Photonic Integration in Indium-Phosphide Membranes on Silicon (IMOS)

Jos van der Tol, Josselin Pello, Shrivatsa Bhat, Yuqing Jiao, Dominik Heiss, Gunther Roelkens, Huub Ambrosius and Meint Smit

COBRA Research Institute, Eindhoven University of Technology (TUE),

Abstract: A new photonic integration technique is presented, based on the use of an indium phosphide membrane on top of a silicon chip. This can provide electronic chips (CMOS) with an added optical layer (IMOS) for resolving the communication bottleneck. A major advantage of InP is the possibility to integrate passive and active components (SOAs, lasers) in a single membrane. In this paper we describe progress achieved in both the passive and active components. For the passive part of the circuit we succeeded to bring the propagation loss of our circuits close to the values obtained with silicon; we achieved propagation loss as low as 3.3 dB/cm through optimization of the lithography and the introduction of C_{60} (fullerene) in an electro resist. Further we report the smallest polarisation converter reported for membrane waveguides (<10 μm) with low-loss (< 1 dB from 1520-1550 nm), > 95% polarisation conversion efficiency over the whole C-band and tolerant fabrication. We also demonstrate an InP-membrane wavelength demultiplexer with a loss of 2.8 dB, a crosstalk level of better than 18 dB and a uniformity over the 8 channels of better than 1.2 dB. For the integration of active components we are testing a twin guide integration scheme. We present our design based on optical and electrical simulations and the fabrication techniques.

Introduction: The IMOS platform is based on a high refractive index contrast InP membrane, which is optically very similar to a silicon membrane and suitable for the creation of high-density, low-power PICs. In the long term, we want the full functionality of classical InP-based PICs to be integrated in IMOS, using a set of standard building blocks. The strength of the IMOS concept resides in its inherent ability to integrate active and passive functions. We follow the same philosophy as the Generic Integration approach for in classical InP-based PICs [1] as illustrated in figure 1. With a limited set of basic building blocks and a generic process we are able to make a great variety of A(pplication) S(pecific) P(hotonic) I(ntegrated) C(ircuits.
Also in IMOS we want to develop a Generic Integration Process in which active and passive components can be integrated.

Figure 1: Comparison of CMOS building blocks and Photonic building blocks and an impression of the basic building blocks in bulk InP

Figure 2: Schematic composition of IMOS chip

Figure 2 represents the typical dimensions and composition of an IMOS waveguide. InP \( (n = 3.17) \) is chosen as the high refractive index wave-guiding material, because of its ability to integrate lattice-matched active and passive materials. The low-refractive index \( (n = 1.5) \) polymer Benzo-Cyclo-Butene (BCB) is used as an adhesive to bond the InP membrane to a Silicon wafer. The thickness of the InP membrane is chosen as 250-300 nm to obtain a strong light confinement, while a bonding layer thickness of 1.9 \( \mu \text{m} \) enables to decouple the InP membrane from the high-refractive index silicon wafer both optically and thermally. The InP membrane can also contain an active layer for active components.

This configuration brings a number of advantages. As demonstrated in the Generic Integration process for classical InP-based PICS, the use of InP and related compounds as the guiding material simplifies the integration of passive and active functions on the platform. Thanks to the high vertical refractive index contrast created in IMOS, light is tightly confined...
in the InP membrane. This allows for very compact photonic devices to be realized in the membrane. Consequently, circuits in IMOS can be made very dense, and with low power consumption. The basic building block for light guiding is a single-mode waveguide with a cross-section of 250 nm x 450 nm, more than one order of magnitude smaller than single-mode waveguides in classical InP based photonics. Furthermore, the use of BCB as the bonding material brings a relatively high degree of flexibility with regard to the carrier wafer composition and topology. An important property of BCB is its low thermal conductivity, which is an advantage for future integration on CMOS circuits with dynamic hot spots. Here the low thermal conductivity of BCB allows for good thermal decoupling between the InP and the CMOS layers, when double-sided cooling is applied. In terms of integration of an IMOS optical interconnect on top of an electronic IC, this means that back-end processing can be used, limiting the interference with the CMOS fabrication to the connections between IMOS and CMOS devices using vias through the BCB layer. Once the platform architecture has been chosen, the next step is to design a set of basic building blocks, which can be combined to realize complex functions in PICs. Ultimately these building blocks should be brought together in a single generic integration process, optimized for providing high performance for all the building blocks.

**Wafer bonding:** As described above the IMOS technology is making use of wafer bonding using BCB to intimately combine CMOS electronics with III-V photonic functionalities. Below the process flow for the wafer bonding step is given. The first step in all fabrication schemes is the fabrication of an Epi wafer with the active and passive areas on the wafer created with one or more epitaxial growth steps. Here, the goal is to have a generic integration process in which all the building blocks can be made in an identical layer stack in which we can integrate active as well as passive components on one chip. In the case of a passive photonic chip the device layer stack consists only of a 250-300 nm thick InP layer on top of a thick InGaAs protection layer, used for selective substrate removal. A more detailed description of the integration of active devices in the platform will be given in section “Integration of Active Components”.

Once the Epi wafer is ready, the processing of the IMOS chip can start, which is schematically illustrated in Figure 3.
The first step (Fig 3a) is the bonding of the InP layer upside down on the Silicon or CMOS carrying wafer with a BCB polymer. If necessary the bottom side of the photonic layer can also be processed prior to bonding, allowing for more complex device designs in the future. To promote adhesion both the carrier wafer and the III-V wafer are coated with SiO$_2$ before bonding. After the bonding and baking of the polymer, the InP substrate is removed by consecutive selective wet etching of the InP substrate and the InGaAs protection layer, the result of which can be seen in Figure 3 b). Finally, as indicated in Figure 3 c), waveguide and grating structures can be defined in the photonic layer using e-beam lithography (EBL) with a ZEP resist and reactive ion etching.

**Propagation loss:** Low propagation loss is crucial for any platform technology. Any roughness created during pattern definition or during etching will severely impact the loss experienced by modes in the high confinement waveguides. A straightforward method to determine these losses is by comparing straight waveguides with varying lengths. In Fig.4, we show the measured transmission through ten IMOS waveguides with length variation of 100, 300, 500 and 700 μm. The waveguides have 400 nm width and 250 nm height. The fitting of the data yields a propagation loss of 3.3 dB/cm$^{-1}$ and a waveguide-fiber coupling efficiency of about 25%. The low value of the propagation loss is obtained by optimizing the electron beam lithography with a resist containing C$_{60}$ (fullerene) [2, 3].
Polarization converter. Polarization handling is a fundamental issue in photonic integrated circuits. Polarization can be taken advantage of, e.g. for light intensity modulation or polarization bit interleaving [5]. For all these applications, an efficient broadband polarization converter is the key component. Here we introduce a polarization converter in IMOS based on triangular waveguides for a short device, but optimized for high tolerance to fabrication errors. This device is the world’s smallest InP polarization converter made to date (0.4 μm × 0.8 μm × 7.5 μm). The device, depicted in fig. 5, consists of two triangular sections of about 2 μm long, realized by a combination of a wet etch, for the sloped side, and a reactive ion etch, for the straight side. Since the gratings used for input and output coupling are optimized for TE-polarization the characterization has to be based on the transmission of the TE-polarized mode. This is done by including between the triangular sections a segment of a straight waveguide. Variation of the length of this segment results in constructive and destructive interference of the converted contributions of each triangular section, and thus in respectively full polarization conversion and zero polarization conversion.

Figure 5: Designed (left) and realized (right) polarization converter structure [6, 7]
Measurement results for sets of devices with different straight waveguide segments are given in figure 6. Fitting an interferometric curve to these data yields the performance parameters of the polarization converter: conversion efficiency of 99.2 ±0.2% (at $\lambda=1530$ nm), and an insertion loss below 1 dB.

![Graph](image)

Figure 6: Transmission through polarization converters, as a function of the length $L_{\text{rec}}$ of a straight segment between the triangular sections [7]

Similar results are obtained with lengths of the triangular sections varying between 2.1 and 2.6 $\mu$m, which demonstrates the tolerance of these polarization converters. The bandwidth over which the conversion is $>92\%$ is 35 nm, which is sufficient for use over the whole C-band. By removing the rectangular section a bandwidth of 200 nm, with a conversion $>98\%$, can be expected according to simulations.

Wavelength demultiplex. One of the fundamental functions required is wavelength (de)multiplexing. Using demultiplexers, the different signals propagating in a given waveguide can be separated (demultiplexed) into different output waveguides according to their wavelength. Conversely, demultiplexers allow signals of different wavelengths propagating in different waveguides to be combined (multiplexed) into a single waveguide. Here we present a planar concave grating (PCG) wavelength demultiplexer realized in IMOS [8].
A PCG demultiplexer (also referred to as Echelle grating, or etched diffraction grating) functions by combining the high dispersion of a large period grating, with the focusing power of a concave mirror. As shown in Fig.7, the light coming from the input waveguide spreads in an un-etched free-propagation region (FPR) and reaches the PCG, where it is simultaneously reflected and diffracted by the grating corrugation, and re-focused by its curvature. Due to the inherent dispersion of the grating, different wavelengths are diffracted in different directions, and can therefore be collected separately by several output waveguides placed on the so-called Rowland circle (see [9] for more detail on the design and behavior of PCG demultiplexers). An eight-channel PCG demultiplexer was designed for TE-polarized light, with a central wavelength of 1550 nm, and a channel spacing of 4.0 nm.

Fig.8 shows the measured transmission spectra of the device. The channel spacing of the device (3.96 nm) is close to the design value of 4.0 nm. The spectrum is blue-shifted by 4.1 nm with respect to the designed spectrum, and the insertion losses (2.8 dB) are higher than the value predicted in simulation (0.2 dB). The shift is caused by a deviation of the membrane thickness from the design value. The extra insertion loss arises due to fabrication imperfections in the PCG's
DBRs. The sidelobes beside the main channels originate from the phase noise created by membrane thickness non-uniformity, and by fabrication imperfections. However, their transmission level does not exceed -21.3 dB, meaning that the cross-talk figure for the device is better than -18 dB. Finally, the power non-uniformity between the transmission of the different channels is below 1.2 dB.

**Optically pumped laser**

So far a number of passive devices realized on an InP-membrane are presented. Active devices are however needed to develop a full platform. To test the quality of processing and materials a DFB-laser structure is developed (see fig.9). This device was realized in a membrane containing a compressively strained quantum well.

This laser was done with pumping from the top in a DFB-type of structure, with the grating etched on top of the waveguide.

![SEM picture of the optically pumped IMOS-POLIS DFB laser. Insets: layer stack and detail of the DFB grating.][11]

With this structure the first laser emission from an IMOS laser is obtained. Figure 10 shows the laser peak and the input-output characteristic, clearly indicating the laser behavior.

After correction for losses in the optical path for the pump light it is found that the threshold pumping power is 6 mW. The next step will be to develop an electrical pumped laser. Based on the results of the optically pumped laser structure a threshold current of 2.5 mA can be expected there.

To combine active and passive structures an integration concept is being developed, which will be presented in the next section.

Integration of Active Components

A key functionality for almost any photonic integration platform is the generation, amplification and detection of light. Devices which perform these functions are so-called active devices, and their integration with passive devices is at the heart of the integration platform. For the integration of active components we develop a scheme based on a twin-guide active waveguide structure. In the following we present the device design based on optical and electrical simulations, as well as details on the device fabrication.

Semiconductor Optical Amplifier in twin guide structure. The twin-guide approach consists of two vertically stacked layer sequences for the active and the passive functions, which are schematically represented in in Figure 11a). After bonding the layer stack is as follows starting from the lowest layer: (1) 300 nm thick InP passive waveguiding layer on top of the BCB/SiO₂ bonding layer; (2) 100 nm thick n-contact layer formed by a n⁺-doped quaternary layer (Q1.25) and a n-doped InP layer; (3) bulk active region formed by 250 nm thick nominally undoped Q1.58 ; (4) p-doped InP cladding layer; (5) contact region formed by quaternary spacer layers and highly p⁺-doped InGaAs. In the passive regions, layers (2)-(5) are removed by a combination of dry and wet etching leaving a high quality surface for the definition of low loss waveguide structures. To define an amplifier 700 nm wide ridges are defined in layers (3)-(5), as sketched in Figure 11a). The p-side of the diode structure is contacted using a Ti/Pt/Au metallization on top of the ridge,
while the n-layers are contacted with Ni/Ge/Au pads parallel to the ridge with a few micrometer separation. A schematic 3D representation of an amplifier and a top view is shown in Figure 11b). We taper the ridges to transfer the optical mode from the passive waveguides to the active region of the amplifier and back. The optical properties will be discussed in more detail below. The choice of the polarity of the lower contact is motivated twofold: firstly, when transferring the mode from the active to the passive waveguiding region through the lower contact the low optical loss due to n-doping is lower than in p-doped layers. Secondly, the higher mobility in n-doped materials results in a lower resistance in the thin contact layer between the contact pads and the amplifier ridge.

![Diagram of active and passive regions of the amplifier](image)

**Figure 11:** a) Layer-stack of active and passive regions. b) Schematic representation of the amplifier.

**Simulated active device characteristics.** We performed optical and electrical simulations to optimize the device design. Figure 12a) shows the electric field profile of the fundamental mode in a 700 nm amplifier ridge calculated with a mode solver (Lumerical). The mode shows TE characteristics and a loss of 54/cm, which is mainly dominated by the interaction with the metal of the top-contact and a confinement factor of 0.55 for the optical mode in the bulk active material. With a width of 700 nm the loss in the waveguide is minimal, while guaranteeing a single transversal mode. To estimate the material gain in the Q1.58 layer we performed electrical simulations using a self-consistent Poisson-solver (nextnano++) with a 1-dimensional model. Additionally, we calculated the series resistance of the ridge structure with 770 Ωµm using a 2D-finite element model. Here, a contact resistance of 1e-6 Ω/cm² was considered for the n- and p-contact. The band structure of the active layer stack under a bias voltage of 1.0 V is presented in figure 12c). The
corresponding I-V characteristics of the diode including the series resistance of the contacts is plotted in Figure 12d) as black line. We calculated the dependence of the Fermi energies in valence (Fig 12c, red line) and conduction band (Fig 12c, blue line) as a function of the current density up to 45 kA/cm². Carrier recombination via Auger (7e-29 cm/s, [17]), radiative (0.98e-10 cm3/s, [17]) and surface recombination (1e5 cm/s, [16]) is considered in the model. Using the values for the Fermi energies we calculate the optical gain with Fermi’s golden rule according to reference [18] using a momentum matrix element of 2|M²/m₀=25.3eV and a temperature of 300 K. The material gain at 1550 nm is plotted in Figure 12(d) as a function of the current density. To compensate the loss in our amplifiers, we need a material gain exceeding 98/cm, which we estimate for a current density of 6.8 kA/cm² at a voltage of 1.05V.

As described above tapered sections are used to transfer the mode from active to passive waveguide sections. We estimated the minimum length and losses of such tapers using 3D-FDTD calculations (Lumerical). Here we chose a critical dimension of 200 nm at the tip, which is easily achievable with e-beam lithography. The result is shown in figure 12(d), where we plot the transmission through the taper as a function of the taper length. A transmission of 0.95, the equivalent of 0.5 dB loss, is reached for tapers as short as 10 µm.

Figure 12: Simulation results for the amplifier structure. a) Normalized electric field of the optical mode. b) Band-structure of the diode layerstack. c) I-V characteristics and material gain at 1550 nm. d) FDTD- calculation of the transmission in the taper.
**Ring laser example.** In the following we will discuss the ring laser schematically presented in figure 13a) to give an example for an active device using the amplifier described above. The laser consists of an amplifier section with variable length, which is embedded in a racetrack cavity, where the bending radius is set to 50 µm to allow sufficient space for electrical contacts. The waveguide exhibits propagation loss of 3.3 dB/cm, as discussed above. Each taper adds an additional loss of 0.5 dB to the cavity. The light is coupled out of the cavity using a 50/50 multimode interference coupler, where we assume an additional insertion loss of 2 dB. Figure 13b) shows the threshold gain and threshold current as a function of the amplifier length calculated from the combined round trip loss. Realistic threshold material gains below 500 /cm are reached for amplifiers exceeding 40 µm in length. The best figure of merit to optimise the amplifier length is the wall plug efficiency of the laser, which is defined as the ratio of optical output power over electrical power consumed. We use the results of our electrical simulations described above to calculate the efficiency as a function of amplifier length for a combined output power of 1, 10 and 50 mW in both output waveguides. The result is presented in Figure 13a). For a low output power of 1 mW an efficiency of 14% at a length of 20 µm is reached. For higher output powers of 10 mW or 50 mW the efficiency exceeds 25% for a length of only 50 µm and 120 µm, respectively. This result demonstrates that extremely efficient and compact laser designs are possible using the IMOS platform.

![Ring laser schematic](image)

**Figure 13:** a) Schematic representation of a ring laser. Intra-cavity losses are indicated. b) Threshold gain and current of the laser shown in a). c) Wall plug efficiency for combined output powers of 1, 10 and 50 mW.
Detector. The active layer stack designed for the amplifier can also be used for a fast and efficient detector, thanks to the use of a bulk active region with high confinement factor and high absorption coefficient on the order of 4000/cm at 1550 nm. As in the case of the amplifier, the active region of the detector is designed as a 700 nm wide ridge. Here, the taper is chosen to be only 5µm long. This is a good compromise between efficient coupling of light from the passive waveguide to the active region and a low footprint to reduce the capacity of the device. The schematic diagram of the designed detector is shown in Figure 14 a). We calculated the responsivity using 3D-FDTD as a function of the length of the detector, which is plotted as squares in Figure 14 b) on the left axis. Responsivities exceeding 0.9 A/W are predicted for detectors longer than 15 µm. We can estimate the cut-off frequency response of the detector with \( f_c = 1/(2\pi R_{\text{det}} + R_{\text{load}})C_{\text{det}} \), where \( C_{\text{det}} = 0.3 \, \text{fF/µm} \) (\( R_{\text{det}} = 770 \, \Omega \text{µm} \)) is the capacity (resistance) of the detector and \( R_{\text{load}} \) is the load resistance of a receiver circuit. The frequency response \( f_c \) as a function of detector size is also shown in Figure 14 b) for \( R_{\text{load}} = 50 \, \Omega \text{ (red)} \) and \( R_{\text{load}} = 1\,\text{kΩ (blue)} \). For \( R_{\text{load}} = 50\,\Omega \) the receiver speed is not limited by the RC constant. Even for a load resistance of 1 kΩ, which is more realistic for integrated receiver circuits [19], \( f_c > 10 \text{ GHz} \) is feasible.

**Figure 14:** a) Schematic representation of a detector. b) Detector responsivity and cut-off frequency as a function of length.

**Fabrication.** The key fabrication steps for the twin-guide SOA/detector structures as well as the passive waveguides are depicted in Fig. 15. The final layerstack of the InP wafer before processing is shown in Fig. 11(a). The fabrication starts with the flip-chip BCB bonding of the III-V layerstack onto a SiO₂/Si carrier wafer, followed by removing the InP substrate and the InGaAs sacrificial layer wet-chemically. After bonding, a 50-nm SiNₓ hard mask layer is deposited on top of the sample by PECDV. The device pattern is defined by using EBL with C₆₀/ZEP resist. The defined pattern from C₆₀/ZEP resist layer is then transferred to SiNₓ layer by means of CHF₃ RIE. The pattern is finally transferred to the III-V layerstack by means of CH₄/H₂ RIE. The etching stops when the n-doped InP layer is reached (Fig. 15a)). The region for making n-type metal contact is defined by a second EBL. The SiNₓ will protect the region where n-InP and n-Q1.25 material are preserved for making n-contact. The n-type material at all the other regions will be removed wet-chemically (Fig. 15b)). Two more EBLs will be performed to realize passive waveguides and fiber-grating couplers with etch depths of 300nm and 120nm respectively, as shown in Fig. 15c). During the final metallization step, the entire sample is
covered by a 50-nm thin SiN₄ layer for electrical isolation and planarized by polyimide. Then two lift-off processes using optical lithography are performed to create p and n metal contacts, as shown in Fig. 15d).  

![Fig. 15 a) The first EBL step to define the SOA structure on the bonded sample. b) The second EBL step to define the n-contact region. c) The third and fourth EBL step to print the passive waveguides and fiber-grating couplers. d) The lift-off step to create p and n metal contacts.](image)

One of the advantages of this fabrication scheme is that all processing is performed after bonding. This ensures a reliable and high-quality bonding, which reduces fabrication difficulties significantly. Another advantage is that this scheme makes the monolithic integration of SOAs and detectors straightforward. The detectors can be processed in parallel with the SOAs as they share the identical layer stack. The critical steps of defining the SOA ridges and the n-contact pads have been successfully tested. The SEM picture in Figure 16 a) shows the fabricated tip of the taper structure, which is sharper than required. As shown in the simulations above the taper coupling efficiency can be as high as 95%, as long as the taper tip width is less than 200 nm. Additionally it is visible that the SOA sidewall is vertical and smooth resulting from an optimized InP dry etching process. The SEM picture in Figure 16 b) shows the definition of the n-side contact regions. The picture was taken right before the wet etch of the n-material in exposed regions. As can be
seen from the figure, the desired n-contact region is well protected by the SiNx layer. Moreover, the entire taper structure is also well protected by the SiNx to prevent any damage to the sidewall of the taper during wet etch.

Fig. 16 a) The SEM picture of the fabricated taper tip structure of the SOA. b) The SEM picture of the SiNx protection layer for the n-side contact definition. The taper tip is also protected by SiNx.

Conclusion
We presented a novel photonic integration scheme based on a thin III-V photonic membrane bonded to a silicon or CMOS carrier chip. We demonstrated low loss passive components including polarisation converters with efficiencies in excess of 95% and wavelength demultiplexer devices with an insertion loss of only 2.8 dB. We presented a scheme for integration of active devices in the III-V photonic layer with a twin guide structure. Based on electrical and optical simulations we expect compact and efficient laser structures, as well as fast detectors with high responsivity. Fabrication is underway.

Acknowledgement:

This work was done in the following projects: ERC grant “No Limit”, “On chip Networks” (TUE internal project) and STW project no. 11361, “IMOS”.
Experimental clean room work was performed in NanoLab@TU/e, the clean room facility of the Eindhoven University of Technology.

References:


