers [17], EDFAs [6], and other semiconductor optical amplifiers (SOAs), which makes it suitable for telecom applications. Thirdly, modulation and detection are not natively available, however, they can be obtained by doping the waveguides and epitaxially growing germanium respectively. The two processes were already available in CMOS production, hence the development of high-speed modulator [18] and detectors [19] did not disrupt the CMOS compatibility. Lastly, the material is at the basis of the CMOS industry, which guarantees the scalability necessary to tackle large volume markets. Indeed, the high-performance transceiver for the next generation interconnects are based on this material platform [13].

Despite the great progress in Si PICs, the platform has a few limitations that pushed the photonic community to explore other materials. A major missing element is the generation of light, which is not possible in Si due to its indirect bandgap. InP fills the gap, with the possibility of developing a complete platform [17], however, the material is expensive and the waveguides are lossy, which is not ideal for large integration of components [9]. For this reason, integrating optical gain materials on other photonic platforms became more advantageous than building the whole circuit in InP [20]. For passive components, the high index contrast of SOI waveguides, which is a great advantage for integration density, induces increased scattering losses [21] that limit the losses of SOI strip waveguides to $\sim$1 dB/cm [22]. Moreover, two-photon and free carrier absorption limit the application of Si to nonlinear applications [9]. A material platform that filled the gap in terms of loss, nonlinearities, and transparency window is silicon nitride, which expanded the field of photonics towards the visible and NIR.

### 1.2.2 Si$_3$N$_4$ photonics

Silicon nitride was initially used as an insulation layer in CMOS processes [23], but it became successful in the photonic field thanks to its large transparency window (400-4000 nm) [24], the Kerr nonlinearities and the lack of two-photon absorption, which makes it ideal for nonlinear optics. Moreover, the possibility to be manufactured on planar Si wafers and the CMOS compatibility of the processes makes it a good candidate to substitute Si as a passive platform.

The transparency window of Si$_3$N$_4$ enabled applications in the very near IR, and visible spectrum, instrumental for biosensing [25], augmented reality [26] and the current standards in data center interconnect, operating at 850 nm. This frequency range is less demanding in terms of film quality compared to the telecommunication range at 1550 nm (C-band) [27], [28], but it is characterized by increased scattering loss due to the shorter wavelength.

In the telecom bands (1300-1600 nm), the moderate index contrast between
Chapter 1. Introduction

Si$_3$N$_4$ and SiO$_2$ reduces the scattering loss due to sidewall roughness compared to SOI waveguides. Indeed, loss below 0.1dB/m can be achieved in ultrathin waveguides with low confinement [29]–[31]. However, the low confinement geometry, requires large bending radii, in the order of mm, which drastically decreases the integration density. Moreover, a single core waveguide cannot be dispersion engineered to attain anomalous dispersion, limiting the possibility to exploit the Kerr nonlinearities of the core material. For this reason, thicker Si$_3$N$_4$ layers were deposited, with a great effort to bypass the crack generated by the tensile stress of the material. Different methods were developed and an overview is reported in Section 3.2.

Overcoming the challenges of depositing thick Si$_3$N$_4$ layers opened new possibilities in nonlinear optics, with waveguides engineered to display anomalous dispersion in the NIR. The generation of dissipative Kerr solitons in Si$_3$N$_4$ waveguide pushed multiple groups in the world to improve the fabrication methods to achieve ultralow loss in dispersion engineered Si$_3$N$_4$, including Cornell (now Columbia) [32]–[34], EPFL [35], [36], Chalmers [37]–[40], Twente [41], Purdue [42], CEA-LETI [43], [44], HKUST [45], and companies like LioniX [46] and Ligentec [35]. The advancement in the fabrication methods propelled integrated nonlinear optics on Si$_3$N$_4$ with the generation of frequency combs [47], which find applications in various fields, such as timekeeping [48], spectroscopy [49], ranging [50], [51], and telecom [52]–[54], where the multitude of lines generated with a single pump laser could substitute tens or hundreds of lasers in an optical communication link.

The excellent performance of Si$_3$N$_4$ in terms of losses, transparency window, power handling capability, and nonlinear optics suggests that this platform could be a strategic complement or replacement for SOI [55]. However, similarly to SOI, also on Si$_3$N$_4$ it is necessary to integrate other materials to expand the functionality with efficient modulation, detection, and optical gain.

1.2.3 Beyond the single platform

A single-platform approach is inherently limiting the performance of the PIC, based on the choice of the material and the core geometry. For this reason, research and industry pushed to integrate multiple materials on the same chip, in an attempt to get closer to the full optical table on a chip envisioned in the 1970s [8]. The integration of other materials in a photonic platform usually follows three main routes, each with its unique advantages.

**Hybrid integration** consists of combining micro-optical elements fabricated on different substrates coupled together by means of packaging techniques, e.g., glue and solder. This includes the integration of semiconductor optical amplifiers [56], [57], lasers [58], and photodetectors [59]. These examples are based on edge-coupling, which is a simple method to combine materials because it is comparable to coupling the signal in an optical fiber. The drawback is that integration density is limited by the available edge space and the alignment of the optical elements. To increase the integration density, flip-chip bonding [60] can be used, where a chip is coupled to the waveguides in the out-of-plane direction. This technique is particularly useful in vertical-cavity...
1.2. Photonic integration

Surface-emitting lasers (VCSEL) [60]. In this case, the available space scales with the area of the chip, allowing for more devices to be integrated. Hybrid integration methods rely on the packaging of devices one by one with an alignment procedure that can be passive or active. The serial nature of the process inherently limits the scalability which will be limited by the packaging time and cost. A technique that could solve some of the drawbacks of hybrid integration is the photonic wire bonding [61], where a polymeric waveguide is patterned in space using a two-photon polymerization technique to connect various optical devices. The alignment is achieved lithographically, hence increasing the throughput and the number of connections possible.

**Heterogeneous integration** combines two or more materials by means of wafer or die bonding techniques. The host substrate is prepared with the PIC and a new material is bonded on top. This can be performed at the wafer, die, or component level. This technique allows exploiting lithographic techniques for the alignment of the different layers, which guarantees high alignment accuracy and high throughput [55]. With this technique, it is possible to integrate III-V dies on Si for lasers on Si$_3$N$_4$ to attain feedback resilient lasers [62], microcombs with an integrated pump laser [63] and LiNbO$_3$ modulators on Si$_3$N$_3$ [64]. The drawback of this technique is that the process requires a highly planar substrate and the majority of the expensive III-V semiconductor is etched away during the process. An alternative is micro-transfer printing ($\mu$TP), where the active material is processed separately and only the working dies are transferred by an elastomeric stamp [65]. This technique can optimize the utilization of the III-V substrate and promises high throughput via the multiplicity of the coupons transferred.

**Monolithic/multilayer integration** is an approach where a material is grown or deposited on the PIC. The integration method is similar and results in a multilayer structure. When a crystalline material is grown on a PIC, it is referred to as monolithic integration. A few examples are Ge on SOI/Si$_3$N$_4$ [66] for photodetection, III-V on Si$_3$N$_4$ [67] for lasing or piezoelectric on Si$_3$N$_4$ [68] for efficient modulation. When a material is deposited by means of chemical vapour deposition (CVD) the monolithic term is often dropped in favour of multilayer integration, which is useful to expand the functionalities of a platform. This approach will be discussed in more detail in Ch. 4. The drawback of integration relying on deposition and growth is that these processes require high temperatures, which can present challenges for the front-end of the line (FEOL) and back-end of the line (BEOL) manufacturing of PICs. For this reason, deposition techniques based on low-temperature processes are actively explored [27], [69].

Expanding the functionalities of PICs, especially by integrating Si$_3$N$_4$ on SOI, is now common in commercial foundries to expand the capability of the circuit.
1.2.4 Foundries

Until the mid 2000s, the photonic field was already mature enough to deliver the first commercial products [10], however, only companies and research groups with access to a cleanroom had the possibility to convert their designs into chips. On the other hand, the semiconductor industry developed a model in the 1980s where companies with an established process could provide fabrication services to fabless groups. This was possible by decoupling the intellectual property (IP) of the fabrication from the IP of the integrated system design, opening a new era of fabless IC companies [70]. In this model, multiple users could get together and purchase areas of the wafer, sharing the cost of the fabrication in multi-project wafer (MPW) runs. The design is based on a process design kit (PDK) defined by the foundry, which does not own the IP of the final circuit but might have the IP of selected building blocks. In this way, the development cost and fabrication of an integrated circuit went from millions of dollars to thousands [70], sparking the development of fabless companies.

In photonics, ePIXfab, a collaboration between CEA-LETI and IMEC hosted by Ghent University pioneered the concept of open-access foundry by providing fabrication services. Initially, the cost was still sustained by a single user that took the risk of the fabrication run [70]. The model later evolved into the MPW approach.

Currently, the foundries that provide MPWs are many and the list keeps on expanding. Most of the foundries base their offer on SOI where passive and active components are defined [15], [71]. Some foundries expanded the SOI portfolio to Si$_3$N$_4$, which is integrated on top of the SOI to exploit the superior propagation loss and further expand the functionalities of a single chip [72], [73]. However, all these foundries utilize Si$_3$N$_4$ for the integration of passive components, with thicknesses between 100 and 400 nm. On these platforms, it is impossible to attain anomalous dispersion, which limits the functionalities in nonlinear optics. A few foundries are filling the gap: LioniX, with their TripleX platform [46], Ligentec, with the photonic Damascene process developed in EPFL [35], Qaleido Photonics [74] and Iloomina AB a spin-off from our group. Moreover, new platforms developed in academia are emerging in the commercial landscape, like Al$_2$O$_3$ offered by Aluvia Photonics for applications in the UV [75] and LiNbO$_3$ by CSEM for high-speed modulation and low loss applications [76].

1.3 This thesis

This thesis focuses on the development of methods for the fabrication of ultralow loss Si$_3$N$_4$ waveguides at the wafer scale, tailoring the properties of the waveguide to the desired application. A multilayer integration approach is used to bypass the limitation of the single platform and expand the capabilities of the photonic integrated circuit. The research is framed in the context of data center optical interconnects, with the goal of addressing the increasing bandwidth demand with power-efficient solutions, but the techniques are sufficiently general so that they could enable applications in different fields.
In Paper A we propose a passive $\text{Si}_3\text{N}_4$ photonic platform for data center optical interconnects operating at 1 $\mu$m, with a reach of a few kilometers and Tb/s capacity. The power budget of an unamplified optical link operating with GaAs VCSEL is studied, setting the target in terms of losses for the integrated optical components.

For long-reach (100 km) and high capacity (>10 Tb/s) interconnects we target a coherent optical link operating in the telecom C-band, based on Kerr frequency combs integrated with LiNbO$_3$ modulator. Paper B presents the strategy to efficiently couple the light from the dispersion-engineered layer to a strip loaded LiNbO$_3$ waveguide for the modulation of the optical signal.

The challenges associated with the development of a multilayer platform pushed me to consolidate the fabrication of ultralow loss $\text{Si}_3\text{N}_4$ on 100 mm wafers, building tools for the design of PICs on our photonic platform. The tools and the reliability of the process enabled the demonstration of dissipative Kerr microcombs with a substantial improvement in terms of power conversion efficiency, reported in Paper C.

In Paper D, the multilayer integration of a dispersion engineered and a linear $\text{Si}_3\text{N}_4$ platform are reported. The study shows the benefit of integrating these two layers of $\text{Si}_3\text{N}_4$ by evaluating the loss of a single-mode waveguide. The multilayer integrated device reported demultiplexes the lines of a power-efficient microcomb demonstrating the first half of a datacom transmitter based on an integrated frequency comb.

Paper E, addresses the scalability of power-efficient microcombs, instrumental for applications in large-scale markets such as data center interconnects. A power-efficient microcomb is replicated 50 times on a wafer and the comb generation yield is measured to be 98%, with average conversion efficiency >50%. This is the first-ever analysis of the yield of chip-scale comb generators.

**Thesis outline**

In Chapter 2 we describe the basic aspects that characterize a real integrated waveguide and how the fabrication affects its performance. Chapter 3 reports the wafer-level manufacturing of $\text{Si}_3\text{N}_4$ photonic integrated circuits and how to apply a foundry approach in an academic environment. Chapter 4 explains how to integrate more than one photonic layer and expand the functionality of the PIC in a multilayer integrated device. Chapter 5 describes how the integration of $\text{Si}_3\text{N}_4$ with other materials such as GaAs based VCSELs and LiNbO$_3$ can address the increasing demand for bandwidth in data center interconnects, following the requirements set by the optical power budget of the communication link. Chapter 6 provides a future outlook.
Chapter 1. Introduction
Chapter 2

Mode engineering in real waveguides

Integrated photonic devices are capable of guiding light in narrow waveguides thanks to the index contrast between the core and cladding materials. The cross-sectional geometry of the core can be engineered to tailor its properties, e.g., the number of supported modes and the dispersion. However, the performance of real waveguides can significantly differ from the simulation due to fabrication imperfections. In this chapter, we describe how a waveguide can be engineered to fulfill the specification required by the intended application and how the fabrication can influence the outcome.

2.1 Waveguide characteristics

Light can be guided in a structure composed of a core material with a higher refractive index \( n_{\text{core}} \) and a cladding material with a lower refractive index \( n_{\text{clad}} \). These structures confine the light in the transverse plane (x and y) so that the field distribution remains unchanged upon propagation in the longitudinal direction (z). In this situation, the electrical field can be expressed with the following mathematical structure:

\[
\vec{E}(x, y, z) = \vec{E}_t(x, y)e^{i\beta z}
\]  

(2.1)

where \( \vec{E}_t(x, y) \) is the field distribution on the plane perpendicular to the propagation and \( \beta \) is the propagation constant. This is a solution for the vector wave equation, which can be found in all the books on the subject, e.g., ref [77]. It is important to underline that this is an eigenvalue equation \( \Box \vec{a} = \beta^2 \vec{a} \), with \( \Box \) being a differential vector operator, and the eigensolutions are called modes \( \beta^2 \) being the eigenvalue. The vector wave equation does not have an analytical solution except in the case of a slab waveguide or cylindrical geometries. However, it can be calculated with finite element computational methods available in commercial software, e.g., Comsol Multiphysics or Ansys Lumerical. The results of the calculation are the propagation constant and the
field distribution, from which we can calculate optical confinement, polarization state, dispersion, and so on.

The propagation constant is a complex value where the real part describes the phase accumulated by the field upon propagation and the imaginary part accounts for the field attenuation. From the propagation constant we can define the effective refractive index:

\[ n_{\text{eff}} = \frac{\beta}{k_0} \]  

where \( k_0 = \frac{2\pi}{\lambda} \) is the wavenumber in vacuum. Guided modes satisfy the relation \( n_{\text{clad}} < n_{\text{eff}} < n_{\text{core}} \), which shows that the mode is guided in the high index core, but part of the field distribution is propagating in the cladding material. In Fig. 2.1 we show two examples of strip waveguides, one with weak confinement (a) and one with strong confinement (b). We can see that the mode distribution in the low confinement waveguide extends in the cladding much more than in the strong confinement one and this is reflected by the low \( n_{\text{eff}} \), approaching the value of \( n_{\text{clad}} \).

The mode area distribution can be evaluated also with the effective mode area, defined as [78]:

\[ A_{\text{eff}} = \frac{\int |(\vec{E} \times \vec{H}^*) \cdot \hat{z} dA|^2}{\int |(\vec{E} \times \vec{H}^*) \cdot \hat{z}|^2 dA} \]  

where \( \vec{E} \) and \( \vec{H} \) are respectively the electric and magnetic fields of the selected mode. This parameter can be used to evaluate the optical confinement, i.e., how much of the field is in the area defined by the core cross-section. This is particularly useful when the interaction of the field with the core material is of interest, e.g., in nonlinear photonics.

The polarization of the mode is defined by which field is oriented in the \( x \) transversal direction (\( \hat{x} \)), following the axis in Fig. 2.1. When the electric field is parallel to \( \hat{x} \), the mode has transverse electric (TE) polarization, while when the magnetic field is parallel to \( \hat{x} \) the mode has transfer magnetic (TM) polarization. It should be noted that pure polarization states are not possible in optical waveguides. Indeed, in the vector wave equation, the inhomogeneity due

---

Figure 2.1: Simulated fundamental TE modes calculated at a wavelength of 1550 nm for core dimensions a) \( 50 \times 5000 \) nm\(^2\) and b) \( 800 \times 1800 \) nm\(^2\).
2.1. Waveguide characteristics

The field is normalized so that the maximum of the field intensity $|E|^2$ is 1.

It should be noted that the mode changes with the wavelength due to two effects. One is the waveguide dispersion, i.e. the variation of the $n_{eff}$ due to the variation of the optical confinement. At longer wavelengths the confinement decreases, hence $n_{eff}$ decreases, approaching the index of the cladding. The other is the material dispersion, i.e., the variation of the optical constants of the material with the wavelength (see Section 2.2). The combined effects results in a $n_{eff}(\lambda)$. The dependency can be studied by expressing $\beta$ in Taylor expansion terms around the central angular frequency $\omega = 2\pi\nu$ as [79]:

$$\beta(\omega) = \beta_0 + (\omega - \omega_0)\beta_1 + \frac{1}{2!}(\omega - \omega_0)^2\beta_2 + \frac{1}{3!}(\omega - \omega_0)^3\beta_3 + ... \quad (2.4)$$

where the $m$-th order dispersion is calculated as:

$$\beta_m = \left. \frac{d^m \beta}{d\omega^m} \right|_{\omega=\omega_0} \quad (2.5)$$

From these equations, we can obtain important parameters of the waveguide such as the phase velocity from $\beta_0$, the group velocity $\beta_1^{-1}$, and the group velocity dispersion ($\beta_2$). These parameters can be calculated from the simulation by the polynomial fitting of $\beta$ as a function of $\mu = (\omega - \omega_0)$.

As mentioned above, all the parameters can be calculated using simulation methods based on finite element calculation. It should be noted that these powerful tools can introduce numerical artifacts, such as excess loss or gain due to the interaction of the field with the boundary condition of the simulation. It is paramount to perform simulation convergence tests to ensure the numerical artifacts are negligible. These convergence tests include sweeping the size of the simulation area, the mesh size, and the mesh type until the results do not change within a certain tolerance. The interested reader can find a more extensive discussion in ref. [80]. However, the parameters of the waveguide can be predicted accurately only if the optical constants of the materials are
evaluated and modeled correctly, and the core dimensions of the fabricated waveguide are close to the designed parameters. These aspects are explored in the next section.

## 2.2 Material refractive index and thickness

The simulation of optical modes requires an accurate evaluation of the material parameters, namely the refractive index and the film thickness. The measurement of the refractive index is based on the refraction of light at an interface, as described by Snell’s law. The direct measurement of the angle of refraction is still used in handheld refractometers to evaluate the sugar concentration of aqueous solution, a common measurement in winemaking. However, these direct measurements rely on the absolute intensity of the light, which is dependent on material imperfections. For this reason, a more sophisticated technique called ellipsometry is used in the CMOS industry.

### 2.2.1 Refractive index measurement

Ellipsometry allows non-contact and nondestructive measurement of thin layers of materials deposited on wafers, yielding the optical constants and the material thickness. Ellipsometry is based on the interaction of polarized light with one or more layers of materials. A schematic of the measurement system is reported in Fig. 2.3. Linearly polarized light illuminates the sample that reflects it to a detector which records the intensity of the beam for every polarization. In spectroscopic ellipsometry, the input beam is broadband and the detector record the signal in a wavelength-dependent manner. The so-called complex depolarization parameter $\rho(\lambda)$ is:

$$\rho(\lambda) = \tan(\psi(\lambda))e^{i\Delta(\lambda)}$$

(2.6)

where $\psi$ and $\Delta$ are the depolarization magnitude and phase respectively [81]. The relation between $\psi$, $\Delta$, and the optical properties of the sample is usually transcendental, hence an inverse approach is used to obtain the index and thickness of the material. A model is defined based on the material properties and fitted to the experimental $\psi$ and $\Delta$. The model includes the substrate and the layer stack, i.e., the sequence of layers present on the substrate with expected thicknesses and refractive index models, and other corrective factors such as roughness layers. These corrections take into account imperfections of the layers and improve the fitting. A detailed description of the corrective factors can be found in ref. [81]. The most important step in ellipsometry is the definition of the index model based on the material. The quality of the model determines the quality of the fitting from which the optical constants are extracted. In the next section, we will discuss the basis of the model and how it can influence the simulation of a waveguide mode.
2.2. Material refractive index and thickness

Figure 2.3: Schematic of the working principle of ellipsometry. A linearly polarized light is reflected by the sample, which changes the polarization state. The input polarization is rotated and detected in a polarization-sensitive manner. $E_{f,i}$ are the electrical field components for each polarization. The polarization states $p$ and $s$ are respectively parallel and perpendicular to the incident plane.

Refractive index model

The refractive index model is built based on consideration of the interaction between electromagnetic radiation and the molecules that compose a material. These can be described as dipoles and behave like harmonic oscillators. The macroscopic effect on the electric field $\vec{E}$ is described by the electric flux [2]:

$$\vec{D} = \varepsilon_r \varepsilon_0 \vec{E} = \varepsilon_0 \vec{E} + \vec{P}$$

(2.7)

where $\varepsilon_r$ is the relative permittivity of the material, $\varepsilon_0$ is the vacuum permittivity and $\vec{P}$ is the polarization of the material. The relative permittivity $\varepsilon_r$ depends on the dipole density and contains information on how a material reacts to incident light both in phase and amplitude. The permittivity is a scalar in isotropic materials like amorphous Si$_3$N$_4$ and SiO$_2$, but its tensorial nature should be included in birefringent materials, e.g., crystalline LiNbO$_3$. The refractive index is defined from the relative permittivity as [81]:

$$n = \sqrt{\varepsilon_r} = n_R - i \, n_I$$

(2.8)

and the complex dielectric constant can be expressed as a sum of oscillators as:

$$\varepsilon_r = \varepsilon_1 - i \varepsilon_2 = \sum_{\text{oscillator}}$$

(2.9)

The oscillators can have different mathematical formulations based on the type of absorption they model. For example, a Sellmeier oscillator can be used to
Figure 2.4: Index models for Si₃N₄: a) comparison between the index calculated using a pole in the mid-IR and a series of experimentally calculated Gaussian oscillators. b) Comparison between the simulated GVD obtained with the two index models and the experimental value obtained from one of the devices in Paper E.

model a transparent material while the absorption close to the energy gap is better modeled by a Cody-Lorentz oscillator. A detailed description of all the oscillator models is beyond the scope of this thesis. The interested reader is directed to ref. [81] for an extensive explanation.

The model based on oscillators is physical only if it fulfills the Kramers-Kronig relation, which couples the real and imaginary parts of the permittivity. However, the Kramers-Kronig integral is defined over the whole spectrum, therefore also oscillators outside the measurement range of the ellipsometer play a role in determining the optical constants of the material [34]. This can result in two effects in the optical constant: an offset and a tilt [81]. Normally, this is compensated by adding a real term \( \varepsilon_1(\infty) \) to compensate for the offset and a simplified Sellmeier oscillator, called pole, to compensate for the tilt [81]. This pole is usually positioned in the mid-IR portion of the spectrum to compensate for the vibrational modes of the molecules. However, this correction does not perform well in our case, so a heuristic correction was adopted.

In our cleanroom, we inspect the Si₃N₄ deposition (see Chap. 3) with a spectroscopic ellipsometer\(^1\) in the wavelength range of 210 to 2500 nm. The blue line in Fig. 2.4a represents the refractive index of Si₃N₄ evaluated from the data collected with the offset-pole correction. This index was used to simulate the GVD of a bent waveguide reported in Fig. 2.4b (blue). If we compare this data to the experimental GVD calculated from the response of a ring resonator (black), we can notice a great discrepancy between the values. For this reason, we apply a different correction to the model by substituting the pole with multiple Gaussian oscillators in the mid-IR region of the spectrum (>5µm). These Gaussian oscillators were determined by a

---

\(^1\)J.A. Woollam R2
2.2. Material refractive index and thickness

The design of a photonic integrated device relies on the simulation based on the index discussed above and the thickness of the core. However, the material thickness varies across the wafer because the deposition rate in LPCVD furnaces is not homogeneous across the wafer. This is due to the fact that the wafers are stacked on a carrier boat closely together, with gaps in the order of 5 mm. The dense packing of the wafers creates a gradient in the concentration of the reactive spices [83], which leaves a concentric thickness pattern on the wafer, as

Figure 2.5: Refractive index of the deposited $\text{Si}_3\text{N}_4$ before and after annealing. The inset shows the increase in the index due to the densification of the layer.

broadband ellipsometry measurement (210-30000 nm) done at J.A. Woollam and are added to the index model as constant. This correction results in the red curves in Fig. 2.4. It is clear that they introduce a significant difference in the refractive index, especially in the most interesting portion of the spectrum for us, i.e., at 1550 nm. This provides more accurate simulation results, as we can see in Fig. 2.4b. A discrepancy of $\sim 10 \text{ ps}^2/\text{km}$ is still present, suggesting that more investigations are needed. A possible effect is the stress introduced by the deposition of the cladding, which can affect the index of the layers below, as reported in ref. [82].

Another important parameter to take into consideration is the influence of annealing on the refractive index. Indeed, in the fabrication process (see Section 3.1), the material is annealed to drive out the N-H bonds that increase the material loss in the telecom band [31]. This densifies the material and increases the refractive index. During the annealing process, the material shrinks by 2%, therefore it is important to deposit a thicker layer to attain the desired thickness. Hence, in the simulation the index and thickness used should be the ones measured after the annealing process. The refractive index before and after annealing is shown in Fig. 2.5.

2.2.2 Material uniformity

The design of a photonic integrated device relies on the simulation based on the index discussed above and the thickness of the core. However, the material thickness varies across the wafer because the deposition rate in LPCVD furnaces is not homogeneous across the wafer. This is due to the fact that the wafers are stacked on a carrier boat closely together, with gaps in the order of 5 mm. The dense packing of the wafers creates a gradient in the concentration of the reactive spices [83], which leaves a concentric thickness pattern on the wafer, as
Chapter 2. Mode engineering in real waveguides

Figure 2.6: Thickness non-uniformity: a) thickness map of the Si$_3$N$_4$ layer deposited on a wafer. b) GVD variation as a function of the core height for a fixed waveguide width of 1800 nm.

shown in Fig. 2.6a. The non-uniformity on the wafer can be calculated as [84]:

\[ \text{non-uniformity} = \frac{t_{\text{max}} - t_{\text{min}}}{2t_{\text{avg}}} \]  

(2.10)

The thickness nonuniformity across the wafer is usually between 2 and 3%, however, not all the wafer is patterned (see Section 3.2). In the area where the devices are patterned, the nonuniformity is usually below 2%. This can still significantly affect the waveguide parameters, e.g., the GVD as shown in Fig. 2.6b. This issue can be detrimental if a specific parameter is of interest, e.g., in supercontinuum generation where close to zero dispersion is important to achieve a broadband spectrum. In this scenario, a difference of 10 ps$^2$/km can change the output spectrum of the pumped device. The variation of the performance of a device due to the thickness nonuniformity is perfectly pictured in the repetition rate of the devices reported in Paper E, which clearly changes based on the thickness of the waveguide.

To mitigate the effect of the fabrication parameters on the performance of the device there are two approaches. The first approach is to measure the thickness and design the devices based on the location of the chip on the wafer. This requires an extensive simulation and design of the chip and removes any flexibility on the fabrication side to avoid areas of the wafer with defects. This is still a viable option, but it is limited to research runs, where the feedback between fabrication and simulation is closer. This approach was used to achieve the results reported in [48]. The second approach is to design fabrication-tolerant devices, where the device itself is resilient to fabrication variability [85]. This approach is much more suited for commercial runs, where the performance of the devices provided in the PDK has to be guaranteed within certain specifications.
2.3 Scattering loss

As we saw in the previous section, the waveguide performance can differ from the simulation due to the evaluation of the optical constant or the variation in the thickness of the material. Another aspect that should be taken into account, is the scattering loss introduced by roughness on the waveguide, which can significantly affect the performance of the PIC.

The waveguide propagation loss can be divided into intrinsic and extrinsic. The intrinsic loss takes into account the material absorption, while the extrinsic loss includes all the effects that couple the propagation mode to an undesired mode, e.g., radiation or backward propagating modes [21]. Among these, we find bending loss, which can be avoided by an appropriate design of the bending radius of the waveguide [2], and scattering loss. Scattering loss is due to the interaction of the propagating mode with an interface with imperfections. These imperfections can be point defects or roughness at the interfaces [31]. They are unavoidable features of real photonic waveguides due to machine tolerances and physical processes [21].

In a transparent material like Si$_3$N$_4$ the roughness at the interfaces is a major contribution to the propagation loss and it should be managed throughout the fabrication process [86]. The roughness at the interfaces can be divided into surface roughness, which arises from the deposition of the material, and sidewall roughness which is introduced by the patterning and etching processes. The surface roughness is usually smaller than the sidewall roughness, therefore, when a platform is optimized, most of the effort goes to the optimization of the lithography and etching. When the process is optimized, the sidewall roughness is barely visible with SEM imaging, as shown in Fig. 2.7. The sidewall roughness is characterized by a striping in the vertical direction. This is due to the etching process that transfers any imperfection of the mask throughout the etching, resulting in waveguides with a quasi-unidimensional roughness distribution in the longitudinal direction [21]. This characteristic simplifies the modeling of the sidewall roughness to a 2D problem.

The roughness can be mathematically described by a zero-mean random
function \( f(z) \) characterized by two parameters: the root mean square (RMS) \( \sigma \) and the correlation length \( L_c \). The former provides information about the amplitude of the roughness, while the latter quantifies the randomness of the distribution. When \( L_c = 0 \), the roughness is comparable to white noise and with \( L_c = \infty \) the pattern is completely correlated. The autocorrelation function \( (R(\tau)) \) of \( f(z) \) provides all the necessary information to evaluate the roughness because \( L_c \) is defined as the half-width at \( R(0)/\varepsilon \) and \( R(0) = \sigma^2 \) [21].

The best technique to measure surfaces with sub-nanometer morphology is atomic force microscopy (AFM) [87]. However, measuring the sidewalls of the waveguide is a more challenging process. The AFM is designed to be sensitive to displacement in the vertical direction, however, for sidewall roughness the displacement introduces a torsion on the cantilever. In this case, specialized cantilevers, microscopes, and processing are needed to detect sidewall roughness [88]. The detection with scanning electron microscopy (SEM) would be beneficial thanks to the nondestructive nature of the process, however, the resolution of SEM is limited to \( \approx 1 \) nm [89] and it provides only a qualitative analysis of waveguides with small roughness. The sidewall roughness can be also evaluated with AFM by tilting the sample [89] or cutting a portion of the waveguide and tilting it on its side to be measured as a flat surface [90]. Both techniques provide excellent measurement results, however, they are suitable for small chips or they are destructive. Given the results reported in [21], [90], evaluating the line edge roughness (LER) at the top of the waveguide should give a reasonable estimation of the sidewall roughness of the waveguide. In an attempt to measure non-destructively the waveguides on the wafer, we tried to image the LER with the side of a triangular-tipped cantilever as shown in figure Fig. 2.8. The idea is that a triangular tip (Fig. 2.8b) could make contact with the top edge of the waveguide as shown in Fig. 2.8a, providing information about the sidewall roughness at the top of the waveguide. The scan obtained with this technique is reported in Fig. 2.8c However, this technique proved to be challenging due to the cantilever losing contact with the surface of the sample. The measurement reported in Fig. 2.8d has \( \sigma = 0.78 \) nm and \( L_c = 13.4 \) nm close to our expectation comparing the results obtained in paper D and the results reported by Roberts et al. [90]. This AFM technique shows potential for nondestructive measurements on planar wafers when more advanced AFM tools are not available, but it still requires further development.

**Scattering loss model**

The interaction of the electric field with the interface roughness can be modeled as the coupling between the guided mode and radiation modes, considering the interface as radiating antennas [91]. This approach led to a simple 2D model developed by Payne and Lacey [91], [92] (PL) which provides an analytical form to estimate the scattering loss of a slab waveguide due to the roughness at the interface. The propagation loss can be estimated as:

\[
\alpha_{dB} = 4.34 \frac{\sigma^2}{\sqrt{2k_0d^4n_{core}}} \alpha V f_c \tag{2.11}
\]
2.3. Scattering loss

The waveguide geometry, index, and propagation constant are included in the parameter \( g_V \), and \( f_e \) describes the integral between the mode and the roughness spectral density and the modal field. This analytical solution found great success in the photonics community because of its simplicity and it has been extended to work with two-dimensional waveguides using the effective index method [93]. More rigorous models to accurately calculate the radiation loss were proposed [94], [95], but the calculation requires solving more computationally heavy integrals. Notwithstanding, quasi-empirical corrections are still proposed, e.g., for rib waveguides [89] and more recently bent waveguides [90], showing good agreement with the experimental data.

The extended PL model is an effective tool to visualize which type of waveguide is more resilient to scattering loss. In particular, the colormap reported in Fig. 2.9 shows how a waveguide fabricated in thick Si\(_3\)N\(_4\) suffers from extra loss when the core width is reduced to achieve the single mode condition. On the other hand, a thin waveguide has lower scattering loss when the single mode condition is achieved, as discussed in Paper D.

Figure 2.8: AFM evaluation of the sidewall roughness. a) Schematic of the tip scanning on the waveguide b) schematic of the pyramidal cantilever used (RTESP-300). The edge makes contact with the side of the waveguide leaving a slope profile in the micrograph. c) AFM micrograph of the scan on a waveguide 200 nm thick. d) profile used to evaluate the sidewall roughness and e) autocorrelation function of the trace.

where \( \sigma \) is the RMS roughness and \( d \) is the half-width of the waveguide.

\[
\text{loss} = \frac{\sigma}{d} \int f_e \rho \, d\zeta
\]
Chapter 2. Mode engineering in real waveguides

Figure 2.9: Payne and Lacey model of the scattering loss. a) Sidewall scattering loss changing the roughness RMS and its correlation length for a waveguide with core geometry (h×w) 740x1800 nm b) Surface and sidewall scattering loss for different waveguide geometries. The surface roughness parameters are reported in Paper D and the sidewall roughness has parameters $\sigma = 0.6$ nm and $L_c = 40$ nm.

It should be mentioned that this model does not predict the cases where the aspect ratio of the waveguide is large, e.g., in ultrathin waveguides, and it overestimates the loss [94]. However, this is a useful tool to engineer the mode, avoiding waveguide geometries that maximize the scattering loss.

2.4 Engineering the mode

Engineering the mode means tailoring the waveguide geometry and material to fulfill the functionality the PIC is intended for. This often boils down to selecting an optical material, based on its properties and the material thickness, which determines the core height. Indeed, the core width becomes the only degree of freedom when we design a PIC in a planar, single-layer platform.

The mode engineering is explained here with a few examples based on Si$_3$N$_4$, but it can be easily extended to other materials. The main parameters that should be considered designing a Si$_3$N$_4$ waveguide are the loss, the dispersion, the number of modes, and the footprint. A waveguide designed for linear operation is usually tailored to have a single mode, which avoids coupling to higher-order modes [2]. Linear waveguides can be designed with ultrathin cores (<100 nm) to minimize the confinement. This has two advantages: the interaction with the sidewall roughness is minimized (see Section 2.3) and the optical field is distributed on a large cross-sectional area, which guarantees low nonlinear loss thanks to the low local intensity [96]. Record low loss ($\sim 0.1$ dB/m) has been achieved on these platforms [29]–[31], however, to avoid bending loss a large bending radius is necessary (>1 mm). These types of waveguides find applications in integrated lasers [97], where the quality factor and the mode volume play an important role in reducing the linewidth [58].
2.4. Engineering the mode

Figure 2.10: Mode engineering in Si$_3$N$_4$ waveguides. a) Colormap of the effective mode area of the fundamental TE mode with highlight the number of modes supported by the waveguide. The solid line shows the single-mode condition, where one mode in each polarization is supported. b) Colormap of the GVD for different core geometries. The single mode condition is marked in black and the separation between normal and anomalous dispersion regimes is marked with a red line.

However, the large footprint due to the bending radius makes them impractical for applications where the device density is also targeted, e.g., in data center interconnect. A possible drawback of this kind of geometry is the size of the mode, e.g., the one shown in Fig. 2.1a. To optimally confine the mode, it is indeed necessary to have thick layers (>10 µm) of high quality SiO$_2$, attained by bonding thermally oxidized wafers [30] or depositing specialized materials like deuterated SiO$_2$ to minimize the absorption of the cladding in the telecom band [98].

Increasing the core thickness allows for a drastic decrease in the bending radius, which improves the density of devices fabricated on the chip. The moderate confinement guarantees a small interaction with the sidewalls, and a small bending radius, in the order of 100 µm. The moderate confinement platforms are versatile because they combine the qualities of the ultrathin and thick waveguides [24]. For this reason, moderate confinement platforms were used in Paper A and D. These waveguides find application in data center interconnects [27], [99], where the mode size and bending loss play a role. Moreover, as we can see in Fig. 2.10a, the single mode (two polarization) condition can be achieved with a relatively large waveguide. This provides single-mode operations with low propagation loss and compact devices, as we discussed in Paper D.

Nonlinear applications require stronger confinement to increase the interaction of the optical field with the core material. This can be achieved by increasing the thickness of the material deposited. Moreover, when a Si$_3$N$_4$
waveguide has a core thickness above 650 nm, it can display anomalous dispersion (see Fig. 2.10b). This is crucial for the generation of dissipative Kerr soliton combs [47] and was used in Papers C, D, and E. Conveniently, the strong confinement and the large waveguide required for this application also minimize the sidewall scattering loss, as shown in Fig. 2.9b. However, the tensile stress in the Si$_3$N$_4$ presents some technical challenges during the fabrication, that will be presented in Section 3.2. Moreover, thick waveguides support a large number of modes, as displayed in Fig. 2.10a, which is not ideal for linear applications. The single mode condition can be attained at the expense of a dramatic increase in the scattering loss, as we demonstrated in Paper D.

These tradeoffs can be overcome by integrating multiple platforms together with one of the approaches described in Section 1.2.3. There, multiple materials can be combined, or multiple layers of the same material. This approach was used in Paper B, where a strong confinement waveguide is used to minimize the scattering loss at the interface with a slab of high-index material in close proximity to the waveguide core.
Chapter 3

Subtractive process and wafer-level manufacturing

The fabrication of photonic integrated circuits in an academic cleanroom presents different challenges compared to fabrication in a commercial foundry. The process flexibility of a research environment is often combined with the lower yield of the fabrication process due to the constant sharing of the tools with completely different processes. Moreover, it is usual to work with smaller substrates ranging from chips to small wafers (2 or 3 inches in diameter). These can be the only substrates available in case of expensive materials, e.g., InP or LiNbO$_3$, however, the small-scale processing can introduce additional difficulties, such as mounting the samples on carrier wafers to match the format accepted by the machine. This can add process time at best, and unreliability of the process at worst. Hence, to attain a high fabrication yield, it is good practice to use the wafer format that matches the majority of the tools of the academic cleanroom. In the development of this work, all the fabrication has been successfully migrated from 75 mm wafers to 100 mm wafers, which is the standard format of all the machines in our cleanroom except for the Bosch etching.

In this chapter, the relevant steps to fabricate Si$_3$N$_4$ devices are reported highlighting the important steps to move the process from the chip level to the wafer level.

3.1 Overview of the fabrication process

The fabrication of Si$_3$N$_4$ devices is summarized in Fig. 3.1. The fabrication starts with the thermal oxidation of a 100 mm Si wafer to obtain $\sim 3 \mu m$ of high quality SiO$_2$. The high-quality oxide offers a low roughness surface on which the Si$_3$N$_4$ layer can be deposited. The patterning of crack barriers is performed via UV lithography and wet etching of the thermal oxide via buffered-HF etching. The pattern and function of the crack barriers will be further discussed in Section 3.2. After cleaning the wafers with solvents and
RCA cleaning, the Si$_3$N$_4$ film is deposited via low-pressure chemical vapour deposition (LPCVD) with precursors dichlorosilane (H$_2$SiCl$_2$) and ammonia (NH$_3$). The film is deposited to a thickness up to $\sim$350 nm and then annealed at 1100 $^\circ$C in N$_2$ atmosphere to outgas the nitrogen from the film and densify the film [32]. The target thickness is reached with a second deposition if the film thickness exceeds 350 nm.

The devices are patterned on the Si$_3$N$_4$ film by electron beam lithography (EBL), and etched by inductive coupled plasma reactive ion etching (ICP-RIE) with trifluorosilane (CHF$_3$) etchant. After cleaning the wafer with a combination of solvents (acetone and isopropyl alcohol) and RCA cleaning, the wafers are annealed at 1190 $^\circ$C in Ar atmosphere. The cladding is deposited via LPCVD with tetraethoxysilane (TEOS) precursor in steps of $\sim$ 500 nm, after which the film is annealed at 1100 $^\circ$C. This step is repeated to attain the desired thickness of SiO$_2$. At this point, a distinction is made between wafers dedicated to multilayer integration and single-layer devices. The former case is
3.2 Crack barriers

Silicon nitride deposited via LPCVD is considered the best quality material for ultralow loss applications in the short wave IR [27] because it yields the lowest absorption loss when combined with high-temperature annealing (>1100 °C). However, the film is prone to cracks when deposited with thicknesses above 350 nm, due to the high tensile stress [100], \( \sim 1.1 \text{ GPa} \) in our depositions. This limits the fabrication yield, especially when the substrate is larger than 75mm, therefore mitigation strategies have been reported in the literature to minimize crack formation. One strategy is the deposition of silicon nitride with reduced stress by plasma-enhanced chemical vapour deposition (PECVD) [101], sputtering [69], or LPCVD silicon-rich nitride [37]. The low stress guarantees a reduction of the crack formation, at the expense of a lower material quality, which negatively affects the waveguide propagation loss. The low-temperature deposition of PECVD and sputtering however makes these processes compatible with the tight temperature budget of CMOS fabrication, hence they are actively investigated.

Early work on thick LPCVD Si\(_3\)N\(_4\) minimized the cracks by thermal cycling the wafers [32] to reduce the stress in the film, a method later improved with the twist-and-grow technique [43]. This changed after the work by Nam et al. [102] which demonstrated how to control crack formation and a strategy to block crack propagation efficiently. In their work, they explain that cracks are generated by releasing the elastic energy stored by the film stress. To stop the crack, areas with low stress should be introduced in the crack path, so that the propagation encounters resistance and stops when its energy is dissipated. With this new insight on the crack dynamics, the Si\(_3\)N\(_4\) photonics

---

1Phosphoric acid (H\(_3\)PO\(_4\)), nitric acid (HNO\(_3\)), acetic acid (CH\(_3\)COOH)
community started to pattern stress release structures via scribing [33], etching in the Si substrate [42], etching in the silicon oxide layer [45], [103] or partially dicing the wafer [44]. Another approach is to use the pattern etched in the silicon oxide as the core of the waveguide itself in a process reported as the Triplex filled box [41] and perfected in what became known as the photonic Damascene process [35]. This additive manufacturing approach, combined with high-temperature reflow of the SiO$_2$ preform [36] can achieve ultralow losses in thick waveguides, however, it requires polishing of the deposited Si$_3$N$_4$. This additional step can increase the complexity of the fabrication process and can introduce defects such as dishing in structures larger than 3 µm [74]. On the other hand, subtractive processes like the one used in this work, present more challenges in crack management and in attaining low roughness on the sidewalls.

The design of crack barriers in large wafers follows a few simple principles related to the phenomenology of crack formation and propagation:

- The cracks start from the edge of the wafer [42], where the wafer is handled or where it is held on the boat to load it in the LPCVD or
3.3 Wafer scale electron beam lithography

EBL is a technique that allows patterning optical waveguides directly on the wafer by exposing a resist via a scanning electron beam. Unlike DUV lithography, it does not require a mask to be fabricated for transferring the pattern, which is written serially. The slow writing speed is largely compensated by the flexibility and feature size [104], which makes it ideal for research and commercially feasible for prototyping and low throughput services. Indeed, EBL-based PIC manufacturing is offered with an MPW approach by Applied Nanotools Inc. [105].

Wafer-level fabrication of ultralow loss PIC requires careful consideration of the pattern and the writing strategy to achieve low sidewall roughness, small stitching error, and attainable exposure time. The latter factor is important in research facilities, where the EBL is shared with multiple users, and even more important in a commercial setting, where the EBL machine cost can add up to 50% of the total fabrication cost of the wafer.
In EBL, the GDS pattern (see Section 3.4) is converted in machine language by specialized software [106], which divides the geometrical layout in writing fields (WFs), i.e., the maximum area that can be patterned without a stage movement (Fig. 3.3b). Within the WF, the pattern is divided into trapezoids (Fig. 3.3c), and the shots are placed based on the beam step size (BSS) ranging between 2 and 4 nm for photonic devices (Fig. 3.3d). The fracturing of the pattern [107], the shot placement [104] and the BSS [108], play a crucial role in attaining ultralow loss, due to their influence on the sidewall roughness of the waveguide (see Section 2.3). The BSS also determines the writing time of the pattern, which follows the equation [107]:

\[ t_{on} = \frac{A_{device}}{BSS^2R} \]  

where \( t_{on} \) is the time of active beam, \( A_{device} \) is the area of the device and \( R \) is the shot rate of \( \sim 100 \text{MHz} \), limited by the electronic of the tool. The writing time is a crucial factor both for optical loss and for wafer-scale fabrication. Indeed, fast writing and small BSS are proven to reduce the optical loss in waveguides [108]. However, a small BSS requires a larger number of shots to cover the same area, and the writing time increases with \( BSS^{-2} \). Ultralow loss can nevertheless be attained up to \( BSS = 4 \text{ nm} \) [39] by averaging the shot position exposing the pattern multiple times with a decreased dose [86], [108]. In this work, the ring resonators used to generate soliton combs are patterned with \( BSS = 2 \text{ nm} \) (Paper C, D, E), while all the passive devices are patterned.
3.3. Wafer scale electron beam lithography

Achieving high throughput with EBL is fundamental to exploiting the real estate on the wafer and multiplying the number of experiments that can be performed on the same substrate. The BSS is not the only factor that affects the time; most of the time-saving can be done during the pattern design phase, with a basic understanding of the EBL writing strategies. Here we present two topics to consider to have a high throughput EBL run: the subdivision in fields of large devices and an example of how to minimize the writing time by removing unnecessary patterns.

3.3.1 Writing strategy of large patterns

The writing field is determined by the maximum deflection that can be applied to the electron beam and it can vary between $160 \times 160 \ \mu m^2$ and $1024 \times 1024 \ \mu m^2$ in the Raith EBPG 5200 machine available in Chalmers. Typically, a PIC extends over multiple WFs, hence, the wafer needs to be mechanically moved. This introduces stitching errors, which consist of the misalignment of the features in two consecutive WFs. The stitching error is the combination of mechanical errors in the translation stage, distortion of the beam at the edge of the WF, and difference in height of the wafer [109]. The effect of stitching error can be mitigated at the machine level, by calibrating the beam [109] and making sure that the wafer does not present a significant bow or tilt when mounted on the sample holder. When possible, a photonic device should be contained within a single writing field, to completely bypass the stitching error. This is the case for resonators, that can be coiled in a small area [40], but this cannot be done in large complex devices, e.g., an arrayed waveguide grating (AWG). The larger device needs to be subdivided into multiple writing fields. This can be performed manually, carefully planning the stitching error to occur where the device is resilient to misalignment, as proposed by Jin et al. [110].
for DUV lithography. However, this can introduce significant deviation in the performance of the device, where the phase is of interest, such as the array of an AWG. In these cases, it is beneficial to make the device as compact as possible maintaining it within the width of a single writing field. Doing so limits the stitching error to a single direction of the mechanical stage, which we measured to be $\leq 5 \text{ nm}$ on our system for the largest writing field and undetectable for a $500 \times 500 \text{ µm}^2$ writing field. In Fig. 3.4 we show a portion of the 200 nm Si$_3$N$_4$ mask used in Paper D, where large devices are subdivided into fields represented by the different colours. The devices have been designed to extend as much as possible in a single direction to minimize the stage movement in the other direction.

### 3.3.2 Time and placement optimization

When a wafer is exposed we want to fill the whole area with designs to maximize the amount of experiments we can supply to our group, where only a few people are trained to work in the cleanroom. It is therefore essential to optimize the writing time to keep the total exposure time as low as possible. The EBL machine can be left unattended during the exposure, hence long exposures can be planned and scheduled over low occupation time, such as the night or the weekend, leaving the prime day hours to samples with writing time of minutes or hours. In long exposure, when the total writing time is less than 12 hours,
3.4 A foundry approach in academia

A research group working in the field of integrated photonics can have access to a cleanroom facility or be fabless. The latter type has to rely on external foundries to manufacture the integrated circuits designed. This gives the group the freedom to pick a technology among the open-access platforms available [71], basing the choice on the material platform, list of available components, and performance. The chips can be fabricated in the foundry with a multi-project wafer (MPW), where multiple users can share the cost of the fabrication. The MPW approach requires the designer to strictly adhere to the parameters of the fabrication, collected in a process design kit (PDK) which includes all the information a designer needs to create a chip design that can be fabricated in the foundry [111]. Normally, PDK includes building blocks, i.e., components that have been designed and tested on the platform by the foundry. The building blocks are provided with a compact model, i.e., its response in frequency, which

the wafer can be loaded in the machine in the evening at 19 and retrieved in the morning at 8, effectively using the night, while longer exposure might require to be moved during the weekend. For this reason, it is crucial to trim the writing time wherever possible, without influencing the performance of the devices.

Markers are essential for alignment following lithographic steps, as we will further discuss in 4.2.2. They are geometric shapes, typically a square or a cross, used to accurately determine the position, rotation, and deformation of the pattern already present on the wafer. In Fig. 3.5a, we show a marker proposed by our cleanroom team to align the pattern with EBL and other machines. The functional part in our case is the centre square, while the arms of the cross are there to visually find the marker and to be compatible with machines that can detect only crosses. The marker should be patterned with the same strategy as the optical devices to achieve the highest grade of alignment, which we report in Paper D, where 3 markers are necessary to fully correct the rotation, shift, and scaling errors. The marker shown in Fig. 3.5a requires 30 seconds to be fully patterned, and if we use 3 markers per chip this adds up to 2.5 hours of writing time in the pattern shown in Fig. 3.5c, with 90 chips. If we only expose the center square of the marker, the writing time is decreased to 3 seconds per marker, with a total marker writing time of 13.5 minutes. With these types of precautions, we were able to write the wafer Fig. 3.5c in ∼18h with high-quality devices, attaining the results reported in Paper E.

Another important feature to fill the patternable area as much as possible is aligning the pattern with the CB. We use markers patterned with the CB to reference the design position on the wafer, but we don’t perform any machine correction on the pattern, since we noticed that it can introduce additional stitching error artifacts, especially if a low-resolution lithography like UV mask aligner is used. Referencing the position of the marker allows us to minimize the spacing between the CB and the EBL pattern.
allows quick simulation of a photonic circuit without the need for an ab initio simulation of the circuit. Providing compact models that accurately represent the response of the fabricated building block can be done only by repeating the same design in multiple wafer runs to build statistics on the device response. The building blocks allow the designer to compose the PIC by connecting the input and output ports of the building blocks together to attain the desired functionality. Once the PIC is designed, the building blocks are converted into a layout file (GDSII), where the performance can be verified and design rule checks (DRCs) are done to guarantee that the layout can be manufactured with the expected performance. The tape-out phase follows, where the layout is converted as described in Section 3.3. After fabrication, the chips are tested, packaged and sent back to the designer. A fabless group can exploit the PDK of the foundry to engineer circuits with advanced functionalities, e.g., programmable PIC [112] or quantum computing PIC [113].

A research group with a cleanroom facility has increased flexibility in the fabrication process, which is ideal for device engineering [80], but it does not take advantage of the process stability required by a PDK. In the academic environment, full PDK development is impractical because of the necessity of exploring new processes and parameters. However, the process stability guaranteed by a semi-fixed platform opens up for the development of new designs by people normally not involved in the fabrication process and device engineering. In this work, I developed the fabrication process on the 100 mm wafer scale and developed the MPW approach at the laboratory level. I developed the tools that my colleagues, previously inexperienced in PIC design, could use to submit a GDSII file compatible with our fabrication process. A proof-of-concept example of this approach yielded the results reported in Fig. 3.5c, Paper C, and Paper E, where a few users collaborated to design various chips to fill most of the area available on the wafers. The process flow that was developed during the duration of this work is summarized in Fig. 3.6. It follows the MPW approach of standard foundries [111], in a simplified fashion, to accommodate the continuous improvement in designs and fabrication processes. The steps are briefly summarized here.

**Simulation.** The single devices are simulated with numerical methods. The simulation can simply calculate the modes supported by a waveguide and its frequency response to determine the parameters of a ring resonator for frequency comb generation. More advanced simulations using EME or FDTD can be used to determine devices such as couplers, splitters and AWGs. The simulation determines the geometry of the single components.

**Parametric device design.** Single devices can be modeled in a GDSII file, which is a database format that includes the polygons that form the device, including the information about the fabrication step (layer) the polygons refer to. Manually placing the polygons becomes cumbersome for more complicated devices, hence parametrizing the design with a Python-based script allows great control of the point positioning, crucial to attain ultralow loss [108]. We based our scripting on the *gdspy* library [114], which allows us to define polygons and other basic GDS structures with full control of the point positioning. The
single device is designed with a parametric approach. Following the example in Fig. 3.6ii, the parameters are the radii of the rings, the gaps between the resonators and the adjacent waveguides, and the relative position between the two rings. When the parameters are defined, the polygons are generated and the position of the optical and electrical ports are reported. The ports are used to connect to other devices or to the inputs and outputs of the chip. Our implementation is based on a general library for GDSII file generation, but it should be noted that other more PIC design-oriented libraries are available, such as Nazca [115], GDSHelper [116] or the commercially available IPKISS, by Luceda Photonics.

**Circuit design.** The components can be connected together by sequentially adding them to the design using the ports of the previous component as a reference location. Connections between devices can be extended by placing path elements, i.e., waveguides (electrical traces), that take the optical (electrical) signal from one device to the other. The fully composed circuit is then included in a chip structure, where the inputs and outputs are predefined to standardize the chip form factor.
Verification. The GDSII files are inspected by an automated script based on the KLayout’s DRC engine [117]. We established a simple set of rules to check the minimum feature size (>250 nm) and the overlap of polygons fabricated in the same photonic layer. This is crucial to standardize the fabrication and guarantee high fabrication yield.

Tape-out. The GDSII files are processed by separating the different process layers. The photonic layer is processed for EBL exposure (see Section 3.3) and the metal layer is processed for maskless UV lithography. All the chips are composed together in a layout that fits the patternable area of the wafer. The tape-out process is usually done manually for a single chip. However, in an MPW, we can process ~100 individual chips. Hence it is beneficial to automate these steps with a script that executes the necessary steps automatically. The automation reduces the flexibility in postprocessing, but drastically reduces human errors, guaranteeing an improved yield.

Fabrication. The fabrication process is covered in Section 3.1.

Testing. The singulated chips can be characterized optically to determine the loss and the frequency response. More advanced experiments, e.g., frequency comb generation and demultiplexing are usually performed after the initial testing, as reported in Papers C, D, and E.
Chapter 4

Multilayer Si$_3$N$_4$ integration

Multilayer integration proved to increase a photonic platform functionality by taking advantage of the properties of materials for light generation and modulation, or the flexibility provided by one more degree of freedom in the design. Examples of the latter case are introducing a second optical layer to minimize the loss in waveguide crossings and achieve higher integration density [82], [118]–[120], exploit the multilayer system to break the symmetry of the mode’s vertical component [121], [122] for polarization rotation [123], [124], or just improve the mode field diameter for coupling with optical fibers [125], [126]. All these examples consider materials that can be easily deposited without the need for heterogeneous integration strategies, e.g., wafer bonding [127] or micro-transfer printing [65], thereby simplifying the process. However, these techniques must be used when the integrated platform cannot be easily deposited on the existing PIC, as for LiNbO$_3$.

One of the most critical aspects of multilayer integration is to achieve an efficient transition between the layers. This can be achieved with adiabatic coupling, as described in the first section of this chapter. In the second section, the fabrication of Si$_3$N$_4$ multilayer integrated devices is described.

4.1 A coupler for multilayer platforms

Coupling two layers efficiently is done with the aid of a 3D coupler. Early work on these devices used the evanescent field of a waveguide to couple the signal to a ring resonator fabricated on a different layer [128], [129] or directional couplers where the two waveguides are fabricated on separated layers [82]. However, these devices rely on a careful phase matching between the two waveguides so that the desired mode profile is attained at the output of the coupler. This requires a careful evaluation of the $n_{eff}$ of the waveguides. Even small variations due to stress gradient in the layer stack can significantly influence the performance of the coupler as reported in ref. [82]. The directional coupler structure was
abandoned in favor of a double taper design [130], where the top and bottom waveguides are tapered in opposite directions, as displayed in Fig. 4.1. The idea behind this coupler is to slowly taper down the input waveguide to adiabatically decrease its effective index. At the same time, the output waveguide is widened, to increase its effective index. If the two waveguides are close enough, which depends on the mode confinement, the mode will transition from the input waveguide to the output waveguide adiabatically. The double inverted taper coupler is also broadband and experimental demonstration in [27] and Paper D show little wavelength dependency in the selected bandwidths. This is due to the fact that the mode matching condition can happen at any position of the taper to transfer the mode. Therefore, at a longer wavelength, the transition takes place closer to the beginning of the coupler, while at a shorter wavelength, the transition occurs close to the end. The coupling is low loss as long as the mode is well guided in the input and output waveguides while being loosely or not guided in the taper tips. Indeed a large tip introduces extra insertion loss due to the effective index discontinuity at the taper entries, as reported in ref. [130] and observed in Paper B.

The inverse taper is a flexible design that can be used to couple multiple layers [99], [120], [131], and it is crucial to match waveguides with high index core materials low refractive index material, e.g., Si-SiN [101], [132], [133], or coupling Si3N4 waveguide with III-V materials [63], which enables the hybrid integration of lasers.

The double inverted taper can be simulated with an eigenmode expansion (EME) method which provides fast and reliable results for devices that rely on mode evolution [130]. The mode is calculated in different longitudinal positions (see Fig. 4.1) and the mode response of the device is obtained with a transfer
4.2 Multilayer fabrication

The fabrication of a wafer with two layers of Si₃N₄ follows the process reported in Fig. 3.1 until step IX. The extension of the process is reported in Fig. 4.3. After the annealing of the cladding, the surface morphology should be removed to attain a pristine surface for the fabrication of the second layer. Chemical mechanical polishing (CMP) is used to planarize the surface and reduce the roughness of the top surface. After wafer cleaning, a second layer of Si₃N₄ is processed following the steps described in Section 3.1. The two layers are aligned with the SEM included in the EBL to markers fabricated in the first layer. In this work, only thin layers of Si₃N₄ have been deposited (∼200 nm), but there is no fundamental limitation to the thickness of the second layer. After the deposition of the cladding of the second layer, the CMP can be repeated to attain another layer or the process can continue as described in Fig. 3.1, from step X.

Figure 4.2: Tolerance analysis of the 3D coupler. a) Misalignment in the transverse direction b) variation of the oxide layer thickness.

matrix method approach. With this simulation method, we can calculate the response of the coupler for multiple taper lengths with a single simulation. If the cross-section is modified, a new simulation is required. The results reported in Fig. 4.2 are obtained with EME and provide a good picture of the tolerance of the 3D coupler to the process variation discussed in the following section.

In Fig. 4.2a, the tolerance for misalignment of the top waveguide in the transversal direction x is illustrated. We can see that for a coupler 300 µm long the misalignment does not play a significant role in the insertion loss. This regime depends on the specific geometry of the coupler (oxide gap and waveguides cross-section) and could present advantages when the alignment accuracy between two layers is low. However, if we consider the results from the oxide gap tolerance in Fig. 4.2b, we can see that a longer taper should be chosen when the oxide thickness is difficult to attain accurately. These considerations went into the design of the multilayer platform in Paper D.

4.2 Multilayer fabrication
The critical steps to attaining good multilayer devices are the pattern density control during the CMP, the cleaning post-CMP, and the alignment of the two layers. We describe in this section the basics of CMP and alignment for multilayer integration.

### 4.2.1 Chemical mechanical polishing

CMP is a technique where the material is removed from a surface by a joint effort of mechanical abrasion and chemical action to achieve a surface with high planarity and low roughness. The technique was developed in the 1980s by IBM to achieve the surface planarity necessary to eliminate distortion defects in subsequent lithography steps [134]. In a CMP machine (Fig. 4.4a), a wafer is pressed by a carrier head on a polishing pad. The polishing pad rotates and a slurry is added to the pad. The slurry is a suspension of abrasive particles in a chemical solution. The carrier head spins and moves on the polishing pad to achieve a uniform removal rate across the wafer. A diamond-based conditioning pad resurfaces the polishing pad before the process starts and guarantees equal process conditions between runs [135].

The slurry composition carries out the chemical action, which depends on the material to be removed. For example, an alkaline solution is used to polish SiO$_2$ because the presence of OH$^-$ ions favors the reaction that converts siloxane bonds Si–O–Si into soluble Si(OH)$_4$ [134]. This is an equilibrium reaction, which means that mechanical action is needed to remove the material.

The two main mechanical actions that affect the polishing rate are the pad-to-carrier relative speed and the pressure applied on the back of the wafer. These two quantities are used to model the removal rate (R) in Preston’s equation [135]:

$$ R = k_p P v $$

(4.1)
4.2 Multilayer fabrication

Figure 4.4: CMP: a) Schematic of the CMP tool: the polishing pad is placed on a rotating platen and the sample is rotated in the carrier, with the face to be polished on the pad. The slurry is pumped on the polishing pad and a pad conditioner is pushed on the pad to resurface it. b) Blanket removal rate measured on an unpatterned wafer.

where $P$ is the pressure, $v$ is the relative speed between the pad and the carrier and $k_p$ is the Preston’s coefficient, which models all the other mechanical parameters, e.g., slurry particles’ interaction with the material, pad hardness, and roughness. This model requires knowing the exact pressure the material experiences from the pad, which depends on the pattern density of the sample. Since this estimation is rather complicated, a simplified model was proposed by Stine et al. [136], where the surface morphology is divided into two areas as shown in Fig. 4.5a. The up area experiences higher pressure compared to the down area and, hence is polished faster. The model for the residual oxide thickness is known as the MIT model and corrects the polishing rate with the pattern density [136]:

$$\frac{dz}{dt} = -\frac{K_b}{\rho(x, y, z)}$$

(4.2)

where $z$ is the oxide thickness above the structures, following the schematics in Fig. 4.5a, $K_b$ is the blanket rate, i.e., the removal rate on an unpatterned wafer (see Fig. 4.4b) and $\rho$ is the pattern density, which depends on the location on the wafer $(x,y)$ and the residual oxide thickness. This model assumes that the polishing rate at the down surface is negligible until $z = z_0 - z_1$ is reached, hence the pattern density can be written as:

$$\rho(x, y, z) = \begin{cases} \rho_0(x, y) & z > z_0 - z_1 \\ 1 & z < z_0 - z_1 \end{cases}$$

(4.3)

with $\rho_0(x, y, z)$ the initial pattern density. The solution of the differential equation provides an estimation of the residual thickness after a polishing time.
Chapter 4. Multilayer Si₃N₄ integration

Figure 4.5: Polishing rate model. a) schematic of the polishing rate model on a patterned wafer. b) residual thickness after polishing with increasing polishing time. c) zoom of b.

\[
    t:
    \begin{align*}
        z &= \begin{cases} 
        z_0 - \left( \frac{K_b}{\rho_0(x,y)} \right) t & t < \left( \frac{\rho_0 z_1}{K_b} \right) \\
        z_0 - z_1 - K_b t + \rho_0(x,y) z_1 & t > \left( \frac{\rho_0 z_1}{K_b} \right)
        \end{cases} 
    \end{align*}
\]

With this equation, we can calculate the residual thickness for different polishing times knowing the initial parameters of the wafer \((z_0, z_1)\) and the blanket rate, which can be evaluated by measuring the oxide thickness before and after CMP. This measurement is reported in Fig. 4.4b.

If we calculate the residual thickness as a function of the pattern density we obtain the graphs reported in 4.5b and c. We can see that for a short polishing time, there are major differences in residual oxide thickness that depend on the pattern density. Once all the structures are removed, the removal rate is simply given by the blanket rate and the thickness variation is given by the pattern density. This suggests that the pattern density should be homogeneous across the wafer to achieve a consistent residual oxide thickness. The MIT model is considered to not be accurate at pattern density lower than 15% [136], however, it provides an estimation of the polishing time in our process, as discussed in the next section.

Advanced filling structures can be engineered to maintain the pattern density constant, however, determining the area over which the pattern density should be maintained is not an easy task and different strategies can be used, as reported in ref. [137]. In this work, the simple approach of determining the planarization length was used. The planarization length is a pattern density interaction length that describes the deformation of the polishing pad. We
4.2 Multilayer fabrication

4.2.1 Laser lithography

Figure 4.6: a) Process flow on dummy wafers based on Si etching and direct laser writing. b) Polishing time required to achieve a residual oxide thickness of 300 nm with a pattern density of 2.5% and a polishing rate of 70 nm/min as a function of the etch depth and the thickness of the cladding deposited. The white area represents the initial regime where not all the features are removed.

experimentally calculate it to be 3 mm by measuring the polishing rate on a wafer with structures with increasing pattern density, as reported by Ko et al. in [138]. The calculation was done by comparing the MIT model to the measured residual thickness and extending the planarization length until we obtained a reasonable fit.

It should be noted that in our PIC design, the pattern density seldom exceeds 3%, which is helpful because the variation of thickness between pattern density of 1% and 10% is less than 100 nm, as shown in 4.5c. This minimizes the requirements of pattern density control, given the tolerance of the vertical coupler (see Fig. 4.1). Nevertheless, dummy chips with 2% load are exposed around the PIC to homogenize the pattern density in the interested areas. More advanced pattern density control might be necessary to achieve the planarity necessary for chip bonding, as required by the LN integration in Paper B.

4.2.2 Multilayer integrated devices

Fabricating multilayer integrated devices requires extensive calibration of the CMP process to attain the results reported in Paper D. Aiming at performing all the calibration on wafers with Si₃N₄ patterned via EBL presents some time and economical disadvantages. A full 100 mm wafer could take more than 24 hours to pattern with EBL, and it requires extensive processing before and after exposure. For this reason, the CMP process was developed mostly using dummy wafers, i.e., Si wafers without thermal oxide, where a pattern exposed with direct laser writing is etched directly in the bulk Si, as reported in Fig. 4.6a. The dummy wafer provides two significant advantages in CMP process calibration.
First of all, a dummy wafer is cheap and fast to fabricate, and secondly, the lack of buried oxide removes any measurement uncertainty when the residual thickness is evaluated with ellipsometry or white light reflectance spectroscopy. The latter technique measures the thickness of a transparent layer with a known refractive index, based on the reflected spectrum of a white light beam. The small beam size can achieve accurate measurements in close proximity to patterned devices, but it requires a good index contrast to distinguish different layers. These advantages make dummy wafers ideal to evaluate the MIT model by testing pattern-density variation and different etching thicknesses. Moreover, dummies can be fabricated with the same pattern of the wafer with Si₃N₄ PIC, to test how the process behaves on those pattern densities and layouts.

The parameters of the MIT model are the blanket polishing rate $K_b$, the pattern density $\rho_0$, the etch depth $z_1$, and the thickness of the oxide $z_0$ used as cladding and buffer material in the polishing process. When the process is developed, the polishing rate is set by the recipe that provides the smoothest and flattest surface, $\rho_0$ and $z_1$ are fixed by the PIC design, so the only free parameter is the thickness of the SiO₂ buffer layer. To decide the cladding thickness before CMP we can invert Equation 4.4 to calculate the time necessary to achieve a certain residual layer above the waveguides. If we calculate the time to target thickness for different etch-depths and initial oxide thicknesses, we obtain the map reported in Fig. 4.6b. This map gives us an estimate of how long the CMP will take for a given cladding thickness and was used to calculate the polishing time for the sample displayed in Fig. 4.7.

In Fig. 4.7a we report the thickness of the buffer layer before polishing. The white dots are the position where the thickness was measured, close to critical structures, e.g., the couplers presented in Section 4.1. The average thickness of
the oxide is 1705 nm and the etch depth is \(\sim 880\) nm, hence a polishing time of 480 s was required to reach a target thickness of 300 nm. After polishing, the measurement of the residual thickness on top of the waveguides is reported in Fig. 4.7b. Two things can be noticed: the residual thickness is larger than the targeted value and there is a large range of thicknesses on the wafer. The first issue is not crucial in our process, since the coupler described in Section 4.1 can tolerate a large range of oxide gaps. The residual thickness target could be improved with a look-up table of the rate for specific types of wafers, including layout, pattern density, and etch depth. The thickness variation is not a major issue for the same reason, but it could be improved by minimizing the polishing time. Indeed, since the blanket rate Fig. 4.4 is not constant across the wafer, different areas of the wafer will experience different polishing rates. The longer the polishing time, the larger the difference in removed material rates.

Once the wafer is planarized, the second layer can be deposited as described in Fig. 4.3. The markers for alignment are reported in Fig. 3.5 and three of them are patterned on every chip. To minimize the writing time only one of the three has the visual aids, the other two consist only of the center square. The alignment is done automatically with the SEM of the EBL system. Despite initial concern about the detection of the markers, we never experienced any issue in detecting the \(\text{Si}_3\text{N}_4\) marker buried under a thin layer of \(\text{SiO}_2\). The alignment accuracy was measured using a Vernier scale, where the main scale is etched in the bottom layer and the vernier scale is etched in the top layer. The vernier scale has an increasing gap with a 50 nm increment per bar. A scale was placed in each direction to evaluate the x and y shift, however, the best-aligned bars were always the zero-shift bars, suggesting an alignment in the order of 50 nm, or better, as reported in Paper D.
Chapter 4. Multilayer Si₃N₄ integration
Chapter 5

Study case: application to datacom

The last decades’ advances in communication pushed society to be more connected than ever. It is estimated that 66% of the global population has internet access, but the network-connected devices exceed the global population by more than three times [139]. Indeed, nowadays most of our daily actions go through internet-based services, such as online meetings, streaming platforms, social media, and web-based applications. The internet traffic generated by these services is directed to data centers (DCs), i.e., the hubs that collect the servers used for computation and storage, clustered together to maximize the efficiency and utilization of resources [12]. Interestingly, it is estimated that 71% of the internet protocol traffic is intra-DC and another 14% is between DCs [140], which highlights how crucial the DC interconnects are in terms of bandwidth. Moreover, the DC industry is constantly growing thanks to our demand and it has been reported that the DCs in Europe required 2.7% of the European electricity demand of 2018 [141], a number destined to increase in the upcoming years. This is an environmental issue for energy production, but also an economic problem for the DC operator. Therefore, to keep up the pace of the increasing connectivity demand, DC interconnects need to evolve towards greater bandwidth and lower power consumption.

Current optical transceivers

The network architecture of a DC is divided into different layers (Fig. 5.1a) to efficiently and capillarily distribute the data among the available resources [142]. It should be noted that the length of the optical cables in the DC network goes from a meter, within the rack, to a few kilometres for the higher layers and the distribution in metropolitan networks. The data packages are distributed over the network by switches commonly referred to as top-of-the-rack (ToR) switches because of their location in the server racks (Fig. 5.1b). The ToR switch has an application-specific integrated circuit (ASIC) dedicated to the routing operation. From the ASIC, the data are sent to the face panel via
copper traces. Commercially available switches with this configuration have a capacity of 51.2 Tb/s divided into 64x800 Gb/s channels\(^1\) using 512x100 Gb/s electrical lines from the ASIC to the panel. These will become a bottleneck in future generation switches because of the high losses in the transmission. Indeed, it is estimated that an electrical interconnect is limited to 100 Gb/s·m\(^1\) [143]. Hence, decreasing the length of the electrical lines is a crucial step to increase the bandwidth and decrease the power consumption.

On the panel, faceplate-pluggable (FPP) transceivers can be connected. These devices consist of small modules with an electrical and an optical connector on opposing sides. They dominated the market thanks to the flexibility of decoupling the electrical and optical interfaces. Indeed, this facilitates upgrading the speed of the transceiver with the growth of the DC [140]. The initial deployment of FPP transceiver was based on vertical surface-emitting lasers (VCSEL) coupled to multimode fibers (MMF) [144] transmitting at 850 nm. This technology is still in use thanks to the low production cost of VCSEL and the packaging cost of active optical cable based on MMFs [145]. However, the chromatic dispersion and attenuation in MMF limit the reach of high bandwidth transmission to \(\sim 300\) m [146]. This is a problem in hyperscale DCs where connections can exceed 1km. FPP transceivers are still a workhorse of DC interconnects and current implementations are based on multiple InP DFB lasers and SiPh-based modulation and multiplexing. This technology, with more advanced modulation formats, is able to transmit 800 Gb/s over a few km. However, there is no foreseen development beyond 800 Gb/s for this transceiver form factor [140].

The strategies to increase the bandwidth in DC interconnects are based on increasing the number of bits transmitted in every symbol with more advanced modulation formats and increasing the number of channels transmitted

---

\(^1\)https://www.fs.com
via multiplexing. Multiplexing can be based on wavelength, polarization, or space (multiple fibers or cores). On the other hand, energy consumption is minimized by using high-efficiency sources and simple modulation formats, which minimize the components on the receiver side [147]. For these reasons, intensity-modulated (IM) VCSELs with pJ/bit energy consumption and direct detection (DD) schemes have been successful. Indeed, minimizing the components in the optical link is also beneficial in terms of the cost of the link itself, a crucial aspect when millions of servers need to be interconnected. However, longer-reach channels lean more towards the spectral efficiency of the channel itself, allowing for a higher initial and running cost in exchange for an optimized channel efficiency. For this reason, we worked on two possible solutions for DC interconnects, one targeted for short-reach interconnects, within the DC with a span of a few kilometers, and one targeted for long-reach interconnects, which cover the longer transmission between DCs and the metropolitan network.

**Basic concepts of optical communication**

A detailed description of optical communication links is beyond the scope of this thesis, but a basic understanding is required to follow the concepts presented in the next section. The interested reader is directed to ref. [3] and [148].

In an optical communication link, the data can be encoded on the optical signal by modulating the amplitude, i.e., pulse amplitude modulation (PAM), or both the amplitude and the phase, i.e., quadrature amplitude modulation (QAM). The baud rate determines the modulation speed, and it is limited by the physical properties of the laser or modulator. However, the final bit rate depends on the modulation format, indeed a symbol can contain more than one bit of information. However, advanced modulation formats increase the requirement on the receiver in terms of signal-to-noise ratio (SNR). This is exemplified by the bit error ratio (BER), i.e., the number of bits decoded erroneously in the transmission, which can be calculated theoretically from the probability of receiving the wrong bit. To achieve error-free transmission, a limit on the BER is set to $10^{-15}$, i.e., one wrong bit every six hours in a 50 Gb/s transmission. Forward error correction (FEC) is typically implemented to mitigate this requirement to $10^{-3}$ by adding a transmission overhead. This overhead is a known set of bits that help to decode correctly the transmission and decrease the BER to $10^{-15}$ after FEC.

QAM formats require more advanced modulators to attain both phase and amplitude, such as IQ modulators formed by two intensity modulators in parallel with a $\pi/2$ phase shift between them [149]. The receiver side has a more complicated structure due to the requirement of recovering both the intensity and phase of the signal. This can be done by beating the signal with a local oscillator, i.e., a reference laser on the receiver side, which is at the basis of the coherent detection. The QAM modulation format is more spectrally efficient compared to simple PAM, hence it is favorable when the cost of deploying new fibers is significant, e.g., in submarine cables. However, with the current increase in traffic in DCs, a use case is foreseen for long-distance connections, carrying tens of Tb/s per fiber.
The number of bits per symbol in a PAM format is related to the number of levels M by $N_b = \log_2 M$.

## 5.1 Short reach 1 µm interconnect

In Paper A we propose an approach to increase the bandwidth of short-reach interconnects by combining 4PAM modulation, wavelength, and space multiplexing. The architecture and power budget is illustrated in Fig. 5.2 and is based on a Si$_3$N$_4$ photonic platform.

The direct modulation of a VCSEL at 25 GHz provides a data rate of 50 Gb/s per wavelength channel. VCSELs have been demonstrated to operate at this modulation frequency at 1060 nm, dissipating only 100 fJ/bit [150]. This wavelength is particularly interesting because MM fibers display reduced loss (∼1 dB) and dispersion compared to 850 nm [151], which is beneficial for longer links. Moreover, SM fibers are available at this wavelength with slightly higher propagation loss at 1.5 dB/km, with the advantage of avoiding modal dispersion.

The link capacity can be increased with a multi-wavelength array of four VCSELs monolithically integrated [152] and flip chip integration [60] can be used to send the signal to a PIC for multiplexing purposes. The most suitable platform to perform the multiplexing is Si$_3$N$_4$, thanks to the low optical loss in this wavelength region. As explained in Section 1.2.1, Si would be a mature platform for large-volume manufacturing, however, it is not transparent in this region of the spectrum [16]. The PIC in this case is used to combine the wavelengths on a single channel and couple the multiplexed signal to a fiber for transmission. However, more functions could be integrated into the PIC, such as filtering with microring resonators or switching using Mach-Zehnder interferometers, which are components demonstrated in Paper A.

The receiver has an architecture similar to the one described for the transmitter, with photodiodes replacing the VCSELs. However, the polarization of the signal could be rotated upon propagation in the optical fiber, hence a polarization-insensitive receiver should be implemented, i.e., the response of the receiver should be independent of the input polarization state.

Integrating four wavelength channels with 8 nm spacing centered around 1045 nm allows for the multiplication of the bit rate, reaching 200 Gb/s, and by integrating five arrays we can easily reach the Tb/s capacity of the optical link. The PIC could easily fit a pluggable transceiver, thanks to the small form factor of the single devices described in Paper A. Moreover, integrating these components on the same PIC enable the co-packaging of the optical engine next to the switch ASIC, which would minimize the length of the copper traces, and reduce the power consumption of the electrical communication.

To achieve the high transmission capacity mentioned above, the transmission needs to be error-free. In an unamplified link based on 4PAM the BER can be calculated from the optical power on the receiver ($P_{\text{opt}}$), in the high SNR regime obtaining the curves reported in Fig. 5.2b. A detailed description of the parameters and calculation is reported in ref. [153].
The power budget is calculated by considering the minimum optical power on the receiver to attain a BER of $10^{-3}$ before FEC, i.e, -9.4 dBm in Fig. 5.2b for a 4-PAM modulation format. Setting the goal of 2 dB insertion loss per device leaves us with a 3.4 dB to use in the link. This guarantees a transmission length above 1 km with MMF or SMF, and some power to spare for implementation penalties or additional components, such as a polarization rotator to take care of the polarization scrambling in the fiber transmission.

It should be noted that Fig. 5.2b is a correction of the graph reported in Paper A, Fig. 2. Indeed, in the paper we erroneously plotted the BER for a 4PAM 50 Gbaud transmission, making the reported power budget more stringent for our photonic platform.

### 5.2 Long reach 1.55 µm interconnect

The increasing size of DCs and the increasing demand for high bandwidth connectivity between them require solutions beyond the kilometer range. In long-reach transmission, increasing the number of channels in the optical link could decrease the number of cables to be deployed, hence, massive wavelength division multiplexing (WDM) is advisable. If we target a channel bandwidth >20 Tb/s, we can combine WDM, with a coherent modulation format such as 16QAM. Moreover, if the transmitter could be integrated on the same chip, it would provide flexibility in the deployment of the device, which could be...
co-packaged with the control IC [140] and installed inside a switch.

To achieve a high-bandwidth communication link, we envision a frequency comb-based data transmitted with an architecture similar to the one reported in Fig. 5.3a. This includes a frequency comb source, integrated with a demultiplexer, a series of modulators, and a multiplexer to combine the signal and couple it to a single-mode fiber for transmission. A frequency comb-based transmitter could present some unique advantages compared to a laser array, in terms of channel spacing, joint detection of the line, and simplification of the detector, as reported in ref [154]. Moreover, a frequency comb source can be integrated in Si$_3$N$_4$ as well as in other photonic platforms [47], [155].

Even though the schematic might seem straightforward, integrating these elements in a single photonic platform presents unique challenges, highlighted throughout this thesis. The generation of frequency combs requires dispersion engineering of the optical mode, while linear operations like multiplexing and demultiplexing benefit from single-mode waveguides. As reported in Section 2.4 and Paper D, this presents a tradeoff in terms of optical loss. However, similar to what was reported for the 1 µm interconnect, the loss has an effect on the BER of the transmission (see discussion below). Hence, the multilayer approach presented in Paper D could bypass this tradeoff, by
integrating multiple platforms on the same wafer to target specific applications. Moreover, it should be considered that modulation on Si₃N₄ is normally based on thermorefractive effect. For this reason, a material for efficient modulation should be integrated on the wafer. The best material platform in terms of modulation speed, driving voltage, and optical loss at 1550 nm is LiNbO₃ [157], where CMOS-compatible voltages (not process!) [158] and integrated high-speed IQ modulators [159] have been demonstrated. The integration of thin film LiNbO₃ can be done by exploiting the planarization described in 4.2.2 and micro-transfer printing of LiNbO₃ coupons to achieve the hybrid waveguide reported in Paper B.

Similarly to the 1 \( \mu \)m, a frequency comb-based transmitter requires a power budget to attain error-free transmission. A 20 Tb/s capacity can be achieved with 50 wavelength channels modulated at 100 GHz with a 16QAM format. If we consider the average super-efficient frequency comb reported in Paper E, we can attain \( \sim 60 \) comb lines with power above -15 dBm with a pump power (on chip) of 11 dBm. The average power per line is \( P_{av} = -11.5 \) dBm, which is the value we consider here to calculate the power budget. The power budget follows the numbers reported in Fig. 5.3.

The demultiplexing and demultiplexing can be done with arrayed waveguide gratings (AWG), which were demonstrated with <1 dB insertion loss [160] and 40 channels [66]. Here we consider an insertion loss of 2 dB for the multiplexing and 5 dB for the demux. This considers the fact that after modulation the linewidth will be broadened, hence it will experience extra loss due to the response of the AWG.

The insertion loss of the modulators is reported to be <0.5 dB in ref. [157] and \( \sim 1.5 \) dB in ref. [159]. However, the optical power after modulation depends on the symbol encoded, which depends on the modulation format. For this reason, we consider 10 dB loss in the modulators, to overestimate the effect and consider the worst-case scenario.

The optical power of each frequency line is -32 dBm, including the 2 dB loss to couple the signal to an optical fiber. With this power level, we can evaluate the optical SNR (OSNR) following the equations reported in ref. [156]. Considering a quantum noise-limited transmission, the off-chip OSNR is \( \sim 30 \) dB calculated with a resolution bandwidth of 12.5 GHz. Based on the graph reported in Fig. 5.3b, the OSNR would be sufficient to use up to 64QAM, however, the signal should be amplified before transmission, which adds 5 dB of noise due to the noise figure of a typical EDFA. This suggests that 16QAM should be easily achievable, even including further implementation penalties in the system.

This transmitter would make full use of the techniques implemented in this thesis, and it would address the increasing bandwidth demand of the DCs.
Chapter 6

Future outlook

In this thesis, I presented the wafer-scale fabrication of multilayer integrated linear and nonlinear $\text{Si}_3\text{N}_4$ PICs and the development of tools to exploit the benefits of the MPW approach in the research environment. Here, some areas of interest for future developments are highlighted:

- The EBL writing strategy presented here yields waveguides with ultralow loss but relies on small BSS (2-4 nm) and writing fields. The results are slow writing time and stitching errors between main fields. It would be worth to explore alternative fracturing and writing strategies, such as single-line edge smoothing [161] with variable BSS and fixed beam moving stage writing [162]. These techniques could improve the shot placement, reduce the writing time, and minimize the stitching error, yielding an overall better performance of the PIC.

- The evaluation of sub-nanometer sidewall roughness is an important tool to calibrate the fabrication process. SEM cannot provide a quantitative measurement when the roughness is below 1 nm [89], and sub-nanometer roughness is crucial to achieve ultralow loss [86]. A non-destructive technique based on AFM, like the one developed in this thesis (see Section 2.3), could provide important information to develop models and fine-tune the fabrication process. However, extensive development and testing are necessary to consolidate the technique.

- The integration of LiNbO$_3$ modulators on the multilayer $\text{Si}_3\text{N}_4$ platform developed in this work complete the components needed for the data transmitter reported in Fig. 5.3a. However, further development and testing of the CMP process is required. In this work, the planarity after polishing was not discussed but it is a crucial parameter for the heterogeneous integration of long LiNbO$_3$ chips, needed for modulation. It is possible that more advanced pattern density control and filling structures are required for this task. This might present some challenges, especially if the filling structures are patterned with EBL.

- Other devices and materials such as III-V SOA [163], Ge [66] and lasers
[97] could be integrated on the Si$_3$N$_4$ PIC, providing a more complete set of components for system on a chip as envisioned at the beginning of photonic integration in ref. [8].

- The integration of lasers and microcombs has been demonstrated in literature [57], [63], [164] and it has been shown that laser performance can be improved by exploiting an integrated platform [58], [97]. The heterogeneous integration of lasers and super-efficient microcombs could provide improved wall-plug efficiency for telecom applications as well as spectroscopy, timekeeping, and many other fields.

- The coupling between the optical fibers and the PIC is still limited to $\sim2$ dB/facet. The integration of a second photonic layer could provide improved coupling [125], [126], which could improve the overall performance of the PIC and enable coupling to SM fibers, without the need for tapered fibers. Moreover, polishing the facet could provide a better optical interface and improve the coupling.
Chapter 7

Summary of Papers

Paper A

Si₃N₄ photonic integration platform at 1 µm for optical interconnects
Optics Express, Vol. 28 Issue 9 p. 13019-13031 (2020)

In this paper, we present an integration platform based on silicon nitride to aggregate signals in a short-reach (∼2 km) DC link. We studied the power budget for an unamplified link based on VCSELs operating at 1 µm wavelength to attain Tb/s speed over a few km. Based on the power budget we set an insertion loss target of 2dB per device. We experimentally demonstrate edge couplers, AWGs, MMIs, and MZI. We show via 2D-FDTD simulation that a grating coupler with an optimized back reflector could achieve the performance target set in the budget analysis.

My contributions: I performed the power budget analysis, the 2D FDTD simulation of the grating coupler, and I took part in the writing of the manuscript. I presented the work at CLEO/Europe 2019.

Paper B

LiNbO₃/Si₃N₄-Bilayer Vertical Coupler for Integrated Photonics
2020 Conference on Lasers and Electro-Optics, CLEO 2020, STu4J.7

In this paper, we propose a 3D coupler geometry to bypass the scattering losses incurred transitioning between a Si₃N₄ waveguide and a strip loaded LiNbO₃ waveguide. The mode is carried under the LiNbO₃ slab interface in a strong confinement Si₃N₄ waveguide and a 3D coupler is used to adiabatically transfer the light to the hybrid mode in the Si₃N₄/LiNbO₃.

My contributions: I defined the core geometries of the two Si₃N₄ waveguides and prepared the base simulation of the 3D coupler based on two layers of Si₃N₄. The simulation was extended to support the LiNbO₃ slab.
**Paper C**

**Surpassing the nonlinear conversion efficiency of soliton microcombs**  
Accepted for publication in *Nature Photonics* (July 2023)

In this paper, we demonstrate the generation of a single dissipative Kerr soliton in an anomalous dispersion cavity coupled to an auxiliary cavity. The generated soliton has a conversion efficiency one order of magnitude higher compared to standard anomalous dispersion regime devices. The high conversion efficiency regime was achieved by shifting the pump resonance with the auxiliary cavity. The dynamic of the soliton is demonstrated experimentally and numerically.

**My contributions:** I developed the tools to design the devices and I fabricated the samples.

**Paper D**

**Multilayer integration in silicon nitride: decoupling linear and nonlinear functionalities for ultralow loss photonic integrated systems**  
Submitted, under review in *Optics Express*, (Apr 2023)

In this paper, we report a multilayer integrated platform composed of two layers of Si$_3$N$_4$. These allow us to surpass the tradeoffs in terms of the number of modes and optical losses characteristic of single-layer platforms. We demonstrate that the multilayer integrated platform maintains the low-loss performance of single platforms, improving functionalities. We experimentally show the integration of a super-efficient microcomb with a second photonic layer dedicated to routing and demultiplexing the comb lines.

**My contributions:** I optimized the thick platform on 100 mm wafers, developed the thin platform, and developed the multilayer integration process. I analyzed the tradeoffs between platforms, designed the circuits, and performed all the linear measurements. I took part in the nonlinear measurements. I presented part of the work at ECIO 2022 and at CLEO/Europe 2023. I wrote the manuscript with the help of the co-authors. A patent was submitted based on this work.

**Paper E**

**Superefficient microcombs at the wafer level**  
Submitted, (Aug. 2023)

In this paper, we analyze the yield of the generation of superefficient frequency combs, testing 50 devices fabricated on different chips on the same wafer. We demonstrate that two coupled cavities operating in the anomalous dispersion regime can reliably generate a microcomb with an average conversion efficiency
>50%. This showcases both a high fabrication yield (98%) and the robustness of the design. We demonstrate a tri-comb spectroscopy measurement as an example of the new opportunities opened by the integration of a large number of microcombs.

**My contributions:** I developed the tools to design the devices, developed the wafer scale fabrication, fabricated the samples, set up the linear measurements, performed part of the linear measurement, performed part of the nonlinear measurement, and evaluated the results. I wrote the manuscript with the help of the co-authors.
Bibliography


[124] W. D. Sacher, Y. Huang, L. Ding et al., ‘Polarization rotator-splitters and controllers in a Si$_3$N$_4$-on-SOI integrated photonics platform,’ *Optics Express*, vol. 22, no. 9, p. 11167, 2014.


69


