Nanophotonic integration in state-of-the-art CMOS foundries

Jason S. Orcutt,1,2* Anatol Khilo,1 Charles W. Holzwarth,1,4 Milos A. Popović,1,5 Hanqing Li,2 Jie Sun,1 Thomas Bonifield,3 Randy Hollingsworth,3 Franz X. Kärtner,1 Henry I. Smith,1 Vladimir Stojanović,1,2 Rajeev J. Ram1,2

1Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, Massachusetts 02139, USA
2Microsystems Technology Laboratories, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, Massachusetts 02139, USA
3Texas Instruments, Dallas, Texas 75243, USA
4Currently with the Department of Electrical Engineering, University of Canterbury, Christchurch, New Zealand
5Currently with the Department of Electrical Engineering, University of Colorado at Boulder, Boulder, Colorado, USA

*jsorcutt@mit.edu

Abstract: We demonstrate a monolithic photonic integration platform that leverages the existing state-of-the-art CMOS foundry infrastructure. In our approach, proven XeF₂ post-processing technology and compliance with electronic foundry process flows eliminate the need for specialized substrates or wafer bonding. This approach enables intimate integration of large numbers of nanophotonic devices alongside high-density, high-performance transistors at low initial and incremental cost. We demonstrate this platform by presenting grating-coupled, microring-resonator filter banks fabricated in an unmodified 28 nm bulk-CMOS process by sharing a mask set with standard electronic projects. The lithographic fidelity of this process enables the high-throughput fabrication of second-order, wavelength-division-multiplexing (WDM) filter banks that achieve low insertion loss without post-fabrication trimming.

©2010 Optical Society of America

OCIS Codes: (230.7370) Waveguides; (250.5300) Photonic integrated circuits; (200.4650) Optical interconnects.

References and links


1. Introduction

If nanophotonic devices and systems could be fabricated using state-of-the-art CMOS processes, with their attendant lithographic fidelity, process control and throughput, a major barrier to integration of photonics and electronics would be eliminated, possibly leading to widespread utilization of their complementary features [1–4]. For electronic-photonic integrated circuits to have maximum impact it is important that the integrated transistor performance and density are equal to state-of-the-art electronics processes. Additionally, no front-end photonic integration solution has yet been proposed for bulk-CMOS processes that...
comprise 92% of CMOS logic production on 300 mm wafers. To date, most silicon-based photonic systems have employed non-standard silicon-on-insulator (SOI) starting wafers in which the buried-oxide (BOX) thickness is an order of magnitude larger than is used in SOI-CMOS processes [5–10]. The thicker BOX degrades the performance of deeply-scaled transistors via drain-induced barrier lowering (DIBL) [11,12]. The resulting low switching-current ratios may prevent sub-45 nm gate-length transistor integration in such a platform [13]. Additionally, the thermal impedance of the thicker BOX limits electronic integration density by reducing the power budget [14]. A previously proposed, non-monolithic solution is to stack a separately fabricated photonic layer on top of the electronic circuit [15–17]. Such a 3D platform also limits electronic power densities by adding thick insulating layers into the thermal path and increases process complexity. Other monolithic integration work to integrate photonics in both bulk- and SOI-CMOS processes has focused on modifying the back-end interconnect stackup to include separately deposited waveguiding, detection, and modulation materials [18–22]. However, the specialized wafer-level processing required prohibits the direct use of standard electronic foundry flows, increasing the process cost and complexity.

In this work, we demonstrate a monolithic front-end photonic-integration platform within a state-of-the-art 28 nm bulk-CMOS foundry process. Our approach avoids modifying any in-foundry processes and adds post-processing to locally remove the Si underlying the photonic devices, thereby eliminating optical-coupling to the Si substrate and its associated loss. By complying with all electronics industry design submission practices, we directly use the existing infrastructure as a normal foundry user. This demonstrates the fabless model that has been established as a goal for the silicon photonics community [23]. The generality of this approach makes it suitable for integration within existing bulk- and thin-SOI-CMOS foundry processes.

2. Platform overview

In our platform, no modification is made to the in-foundry CMOS process, and the performance of the included state-of-the-art electronics is not compromised. As first demonstrated in [24], our monolithic front-end integration platform enables electronic-photonic integrated circuit (EPIC) fabrication using the same low-cost foundry infrastructure that has been developed for electronic circuit prototyping and production. The nanophotonic devices demonstrated in this work were integrated alongside over a million transistors into the 2.2×2.0 mm test chip, shown in Fig. 1a. On the process development wafers used for this work, not all of the transistor source/drain doping steps achieved required targets to enable electronic circuit functionality.

Photonic devices were defined using the standard electronic process design kit (PDK) layers in Cadence Virtuoso, a common VLSI electronics CAD environment, as described in [25]. Utility design layers, which modify the default foundry data processing of the submitted layout, were inserted to exclude the photonic regions from optical-proximity correction (OPC) during standard data preparation. Our design shared a 33×26 mm mask set and all in-foundry processing with standard electronic projects in an unmodified Texas Instruments 28 nm bulk-CMOS process on 300 mm wafers, as shown in Figs. 1b and 1c, respectively. By
Fig. 1. (a) Optical micrograph of 2.2×2.0 mm photonic die fabricated in a 28 nm bulk-CMOS process containing 384 optical test ports and over a million transistors. Integrated front-end photonic and electronic features are exposed by silicon substrate removal and back-side imaging. The photonic die shared a 26×33 mm reticle set with standard electronic projects shown in (b) and was fabricated using the standard process flow on a 300 mm wafer shown in (c).

mask sharing with electronics, the prototype cost was reduced by leveraging the standard CMOS economies of scale.

In the SOI platform, photonic devices can be built in the single-crystal Si layer. This is attractive due to the relatively low optical loss of single-crystal Si. Bulk-CMOS, on the other hand, does not provide a patternable single-crystal Si layer for photonics. Therefore, we use the polysilicon, deposited for the transistor gates and local electrical interconnects, as the high-index waveguide core. The layer thickness of roughly 80 nm yields a suitable strong confinement core for the transverse-electric-polarized light from 1.2 μm to 1.6 μm. Due to the thin core layer, transverse-magnetic-polarized light is not well guided for single-mode waveguide geometries. Because default doping and metallization steps introduce optical losses greater than 1000 dB/cm, we employ a combination of design layers, available in the standard CMOS process flow, to locally block such processes for waveguide formation. A second problem is that the oxide below the polysilicon layer, known as the shallow trench isolation (STI) in both bulk- and SOI-CMOS processes, as well as the BOX below the single-crystal Si layer in modern SOI processes, are all thinner than 400 nm. These thin undercladding layers would cause leaky optical modes in these front-end waveguide structures, with propagation losses in excess of 500 dB/cm [5]. To circumvent this problem,
we use post-foundry processing to locally etch out the Si underlying the photonic devices [26,27].

To locally remove the Si substrate underneath the SiO$_2$ layer on which the photonic devices were located, vias were etched from the top surface of the chip down through the dielectric stack to the Si, which was then etched using XeF$_2$. A 10 μm-thick layer of photoresist was spun on the chip and rows of holes, each measuring 10×10 μm, were exposed using contact photolithography. These holes were aligned to in-process dielectric windows where the standard metal fill was excluded adjacent to the photonic devices as shown Fig. 2a and Fig. 4. Reactive-ion etching in CF$_4$ gas, at a bias of 250 V, etched through the SiC, Si$_3$N$_4$, SiON and SiO$_2$ layers of the dielectric stack. To prevent overheating of the photoresist, the 2 hour total etch time was broken up into 5 min segments with 5 min breaks in between. This long processing time can be reduced significantly by using a more powerful etch system such as the inductively-coupled plasma (ICP) etchers with back-side cooling that are common in CMOS foundries. Once the etch reached the Si substrate, the photoresist was removed in acetone. The exposed Si on the backside and sides of the chip was then coated with a protective layer of Crystalbond 509 leaving only the Si at the bottom of the vias accessible. The chip, mounted to an oxidized 100 mm Si wafer for thermal management, was then placed in a chamber that supplied XeF$_2$ gas to isotropically etch the Si, removing it as the volatile product SiF$_4$. Etch selectivity of over 1000:1 allowed the thin STI SiO$_2$ layer to act as an etch mask for the undercut. A pulse etch technique was used, where etch steps of 10 s were

![Diagram](https://example.com/diagram.png)

**Fig. 2.** (a) Optical micrograph showing relevant dielectric window openings for the dielectric etch as well as optical access. (b) Cross-sectional scanning electron micrograph (SEM) of die after localized substrate removal in the photonic region. To demonstrate the film planarity and stability, the undercut shown here is roughly five times wider than required. A die-saw was used to section the processed chip through the undercut region resulting in the rough CMOS layer stack edge.
interleaved with 50 s steps to pump out the reaction products. The undercut shown in Fig. 2b took approximately 430 s of etching time, which corresponds to a Si etch rate of about 315 nm/s.

The localized nature and low temperature (less than 200 °C) of this process ensure that the electrical performance of the neighboring electronic regions is unmodified. Placing a ground-connection guard ring at the perimeter mitigates any impact that the substrate discontinuity may have on the electronic substrate while allowing electronic devices within 5 μm of the photonic region. The total undercut width of 271 μm that is shown in Fig. 2b is more than five times wider than required to release the photonic regions. Even over this large span, the CMOS metal and dielectric film stackup remains planar and stable without special handling conditions.

Although photonic integration is a new application of localized substrate removal, such technology is well proven for CMOS integration within other fields. The microelectromechanical systems (MEMS) community has utilized similar post-processing techniques to create a variety of sensors within standard CMOS processes for over a decade [28]. Recently, Akustica, a subsidiary of Bosch Sensortec GmbH, has demonstrated the commercial viability of such an approach by the sale of over 5 million suspended-layer microphones integrated alongside the necessary interface circuitry in standard CMOS processes [29].

3. Photonic device performance analysis

Utilizing previous generation projection lithography steppers, prior work has demonstrated a wide variety of highly-resonant photonic devices that require post-fabrication trimming to align resonances for adequate device performance [5–10,30–33]. Previously, scanning-electron-beam lithography (SEBL) was used to fabricate structures with sufficient resolution and process control to enable the required resonance frequency matching required for highly-resonant devices [34–36]. However, the substrates used were not optimal for electronics, and SEBL is incompatible with standard CMOS processing for a number of reasons, including throughput, which for SEBL is many orders of magnitude lower than that of the optical-projection lithography used in modern CMOS facilities and foundries. In this work, state-of-the-art ArF 193 nm immersion lithography scanners with 1.35 numerical aperture (NA) performed standard front-end lithography on 300 mm wafers. This technology is a significant improvement over the most advanced prior silicon photonic work [32] where non-immersion 193 nm ASML PAS5500/1100 scanners with a 0.75 NA were used on 200 mm wafers.

Grating couplers [37–39] provided surface-normal optical input and output for 150 integrated microring resonators. Figure 3a shows a grating coupler that provided a minimum insertion loss of 4.8 dB at 1560 nm with a 1 dB bandwidth of 93 nm. Coupling efficiency was measured using lensed SMF-28e fibers to match the 5 μm mode size of the coupler. An Agilent 11896A polarization controller was used to align the input polarization linearly with the long direction of grating bars by maximizing transmission. The resulting transverse-electric-polarized light in the waveguide is used as the single operating polarization for the integrated photonic platform. In applications requiring polarization-independent interfaces, alternative coupler design is required to decompose the arbitrary input polarization into two transverse-electric-polarized waveguides [40]. The coupler was designed with fully-etched gaps of 480 nm between 590 nm bars and simulated to have an insertion loss of 5.5 dB. The small discrepancy between the simulation and the measurement is attributed to incomplete information of the exact dielectric layer thicknesses and refractive indices in the CMOS back-end. The resulting uncertainty in the reflection from these layers is larger than the difference between theory and experiment.

The propagation loss in our waveguides – cross-section shown in inset to Fig. 3b – was determined by using the cut-back method and by measuring the intrinsic quality factors (Qs) of weakly-coupled ring resonators. For the wavelength range near 1550 nm, 670 nm wide
waveguides were chosen to strongly confine transverse-electric-polarized light while remaining single-mode at the thickest and widest dimensional tolerances of the polysilicon core. As shown in Figs. 3b and 3c, we obtained approximately 55 dB/cm from 1520 to 1580 nm, and a Q of approximately 8000. This loss is significantly higher than previously reported for silicon photonic devices due to our reliance on deposited polysilicon that has not been optimized for photonics. The top surface roughness of end-of-line polysilicon is approximately 6-8 nm rms with a correlation length of 100-200 nm as measured by TEM, which is consistent with theory [41]. Still, the measured Q is suitable for devices such as ring-resonator WDM filters and modulators designed for 10 Gb/s datacom. Additionally, for key

Fig. 3. (a) Measured insertion loss of the vertical grating couplers (inset: SEM of coupler). Waveguide propagation loss (b) calculated by the differential loss through two waveguide structures (inset: TEM of waveguide cross-section for 670 x 80 nm polysilicon core, clad with a conformal 50 nm silicon nitride liner and surrounded by oxide) with a straight section length difference of 2.72 mm and identical bends. Error bars calculated as the standard deviation for 4 samples. (c) Transmission through the drop port of a weakly coupled 670 nm width, 20 \( \mu \)m radius ring resonator. SEM of resonator containing fill shapes in center for process compliance shown in inset. The measured quality factor was 7960. Measured data (blue dots) is most closely fit with a simulated ring resonator transmission response with a 55 dB/cm waveguide loss. The sensitivity of this technique is illustrated by the divergence of the 50 dB/cm and 60 dB/cm simulated responses.
applications that require high-density, low-energy off-chip interconnect with minimal on-chip routing distances [4], the demonstrated device set meets system requirements.

The waveguide loss of 55 dB/cm measured in this work is consistent with losses measured by early material optimization attempts to reduce polysilicon loss below the 350 dB/cm initially measured for as-deposited polysilicon [42,43]. There is a path to reducing this loss further – enabling higher Q resonators and longer distance waveguide routing – through optimizing the in-foundry polysilicon deposition conditions. Similar work performed for micrometer-sized waveguides [44] and then for single-mode nanowire waveguides [45], successfully demonstrated polysilicon waveguide losses below 10 dB/cm. Such an approach, however, would limit the ability of photonic devices to leverage existing infrastructure. Alternatively, the localized substrate removal post-processing presented here can be leveraged to fabricate photonic devices within a standard thin-SOI-CMOS foundry where the buried oxide layer thickness is below 200 nm in modern processes. As mentioned in Section 2, the presence of the patternable single-crystalline silicon layer traditionally used for the transistor body enables the possibility of low-loss integrated waveguides in that platform. The typical layer thicknesses for the single-crystalline silicon layer range from 50 to 100 nm in deeply-scaled processes. Since a common thin-SOI layer thickness of 80 nm for state-of-the-art processes matches the polysilicon layer used in this work, a direct transfer of device geometries is possible. Although the thin-SOI technology represents a smaller fraction of the total CMOS market, such a platform would enable high-performance electronics integrated alongside low-loss waveguides within the existing foundry infrastructure.

Figure 4 shows the four-channel second-order ring-resonator filter bank we fabricated and tested as one of the primary device demonstrations on this chip. Similar filters have been demonstrated in a variety of materials to enable complex wavelength routing [34,46–48]. The transmission functions for all ports in this filter bank, shown in Fig. 5a, were measured without any thermal tuning or post-fabrication trimming. The drop-port insertion losses were below 5 dB, and the crosstalk between adjacent channels was less than 15 dB. The mean channel spacing for the four filter banks measured was 137 GHz. As far as we are aware, this is the first time that high-index-contrast second-order filter banks have been repeatably fabricated to yield untuned insertion losses below 5 dB with channel spacings below 200 GHz.

The precision of the unmodified CMOS process can be quantified by analyzing several copies of this filter bank across the 300 mm wafer. Over the short length scale of a single filter, the polysilicon thickness is approximately constant and the average lithographic linewidth control can be determined from the frequency matching of the two rings. The mean frequency mismatch – extracted by simultaneously fitting the through-port and drop-port responses to a model – was 30.9 GHz, as shown in Fig. 5b–d. With a simulated resonant-frequency dependence on microring waveguide width averaged along the circumference of the resonator of 38 GHz/μm, this corresponds to an average linewidth mismatch of 810 pm. The standard deviation of this mismatch, representing the stochastic variation of this process was 680 pm, less than six times the stochastic-process variation limit of 120 pm previously
Fig. 5. (a) Measured transmission normalized to off-resonance through port transmission for a four-channel second-order filterbank. No thermal tuning or post-fabrication trimming was performed for these measurements. Port line colors and naming convention correspond to labels in Fig. 4. To extract the resonant frequency mismatch for the two rings in the second-order filters, measured through and drop transmission functions were fit to ideal filter model with the following free parameters: bus-ring coupling coefficients, ring-ring coupling coefficients, ring round-trip loss, and separate resonant frequencies for each ring. (b,c) Resulting model fit (dotted black) lines for through and drop responses overlayed with measured transmission (solid red) lines for two example filters. Extracted bus-ring coupling coefficients of 10.2% ± 1% and ring-ring coupling coefficients of 0.63% ± 0.08% differ from design values due to thinner polysilicon and thicker nitride layers in fabricated filters as compared to simulated couplers. (d) Histogram of resonant frequency mismatch between the two rings in the second-order filters from four die from different wafer locations.

achieved using SEBL in which the intra-field distortion was corrected [35,36]. The demonstrated precision establishes the feasibility of high-yield fabrication of such resonant nanophotonic devices in state-of-the-art CMOS. This demonstration is significant because yield due to fabrication variations is a primary reason that frequency-matched resonators have not been widely adopted in applications such as demultiplexers and modulators where they otherwise promise major technical advantages including higher density and energy-efficiency.

In contrast to the precise frequency matching and channel spacing within a filter bank, the absolute frequency of each filter channel varies by as much as 600 GHz across the 300 mm wafer, presumably due to variation in polysilicon thickness. This variation is expected to be even higher from wafer to wafer. At first glance, this appears to be a major yield barrier. However, systems utilizing dense channel packing of the full filter free-spectral-range reduce constraints on absolute frequency control by allowing fabricated filters to be locked to a nearest-neighbor wavelength grid. This locking can be achieved through thermal tuning using the filters’ effective thermo-optic coefficient of 7.9 GHz/°C, as shown in Fig. 6a. The undercut photonic region offers a 24-fold increase in thermal impedance, as shown in Fig. 6b, and therefore over an order-of-magnitude reduction in tuning power. If heaters are directly
integrated into the ring filters, as has been demonstrated previously [49], low tuning efficiencies of 3 μW/GHz would be achievable. Recently, this approach has been demonstrated to produce record low tuning powers where localized substrate removal was not required for optical mode isolation [50].

Although the high thermal impedance is useful for reducing passive device tuning power, it can also result in excessive operating temperatures for power dissipating active devices such as modulators. This temperature increase for an integrated active device can be calculated by multiplying the thermal impedance by the power dissipation. For example, the measured thermal impedance of the heaters that are well insulated from the neighboring circuitry, 44 °C/mW, would cause an energy-efficient modulator with a power dissipation of 0.5 mW to reach an operating temperature of 22 °C above the surrounding environment. To reduce the temperature rise for a given power dissipation, local environment engineering can reduce the thermal impedance. If modulators are instead placed at the edge of the localized substrate removal region, contacted with wide copper lines and surrounded by thick, substrate-connected metallization, the effective thermal impedance would be reduced to enable even lower operating temperatures.

4. Conclusion

In this paper, we demonstrated a wavelength demultiplexing filter bank integrated in the front-end, electronic device layer of a state-of-the-art 28 nm bulk-CMOS process. This device demonstration served as a vehicle to demonstrate our proposed foundry CMOS electronics-
photonics integration scheme and to evaluate its feasibility by quantifying critical photonic device performance parameters. The dimensional precision demonstrated indirectly through optical measurements of the filter banks, combined with the potential of ultra-low-power wavelength locking, provides the basis for a scalable nanophotonics-electronics integration platform. Since the waveguide polysilicon layer is also the transistor gate and local interconnect layer of the standard bulk-CMOS process, available doping and metallization steps allow active devices such as carrier-injection modulators to be built upon this foundation [6]. On the detection side of the link, deeply-scaled CMOS processes already include a silicon-germanium layer for stress engineering the p-type transistor [51]. This lower bandgap layer may then be leveraged to integrate front-end photodiodes and form a complete photonic device platform at short operating wavelengths such as 1.2 μm where the SiGe alloy ratio provides a sufficient absorption coefficient. Since this platform is built into a state-of-the-art CMOS process, a major step in electronic-photonic circuit integration is enabled. The existing electronic CMOS infrastructure already demonstrated to fabricate 2 billion transistor circuits with high yield [52] could now be used to simultaneously fabricate nanophotonic circuits with high yield as well. By complying with all in-foundry processes, no further infrastructure investment is required. Additionally, sharing the mask costs and all wafer-level processing on multi-project wafer runs with the large number of electronics industry projects significantly lowers the incremental cost of developing systems and devices [23]. This work may also be carried over to thin-SOI-CMOS foundries where there is the potential of low-loss waveguides to enable further system applications.

Acknowledgements

The authors acknowledge Dr. Jagdeep Shah of DARPA for funding under contract numbers W911NF-06-1-0449 and W911NF-08-1-0362. The authors also acknowledge J. L. Hoyt and M. Schmidt of MIT for discussing the localized substrate removal method as well as D. Buss, S. Yu and R. Khambarkar of Texas Instruments for discussing relevant CMOS processing steps. M. Georges, J. Leu and B. Moss of MIT designed the electronic circuits contained on the die shown in Fig. 1a.