Back-End Deposited Silicon Photonics for Monolithic Integration on CMOS

Yoon Ho Daniel Lee and Michal Lipson, Senior Member, IEEE

(Invited Paper)

Abstract—We present the vision of back-end deposited silicon photonics (BDSP) and review works that have been done in this field. Individual aspects of BDSP platform including excimer-laser-annealed polycrystalline silicon, low-loss plasma-enhanced chemical vapor deposition silicon nitride waveguide, modulator, detector, electrical interface, back-end CMOS compatibility, and benefits of the platform are discussed in detail.

Index Terms—CMOS technology, excimer laser annealing (ELA), integrated optics, optical modulators, photodetectors, polycrystalline silicon, silicon nitride (SiN), silicon photonics (SiP).

I. INTRODUCTION

Advances in silicon photonics (SiP) have produced high-performance building blocks such as modulator, detector, switch, and Mux/Demux that are highly desirable for optoelectronic integration with CMOS system for optical interconnections [1]. Silicon modulators operating at 40 Gb/s have been demonstrated by multiple groups [2]–[4], and comparable detectors exist as well [5], [6]. While the push for high-performance device designs continues, performances suitable for on-chip optical interconnects have already been reached, satisfying architectures that optimize energy and bandwidth. In an on-chip interconnect setting in which total available power is limited due to limited heat extraction from the chip and the associated cooling cost, energy comes at a premium. In such a scenario, where energy per bandwidth is an important figure of merit, multiple slower channels operating at small multiples of the system clock rate (10 Gb/s × 4 λ) are favored over one fast channel (40 Gb/s × 1 λ) due to reduced electronic power and circuit complexity overhead paid in serialization and deserialization (SerDes) and optoelectronic transceiver [7].

While these devices meet the performance needs, they remain incompatible for integration with the standard CMOS processes used in fabricating the latest generation of microprocessors and memories. Guiding of light requires sufficient optical isolation from the surrounding, i.e., a separation between the waveguide and the silicon substrate whose thickness depends strongly on the refractive index contrast, geometry, and wavelength. This separation is typically on the order of 1 μm in a Si–SiO₂ material system in the telecommunication wavelength. This requirement conflicts directly with the lack of native optical isolation from the substrate in a typical CMOS process, in which bulk process offers none and modern 45-nm silicon-on-insulator (SOI) process offers less than 200 nm of buried oxide (BOX) [8]. Furthermore, BOX thickness of SOI CMOS devices is projected to shrink further down to 10–30 nm. Guiding of light also requires minimum dimensions in order to ensure optical confinement, typically on the order of 150–400 nm in silicon. This requirement conflicts directly with predictions that the thickness of SOI will shrink to 5–10 nm [9] for fully depleted SOI transistors due to device electrostatics. Therefore, there is a strong need to address these limitations of silicon photonics incompatibility with CMOS, especially so for the more advanced process nodes to which photonic interconnect is geared toward. Attempts to address some of these issues have been made recently by removing part of the substrate to address leakage of light from the waveguides [8], [10].

Back-end integration of photonics has been suggested recently in [11]–[13]. In [11], the authors use electrooptic polymer and germanium, and authors in [12] use III–V as the active material. In this paper, we show that back-end integration is possible without departing from standard SiP material system, which enables the use of CMOS foundries for both optics and microelectronics independent of the underlying microelectronic fabrication process.

II. BACK-END DEPOSITED SILICON PHOTONICS (BDSP)

CMOS back-end deposited photonics have two main aspects to it: low-temperature excimer-laser-annealed (ELA) polysilicon for active devices, and low-loss silicon nitride (SiN) for passive waveguides. We combine these two technologies into the BDSP platform as in Fig. 1, which clearly delineates the deposited photonics on top of the CMOS back end from the underlying CMOS. CMOS microelectronics consists of the front end of line (FEOL), which includes the transistors and other active devices fabricated on the silicon substrate at the bottom in green, and the back end of line (BEOL), which is the system of multiple layers of metal (as many as ten or more in modern logic process) and interlayer dielectric that connects the front-end devices together to form a circuit. BEOL traditionally ends with the last metal layer that interfaces with the outside and the passivation layer on top to protect the BEOL, but BDSP augments this BEOL with multiple photonic layers.
In the upper deposited photonics layer in Fig. 1, we show two layers of SiN waveguides in blue, and one layer of ELA polysilicon in green for clarity. As in any photonics platform, waveguide needs optical isolation, and this isolation is provided by a layer of SiO$_2$ deposited using plasma-enhanced chemical vapor deposition (PECVD), depicted in a light shade of gray. The SiN waveguides in multiple layers traverse in orthogonal direction in order to minimize unwanted interlayer crosstalk and crossing losses, and a ring resonator can be used to couple from one layer to another very efficiently as an optical via as we demonstrated in [14], with crossing loss as low as $-0.04 \text{ dB/cross}$ and interlayer coupling insertion loss as low as $-0.6 \text{ dB}$. In order to modulate and detect optical data, we propose separate active layers that are placed in between any of the multiple SiN waveguide layers to efficiently couple to and from the bus waveguides.

A. Conditions for CMOS BEOL Compatibility

Integration in CMOS BEOL requires compatibility with not only the CMOS materials and processes, but also requires a strict thermal budget limit to prevent performance degradation. CMOS compatibility is considered a gold standard in SiP because its basis lies in leveraging of the CMOS fabrication infrastructure and process. However, notion of compatibility is often vague to the point where any material not explicitly named in incompatible with CMOS such as gold is phrased to be CMOS compatible. For a straightforward adoption by the CMOS industry, we adopt definition of compatibility as consisting exclusively of materials already in use in commercial CMOS process, which SiN, polysilicon, SiO$_2$, and Ge satisfy. Furthermore, the process flow of BDSP consists of PECVD, photolithography, inductively coupled plasma etch, and chemical mechanical polishing (CMP).

In addition to material and process criteria, thermal budget, i.e., the duration and temperature of thermal processing, is a very important factor in a CMOS process. Modern process flow is very complex with intricate device doping profiles, gate oxides approaching atomic scale, exotic silicides, and metallization diffusion barriers to name a few. The whole process is only strong as the weakest point, and the weak links can be identified as widely used nickel silicide, which can go through metallurgical phase change around 750 °C causing contact resistance to increase [15], and copper diffusion that deteriorates transistor characteristics at a temperature as low as 600 °C [16] in a modern Cu BEOL process. There are reports of aluminum metallization degrading at a budget as low as 1 h at 450 °C [17], [18], but both works attribute these degradation to the low melting point of aluminum BEOL, and not the FEOL. While there are no definitive works on the thermal budget limits of a modern Cu BEOL process, it is expected to be higher than 450 °C of Al BEOL because of 64% higher melting temperature of Cu relative to Al and the aforementioned temperatures. Here, we remain conservative and propose a platform that maintains the thermal budget below 90 min at 450 °C.

B. Properties of Polysilicon

To date, most SiP device works rely on single-crystalline silicon (c-Si), which prohibits its use in the back end. C-Si is of very low optical losses in the telecom wavelength range and another, perhaps even more important quality of silicon is its electrical properties such as carrier mobility and carrier life time. C-Si has both mobility and carrier life time that are much higher than that of a-Si due to absence of defect-assisted scattering and recombination. These excellent electrical characteristics enable low resistivity and dynamic control of free carriers in photonic devices which makes silicon great at high-speed modulation.

Polysilicon exists in the regime between c-Si and a-Si, embodying electrical properties and optical properties in between that of the two phases. Polycrystalline silicon, as its name suggests, exists as an aggregate of small grains, which are packets of c-Si. Polysilicon inherits properties of c-Si, modified by the existence of grain boundaries, which are atomically thin layers of a-Si at the grain interfaces. Therefore, polysilicon appear more and more like c-Si the less a device crosses a grain boundary.

Despite its potential as an alternative to c-Si, polysilicon has rarely been used in SiP due to its inherent high losses [19]. These losses originate from scattering due to surface roughness, grain boundaries, and dangling bonds. Surface roughness can be somewhat minimized using CMP; while grain boundaries can be minimized by maximizing grain sizes, and dangling bonds can be somewhat minimized by intentionally terminating them with hydrogen, as often done for a-Si to reduce losses. With such advances in polysilicon waveguide fabrication using hydrogenation and 16 h of 1100 °C anneal, waveguide loss as low as 9 dB/cm has been demonstrated [20]. However, out-diffusion of hydrogen at temperature above 300 °C causes increase in loss [21], and hydrogen dopant complexes decrease dopant activation efficiency leading to lower electrical conductivity of the film [22]. Therefore, hydrogenation, while useful for making low loss waveguide, may not be optimal for making stable and high-performance active devices.

Recent advances in nanophotonics enable the use of polysilicon in high-performance photonics devices since the sizes of devices have become small enough such that photonic devices span across only a handful of grain boundaries. In the limit where grain sizes are much larger than the device of interest, the device behaves essentially as if fabricated on c-Si. A ring
resonator as small as 1.5 \( \mu \text{m} \) in radius has been demonstrated [23], and with less than 10 \( \mu \text{m} \) of circumference in such device, the device would traverse only a few grain boundaries when fabricated in polysilicon with grain sizes on the order of 1 \( \mu \text{m} \). This ambitious promise would not be feasible with furnace-annealed polysilicon, in which grain sizes are limited by crystallization kinetics—the grain sizes were less than 300 nm even in the aforementioned low-loss polysilicon waveguide. What makes micrometer-sized grain possible is the use of ELA.

C. ELA polysilicon

Large grain polysilicon can be formed with low thermal budget by ELA of a-Si, a technology widely used in fabricating thin-film transistor on glass to manufacture touch screens and LCD screens. This industry-proven technology has throughput of 100 cm\(^2\)/s, exceeding even that of state-of-the-art CMOS lithography tools, corresponding to over five hundred 300-mm wafers/h [24]. The theory behind ELA lies in the use of a pulsed UV excimer laser, which is absorbed completely within the first tens of nanometers from the irradiated surface due to very high absorption coefficient of silicon in UV wavelength around 300 nm or below. Due to the short pulse duration on the order of 10 ns and localization of this energy in the thin film of a-Si, the absorbed energy is enough to transiently heat up the silicon above its melting temperature. The key benefit of ELA is that while this transient heating is enough to cause full melt of a thin a-Si film, total thermal energy transferred to the BDSP stack is negligible [25]. Then, subsequent cooling that follows after the pulse causes small fraction of silicon to solidify first, which acts as nucleation sites aiding in growth of polycrystalline grains. Optimization of this technique produces controlled grain growth as large as 7 \( \mu \text{m} \) in length [26], providing grain sizes large enough to realize the promise of quasi-single-crystalline photonic devices on ELA polysilicon.

Dopant activation and silicidation are another high thermal budget processes in active device fabrication that can be achieved using ELA. Activation step is a critical step that ensures that the dopant atoms find their substitutional sites within the silicon lattice so that they can contribute electrically in forms of donors or acceptors. Traditional process uses furnace annealing or rapid thermal anneal to provide the necessary activation energy, but they are very high thermal budget processes, incompatible with BEOL. Instead, ELA can be used to activate dopant as demonstrated in [27], as well as forming of silicide [28] with low thermal budget.

We have already demonstrated a CMOS BEOL compatible, low-temperature ELA polysilicon ring resonator using a-Si deposited by electron gun evaporation followed by ELA and CMP. We measured quality factor of 2900, translating to 65 dB/cm loss [29]. This relatively high loss is likely due to metallic impurity contamination in starting material from multimeaterial evaporator. We previously demonstrated ring resonator from thermally annealed polysilicon with 300 nm average grain size using LPCVD a-Si starting material to have intrinsic Q-factor of 40 000 [13]. Using electronic grade, contamination-free a-Si deposited using PECVD, we envision that comparable performance from ELA polysilicon.

Fig. 2. Optical eye diagram of deposited polysilicon modulator for nonreturn zero 2.5 Gb/s 27-1 pseudorandom binary sequence signal.

D. Modulator

The polysilicon modulator is based on a similar geometry as the single-crystalline silicon modulator, but one must adjust the doping level to account for the effects of grain boundary traps that reduces doping efficiencies. With careful design of doping, we have already demonstrated modulator in high-temperature polysilicon for initial proof of concept [30]. We demonstrated 2.5 Gb/s data modulation with 10-dB modulation depth by charge injection as shown in Fig. 2. We have extrapolated the carrier mobility from the \( I-V \) characteristics to be at least 100 cm\(^2\)/V\( \cdot \)s and successfully simulated the transient electrooptical behavior of the modulator using this parameter. This simulation will be used in designing the next generation of polysilicon modulator.

E. Detector

There are two candidates for photon absorption in deposited silicon photonics: germanium, and silicon itself. While germanium (Ge) is the most attractive candidate for making detectors in the telecommunication wavelength, silicon can be engineered to produce detectors at the wavelength of interest. Silicon bandgap prevents efficient absorption in the telecom band, but detectors can be made by making use of, or intentionally creating, defects that gives rise to mid-gap state that allows subgap absorption. We have demonstrated responsivity of 0.15 A/W at 1550 nm in a compact polysilicon PIN ring resonator [31], and Geis et al. demonstrated millimeter-scale waveguide photodetector with high responsivity of 0.5–0.8 A/W using Si implantation [32]. While these detectors do not match the performance of dedicated Ge detectors at the moment, pure silicon detector can be beneficial when trying to reduce process complexity.

Germanium is also an excellent candidate for making detectors in the back end. While BDSP platform champions ELA polysilicon as its active material, Ge can be just as easily processed through ELA. While Ge CVD produces high-quality material capable of achieving high responsivity, it comes with a prohibitively high thermal budget for BDSP, requiring temperature beyond 700 °C [33]. Instead, one can deposit Ge at low temperature using evaporation or sputtering which can then in principle be ELA [34].
F. Low-Loss Silicon Nitride Waveguide

SiN as a material for passive optical waveguides is attractive for long, centimeter-scale low-loss optical interconnects. Traditionally, low-loss SiN waveguide has only been used for visible wavelengths due to stress issues complicating the deposition of nitride films thick enough for guiding in the telecom wavelength range [35]. We have recently demonstrated high-confinement SiN waveguide for the telecom wavelengths, with losses as low as 0.065 dB/cm [36] based on annealing and thermal cycling of LPCVD SiN. In [14], we have also demonstrated low-temperature PECVD SiN multilayer 3-D integrated SiN waveguide system as illustrated in Fig. 3, with losses slightly over 1 dB/cm in the L-band increasing to 6 dB/cm at the lower bound of the C-band, and excellent adjacent layer crossing losses of $-0.04$ dB/cross and interlayer coupling insertion loss as low as $-0.6$ dB. The increase in loss in C-band is due to Si-H and N-H bond absorption harmonics and it can be lowered with process optimization. The multilayer stack formed by PECVD of SiO$_2$, SiN, and CMP planarization process also allows for precise control of vertical coupling gap, which can help mitigate coupling coefficient variability present in typical lateral coupling scheme.

G. Electrical Interface

Electrical connection with low parasitic capacitance and resistance is needed to maximize the performance of BDSP active devices. As seen in Fig. 1, the structure that connects the last metal layer of CMOS BEOL to active devices resembles a through-silicon via (TSV) in that it penetrates through the entire stack of photonic layers. For BDSP, one needs a via of length as short as 3 μm to as long as 10 μm or longer depending on the number of photonic layers, with low resistance and robustness. Fortunately, these requirements are easily met with the existing TSV technologies that have been under active research for purpose of 3-D stacking in microelectronic industry [37]. However, these long vias introduce additional fringing capacitance that can limit systems RC time constant and may also be susceptible to capacitive coupling between unwanted neighboring signals, which must be carefully mitigated. Literature suggests that the capacitance and coupling can be minimized with judicious use of shielding and spacing [38].

III. EXPERIMENTAL VERIFICATION OF CMOS BEOL COMPATIBILITY

CMOS BEOL integration hinges on maintaining CMOS integrity after the low-temperature BDSP processing steps, but it is not clear what the level of integrity of sub-100 nm CMOS process is after going through various thermal processing conditions and plasma processes. We show here that CMOS is indeed unaffected by the various relevant processing conditions used in BDSP.

We designed a custom CMOS chip in a commercial, 90-nm bulk CMOS process offered by IBM foundry services. After applying 90 min of thermal processing in nitrogen ambient in annealing furnace at three different conditions: 400 °C, 500 °C, and plasma processing in addition to 400 °C furnace anneal to model a typical BDSP process flow, we measured the drain current versus gate voltage of transistors 5 μm wide and 1 μm long from each condition and compared them against an unprocessed baseline device for changes. In Fig. 4, we see that the drain currents as a function of gate voltage for both P-type and N-type transistors match with one another extremely well even after going through different thermal processing conditions.

IV. OUTLOOK FOR BDSP

BDSP offers multiple benefits: independence from complex CMOS front-end processing, reduced constraint in photonic
footprint, and multilevel architecture. In a modern CMOS process, it is not uncommon to find a process flow with more than 40 mask layers. In such a complex set of processes, every small tweak to a given processing step can lead to unintended compounding of side effects that can adversely affect yield or even render a process unstable. It does not help that the industry’s profit margin is thin; so, it is almost natural for the CMOS foundries to be very risk adverse and unresponsive to bringing new processes or modules into their facility, including photonics.

The FEOL of a CMOS is the most sensitive part of the process, and thus foundries are rightfully opposed to making changes at the front end to accommodate photonics. BDSP decouples photonics from the most sensitive part of a CMOS process, and adds the whole photonics module after the very end of a CMOS process, so that foundries are not required to change their process. In fact, back-end photonics processing can in principle be done in a different foundry from which the CMOS wafer was fabricated, since the photonics process is its own complete module that does not intrude upon, or depend on other processing steps of the underlying CMOS. This aspect greatly lowers the barrier of introducing silicon photonics into manufacturing.

The cost of adding the photonics module is kept low by use of i-line or 248-nm lithography used in noncritical back-end layers. The SiN waveguide has a width of 1 μm, and polysilicon active waveguide is 700 nm wide, well within capability of i-line lithography. Furthermore, the lateral alignment requirement across layers is expected to be around 100 nm depending on specific extinction ratio requirements, which is easily met even by i-line tool at 12-nm overlay [39]. Photonic module will add approximately seven mask layers per active layer and one layer per passive SiN waveguide, where much of active layer masks can be reused for patterning additional devices in different layers in some scenarios to reduce cost. Note that the masks become exponentially more expensive as the technology node becomes smaller, and by using back-end process lithography, which is a generation or two behind the node of the process, total cost of the photonic module can be kept to a small fraction of the total mask cost process cost [40].

Deposited silicon photonics also greatly alleviates the constraints on footprint of photonic devices. The front-end silicon real estate is considered a highly valuable commodity, since every savings in area translates to more dies, hence revenue, per wafer. This is the reason why the microelectronics industry has pursued larger wafers and smaller transistors. If integrating photonics in the front end means that total die area is going to increase significantly, one takes a hit not only because there are less dies per wafer, but also because yield of a die decreases nonlinearly as a function of the die area [41]. Therefore, if photonics is to be introduced in the front end, its footprint is critical. While a ring resonator is one of the most compact photonic structures short of photonic crystal cavities, a typical ring resonator is still several micrometers in radius, which translates to hundreds of micrometers squared of footprint once optical isolation is considered. In addition, typical photonic transceiver circuitries are several hundreds of micrometer squared per channel, which adds significantly to the total area. Therefore, moving the photonic devices out of the front end significantly decreases the total real estate needed for photonic interconnects, enhancing its area competitiveness. This competitive edge becomes even more apparent when we consider other common designs like Mach–Zehnder interferometer-based modulators which can easily approach a millimeter in length in order to achieve sufficient extinction ratio at CMOS voltages. Therefore, by separating the photonics to dedicated layers, we greatly alleviate the issue of photonic footprint.

Similar to the multiple metal layers in CMOS back end, deposited silicon photonics naturally lends itself to multilayer optical routing, but it goes even further by enabling multiple layers of active devices. A network-on-a-chip that supports communication between cores in a massively multilcore chip multiprocessor, for example, requires a closely knit network that can only be realized with many waveguide crossings. In-plane waveguide crossing is inherently lossy, and even relatively low loss of 0.7 dB/cross demonstrated in [42] accumulates quickly and renders a network topology infeasible. However, in BDSP with multiple layers of low-loss waveguides with very low crossing losses as discussed earlier, such network is perfectly feasible. Another benefit of having photonics on the back end is its easy access to end fire coupling from the periphery of the die. In a logic die where top side of the chip is completely covered in arrays of bumps for I/O, accommodating fibers vertically among arrays of bumps may be very difficult. However, side of the die remains clear, and by use of our plasma-etched facet in [43] to define the smooth facet required for end fire coupling, we can achieve very efficient side coupling while being compatible with both flipchip packaging and mass manufacturing in just a single dielectric etch process followed by dicing. In addition, on-wafer testability can be maintained by use of grating couplers in SiN layers for optical testing before bump metallization [44].

V. CONCLUSION

BDSP has a multitude of benefits including reduced constraint in photonic footprint, multilevel optical routing, potential for unique device and system architecture that makes use of its 3-D nature, and most importantly its independence from CMOS. The combination of mass-production-compatible multilevel silicon active layers, modularity, and strict CMOS compatibility makes BDSP an appealing solution for both photonics designers and CMOS foundries. This opens up a different dimension to silicon photonic integration, potentially transforming what photonic integration on CMOS means and help more rapid adoption by the CMOS foundries thanks to its fundamentally nonintrusive nature to the CMOS process.

REFERENCES


Yoon Ho Daniel Lee received the B.S. (magna cum laude) degree in electrical and computer engineering from Cornell University, Ithaca, NY, in 2010, where he is currently working toward the M.S. / Ph.D degree in electrical and computer engineering.

His research interests include integrated on-chip optical interconnects, polycrystalline silicon photonic devices, and CMOS optoelectronic circuit design.

Michal Lipson (SM’07) received the B.S., M.S., and Ph.D. degrees in physics from the Technion—Israel Institute of Technology, Haifa, Israel, in 1992, 1994, and 1998, respectively.

In December 1998, she joined the Department of Material Science and Engineering, Massachusetts Institute of Technology, as a Postdoctoral Associate. In 2001, she joined the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, where she is currently an Associate Professor. Her research at Cornell University involves novel on-chip nanophotonic devices. She is the inventor of 15 patents regarding novel micrometer-size photonic structures for light manipulation. She is the author or coauthor of more than 150 papers in the major research journals in physics and optics.

Dr. Lipson is a McArthur Fellow, and the Fellow of the Optical Society of America. She was the recipient of the National Science Foundation CAREER Award, IBM Faculty Award, and Blavatnik Award, New York State Academy of Science.