The industrialization of the Silicon Photonics: technology road map and applications

Maurizio Zuffada
STMicroelectronics
Milan, Italy
maurizio.zuffada@st.com

Abstract— The R&D maturity level reached by the Silicon Photonics technology envisions a clear road map fitting the needs and the challenges of the future ICT (Information Communication Technology) systems and services. Four key applications will drive the evolution: intensive computing, broadband communication, mass storage and consumer multimedia. The Silicon Photonics technology, ported in the 300 mm silicon wafer fabs, together with the most advanced package techniques, will offer a low cost manufacturing infrastructure for 3D heterogeneous system integration. Starting from a summary of the peculiarities of this technology, the unprecedented level of miniaturization and power consumption reduction, this paper will address several disruptive applications: starting from active optical cables, going to optical modules, to inter-chips communications, up to intra-chip applications, it will be shown how the Silicon Photonics technology will progressively evolve from 40 Gbps to 3200 Gbps rate, from 20 pJ/bit to 1 pJ/bit dissipation, from 4 $/Gbps to 0.04 $/Gbps cost and from 250 mm ³/Gbps to 0.05 mm ³/Gbps volume. Furthermore, a picture on how industry is preparing for the massive adoption of Silicon Photonics, by exploiting the existing large scale semiconductor manufacturing environment, will be provided.

I. INTRODUCTION

Data communications through metallic interconnections (cables, back-planes, boards and Integrated Circuits) show relevant attenuation and electrical crosstalk even at few GHz of bandwidth and relatively short distances [1]. Intensive computing applications in data centers are being more and more limited [2]. In particular, power consumed for I/O data transfer and for system cooling are increasing at too fast pace. While metallic substrates supporting electrical data transfer are unable to offer a reliable long term solution, photonics, up to date used for long distances, is a good candidate also for applications entailed at short distances, by taking advantage of the huge bandwidth and low energy consumption. The push toward ever increasing data rates coupled with low power consumption suggests integration of optics together with electronics on silicon giving rise to Silicon Photonics, the technology discussed in this paper. The chart, shown in fig.1, synthesizes the foreseen electrical limits versus data communication distances. Traditional hybrid photonics, based on the 2D/3D assembly of III-V photonic devices with the traditional silicon electronics, may address problems of electrical interconnects; nevertheless it will not fill the gap for photonics electronics integration. The main limit of the hybrid solution comes from the miniaturization required by the long term intra-chip communication. The emerging Silicon Photonics technology promises a low cost solution by allowing close proximity integration of photonics with electronics.

![Figure 1 – Electrical Data Rate limits for different data communication distances](image)

This article is organized in four Sections. Section II will describe an electro optical transceiver based on the Silicon Photonics technology and its characteristics. The four key figures of merit, namely: bandwidth, energy, size and cost are introduced through an example. Section III describes the necessary breakthroughs for the evolution of Silicon Photonics in order to address application specifications. Main technological challenges to be addressed and solved are also discussed. Section IV introduces a possible Road Map vision of the Silicon Photonics Technology and its applications. Finally, Section V provides a summary and draws the main conclusions.
II. SILICON PHOTONICS ELECTRO-OPTICAL TRANSCEIVER

The Silicon Photonics CMOS process starts from silicon-on-insulator (SOI) wafer. The buried oxide and the epitaxial layer thicknesses are chosen and optimized according to the range of the selected operating wavelengths. SOI is required for the confinement of the light in the optical waveguides. First process steps are instrumental to realize the photonic structures inside the silicon epitaxial layer. There are six basic photonic structures necessary to build up an optical transceiver: two types of Grating Couplers (GC), the single mode waveguides, the beam splitter, the modulator and the Photo-Detector (PD). The GC structures allow the coupling of the light, either from the laser source or the I/O single mode fiber arrays with the silicon waveguides.

Figure 2 – CMOS Silicon Photonics process cross-section

As shown in fig.2, the light is confined and guided into the silicon strip layer, thanks to the difference of the two refractive index of the silicon epi-core (nSi ≅ 3.5) with respect to the cladding silicon oxide (nSiO₂ ≅ 1.5). Selective P-type and N-type implantations form the P-N and the P-I-N junctions as well as the diffused resistors, essential to implement optical phase modulators. After the definition of the CMOS transistors, a pure germanium layer is selectively grown over small areas of the waveguides. These small Ge islands enable the monolithic integration of high-performance photodetectors within the silicon waveguides. The input/output light is coupled to the Silicon Photonics die through the Grating Couplers [3] which are passive optical I/O structures, shown in fig.3. The GC’s are designed to operate with tilted incident light angle; this improves the efficiency of the coupling. The GC’s are diffractive (periodic) structures etched in the silicon material. Vertical optical interfaces allow very low loss optical coupling from/to the surface of the chip. Low loss is enabled by the special design of the structure so that the light coupled to the diffractive structure is “mode matched” with a single mode optical fiber. Vertical optical interfaces have significant advantages being fully compatible with the CMOS process and the single mode fibers assembly. Another big advantage offered by GC’s is in the testing of the optical circuits at wafer level with high speed optical wafer probing. Because of the strong dependence from the signal polarization state, Single Polarization Grating Couplers are used to couple the light from the (polarized) light source into the die and the light from the die into the single mode fiber. Instead, the Polarization Splitting Grating Couplers are used to couple the light from standard single mode fiber into the silicon wave guides. These couplers are more complex and consist of a tapered bi-dimensional diffractive structure. The Single Polarization Grating Couplers are used in transmission direction when a laser source, which has a defined polarization state, is directly coupled to a waveguide (without the transit of the light through an optical fiber which normally rotates the polarization randomly).

Figure 3 –SPGC - Single polarization grating coupler

The Polarization Splitting Grating Couplers are used in the receiver direction because the polarization state of the received light signal is not defined. Propagating mode in silicon waveguides is characterized by the transverse electric (TE) field. The typical WG losses are below 0.5 dB/cm. The beam splitters are usually cascaded to distribute optical power over many waveguides. The commonly used electro-optical modulator is based on the Mach-Zehnder interferometer, whose structure is shown schematically in fig. 4. The narrowband input field E₀ enters in the MZ through a single-mode waveguide, then it is split into two equal parts by a directional coupler. The phase velocity of the light in the silicon waveguides is equal to c/nₑ, where nₑ is the effective group refractive index that depends on temperature and dopants concentration in the silicon core. By controlling the reverse biasing of the depletion region in the P-N junctions of the two arms, the c/nₑ phase velocity of each arm can be varied and, over the MZ length, the optical phase shifted. After two symmetrical paths, the CW beams are recombined using the same directional coupler structure to produce the output field E_out. The phase of the balanced 50/50 CW light in the two arms is controlled so to achieve a relative phase shift either equal to 0 or π. When the two arms are combined together, the light interferes constructively or destructively and hence the light is OOK (On Off Keing) modulated at the exiting waveguides. Usually only one output is used, the other output has the inverted optical data. In order to set the operating point of the MZI modulator, a precise static offset control of the phase difference between the two arms is necessary. This is possible by means of a low frequency feedback loop. The phase mismatch compensation can be realized with a relatively short time constant, in the order of ms, therefore the phase modulators can be based either on the carrier plasma dispersion effect (PIN type) or on the thermo-optical effect. In the carrier plasma dispersion effect the variation of the waveguide refractive index is caused by
carrier injection while in the thermo-optical effect the index variation is caused by local heating.

![Figure 4 – Mach-Zehnder Interferometer](image)

The Ge photodetectors are P-N diodes, biased at a reverse voltage of about \(-1 \div -2\) V. These Ge diodes are very small micro-strips (see fig.5), about \(1 \mu m\) wide and \(10 \mu m\) long, with a thickness of about 300-500 nm, built on top of the silicon waveguide. Thanks to the higher refractive index of the germanium, the light, while traveling along the z axis, is totally absorbed by the Germanium strip. The absorbed photons generate hole-electron pairs inside the depletion p-n region, which produce the photo current. Typical values of Ge photodiode responsivity, for wavelengths from 1300 nm to 1550 nm, range from 0.7 A/W to 1 A/W. Being these devices very small, their parasitic capacitance is below 10 fF, ideal to achieve a large BW at the receiver side. The typical BW (-3dB) of available Ge photodiodes are in excess of 30 GHz [4].

The electro-optical transceiver, shown in fig.6, is the fundamental sub-system through which the electrical data are up-down converted into optical signals. In this section we will limit our analysis to an electro-optical transceiver that is exchanging the data with an IC placed at a distance of few centimeters through serial ports at a given Data Rate. On the TX side, a continuous wave light beam, provided by an external laser source at a given wavelength, is coupled to a Silicon Photonics waveguide, through a GC. Then, the beam is split by a factor of 2 or 4 depending on the number of links/fibers the optical module has to be compatible to. After the splitters the light beam enters in the high speed MZ modulator. The serial data, received by the input buffer, are equalized and resynchronized (if necessary). The NRZ digital inputs modulate the light beam phase delay in the two arms of the MZ. As a consequence, the output of the MZ shows an amplitude modulated beam with two level of light power, P1 and P0. The ratio P1/P0 represents the Extinction Ratio, while the difference (P1-P0) represents the OMA (Optical Modulation Amplitude). Today CW lasers electrical power consumption, operating at wavelength between 1310nm and 1550nm, sustaining an optical output power of 13dBm, in nominal conditions, shows a compliance voltage of about 1.4 V with a drive current of about 80 mA, hence the electronics power spent for the driver is close to 150 mW. The receiver part of the electro optical transceiver, shown in fig.6, includes an external SM fiber which is coupled to a PSGC (Polarization Splitting Grating Coupler). The PSGC couples the input modulated beam light to the silicon waveguides and to the Ge PD. The PSGC ensures that no penalty will be paid for the optical signal energy carried on the two polarization modes.

![Figure 6 – Silicon Photonics Electro Optical transceiver functional block diagram](image)

The electrical current signal, produced by the Ge PD, will be equal to the OMA input power multiplied by the PD responsivity. For a given BER (Bit Error Rate), the OMA sensitivity at the receiver is proportional to the total rms noise at the input of the TIA, and inversely proportional to photodetector responsivity. The OMA sensitivity reached with the Silicon Photonics at 10Gbps has been shown to be better than -22dBm with a BER< 10^-12. These results have been obtained with a TIA-LA implemented in CMOS 130 nm with power consumption below 25 mW and a size of 0.9 mm² [5]. A monolithic Silicon Photonics electro-optical transceiver (4 x 10Gbps) has been commercialized, since 2010, by Luxtera within Active Optical Cable applications. The Silicon Photonics transceiver has a size below 40 mm² and is assembled in a QSFP housing with a total power consumption of 780 mW. The QSFP housing contains the Silicon Photonics IC (see fig.7), the laser source and the fiber input/output array plus the voltage regulators and few discrete electronic components. This electro-optical transceiver represents the reference starting point for the next Section II, setting the following performances: a DR of 10 Gbps/link, an energy per bit below 20 pJ/bit, a form factor of 40 Gbps/10,000 mm³ and a reference cost in the order of 4 $/Gbps.
III. SILICON PHOTONICS BREAKTHROUGHS

The Silicon Photonics technology, analyzed in Section I, has been developed on 200 mm wafers with lithography and metrology process node of 130 nm. By porting such a technology into 300 mm wafers with lithography and metrology process node of 65 nm, an increase of the data rate per link of a 3x factor can be expected. This improvement is due to CMOS electronics which allows to scale down the parasitics and to increase the $f_T$ of the transistors by 3x ($f_T=180$GHz). The 300 mm, 65 nm process environment will also improve the matching of the photonic devices as well as the losses due to the light’s scattering. Matching and reduced losses will also imply less required laser output power.

The die size will not significantly scale down because the chip is dominated by the photonics device occupancy and their constraints, however with some optimization, the total transceiver die size could be squeezed to about 35 mm$^2$ according to the Pareto block diagram shown in fig.8. The 65 nm CMOS transistors will also allow some improvement in the electronic MZI driver’s efficiency. Based on these assumptions we could reasonably predict that a monolithic Silicon Photonics electro-optical transceiver, similar to the reference described in section I, ported in a 300mm wafer CMOS 65 nm SOI process, will achieve a total aggregate BW approaching 4 x 30G, with an energy per bit below 12 pJ/bit (compatible with the QSFP size and power dissipation capability). This technology will be matured for massive production in three years time frame offering a transceiver cost below 1.33 $/Gbps. These characteristics will allow to increase the Silicon Photonics market share of the active optical cables and enter in new application fields such as the optical back-planes for board to board communications. Meanwhile some high end inter-chip interconnections based on Silicon Photonics electro optical interoperability (Silicon Photonics interposer) could also occur.

The main challenge of these devices is in the thermal stabilization, because their optical parameters, like the resonant wavelength, are very sensitive to the temperature variations. Their maximum driving voltage swing (1V) is compatible with modern CMOS technology and the ER is in excess of 4 (6dB). By using a CMOS 28 nm electronics, and assuming an $f_T=300$GHz, a modulation speed of 50 Gbps/link per wavelength could be sustainable. This new class of modulators will dramatically reduce the modulation energy per bit because both the capacitance and the voltage will scale by 24x and 2.5x respectively when compared to the MZI structure. Thus, at the same DR the energy/bit will decrease 150 times. The modulator size will become negligible (0.002 mm$^2$).

The 12 pJ/bit are distributed according to the Pareto diagram shown in fig.9. In order to make significant advances in the third generation of the Silicon Photonics technology, it will be necessary to introduce a new class of modulators (based on the micro ring resonators) and the Wavelength Division Multiplexing (WDM). The micro ring modulators have already been demonstrated and well characterized [6]. Micro rings are optical structures [7], 200 times smaller in size with respect to the today MZI. These devices are suitable to realize DWDM (Dense Wavelength Division Multiplexing) filters mux’s and demux’s (multiplexer, demultiplexer) as well as optical switches (see fig.10 and fig.11).
mm²/micro-ring). We can imagine to use 4 lasers each with a different wavelength spaced by a $\Delta \lambda$ (a coarse spacing could mitigate the thermal constraints) among each other. These improvements allow to increase the transceiver aggregate bandwidth to 800 Gbps on 4 links in an estimated transceiver size of about 52 mm².

Figure 10 – Ring resonator schematic and frequency response

This third generation of the Silicon Photonics technology could be mature for massive production in 2018. For this generation, by using 4 lasers, we can estimate a total energy per bit from 2 to 2.5 pJ/bit, still being compatible with the same module form factor and power constraints. Based on these assumptions, the third generation could achieve a cost scaling factor of 6.66x, by allowing 0.32 $/Gbps. By changing the fiber in/out assembly from vertical to horizontal, we could also expect the module thickness to scale down by a factor 2. Thus an opto transceiver module of 800 Gbps could fit in about 5000 mm³. This last size compression could be achieved in two ways: either by modifying the end finishing of the fiber block, realizing a 45° mirror that couples the SM fiber light to the GC (this will also eliminate the fibers curvature issue), or by using the low-loss inverted taper coupler for silicon-on-insulator ridge waveguides [8], as represented in fig. 12. So, in 2018, we could imagine the Silicon Photonics pervading the on board inter-chip applications as well as gain market share in the small form factor 2D-3D opto-asic applications. When this will happen, a further growth in volumes will occur being fully compliant with 300 mm wafers fabs. The third generation of the Silicon Photonics technology will get a Pareto diagram (energy/bit) close to the one shown in fig. 13. It is important to notice that the Silicon Photonics transceiver size will be now dominated by the optical I/Os, the lasers micro-packages and the electrical I/Os. The fourth generation of the Silicon Photonics technology will find its applications in the intra–chip communications [6]. Thus, for this generation, the Silicon Photonics will be used to realize an optical communication layer capable to satisfy the required bandwidth for intra-chip communications with almost zero latency. The integration of the multi-wavelength lasers sources must be introduced in this generation by solving all the issues related to the realization of low cost, highly integrated external optical DWM optical sources. The electrical I/O’s will consist of very dense vias connecting the CMOS electronics to the photonic devices. The communications, at distances inside 1mm radius, at clock frequency between 3-5GHz, could remain based on traditional metals, while the wide bandwidth interconnections among clusters of processors and memories will be based on a Silicon Photonics layer.

Figure 11 – Transmission spectrum of a 40-µm-radius micro ring resonator and zoomed-in spectrum of an optical mode with a 1.2 pm line width

That layer could be back end post processed after the wafer to wafer molecular bonding [9]. Let’s imagine to interconnect 16 clusters with a clock of 4 GHz and I/O data of 64 bits. For each cluster we need an aggregate bidirectional bandwidth in excess of 256 Gbps. The numbers of nodes will be 120, to stay in a sustainable transceiver power budget of 20W we need to achieve an energy/bit below 700 fJ/bit. Advanced R&D Labs [10] have already shown hybrid micro-ring lasers with electrical power consumption 3x times lower than present DFB lasers. The electrical and optical characteristics of these micro-lasers are shown in fig. 14. This may be the way to integrate the lasers and simultaneously scale down the laser energy/bit below a value of 250 fJ/bit. That will relax the photodiode and TIA noise constraints. Thus a reduction of the power consumption of 3x may be expected from less requirements of TIA and LA (Limiter Amplifier), while the output buffer will simply disappear.

This fourth generation of Silicon Photonics could be envisaged in massive production by 2021. The total volume,
included the 3D package constraints, has been estimated to be below 150 mm³. At this time, the Silicon Photonics could finally enter in the more traditional high volume silicon VLSI domain.

The single Silicon Photonics transceiver, based on 16 μ-ring lasers, each with a power of 50 mW, could achieve 800 Gbps/WG (3200 Gbps on 4 + 4 wave guides) by fitting a total estimated size of 12.5 mm².

The cost/Gbps of an equivalent transceiver, after molecular bonding and post processing, could be estimated in 0.04 $/Gbps. The main challenges will be in the thermal IC control that will need a temperature stabilized inside T_{ref} +/- 10 °K, and the wafer to wafer alignment and planarization. On the temperature issue, advanced R&D labs are already working on the fluidic cooling [11] and SoD (Silicon on Diamond) platform [12]. Other issues to be addressed are related to EO VLSI packages. As a conclusion of this section, we can summarize that by 2021 a fourth generation of the Silicon Photonics technology could be ready for the massive production by allowing an energy/bit of 600-800 fJ/bit, a total aggregate bandwidth of 3200 Gbps with and a cost below 0.04 $/Gbps in a total size of about 150 mm³.

IV. SILICON PHOTONICS ROADMAP

Based on the assessment of the today Silicon Photonics technology, described in section I, and the new class of photonics devices already demonstrated in the R&D labs, we could envision a Silicon Photonics technology road map as synthesized in fig. 15. The reference generation is based on the assessment of the today technology built on a 200nm SOI wafers with CMOS 130nm. This technology is already in production by meeting its main application in the active optical cables. The reference key characteristics are summarized in a total aggregate bandwidth of 40 Gbps on 4 links, with energy per bit below 20 pJ/bit. The Silicon Photonics chip is integrated in a QSFP module form factor with a total cost close to 4 $/Gbps. Based on these characteristics and by porting this technology on 300nm SOI wafers, with CMOS 65nm , we can expect to improve the four key figures of merit by a factor 3x. Thus, we can increase the aggregate bandwidth and improve the optical losses. We expect the second generation to be mature for industrial application in 2015. The driving applications of this generation can be envisaged in HPC and modern data centers where the intensive computing and huge routing are strongly requiring wide bandwidth at low power and low latency. The volumes are expected to grow with the consequent reduction of the overall cost. The cost is scaled down by the aggregate growth and by the energy reduction that is indirectly impacting the cost of maintenance and cooling. The technology breakthrough offered by the micro ring modulators will dramatically scale down both the modulator size and the modulators energy. The micro ring resonators will also be very efficient for the introduction of the coarse wavelength division multiplexing. We expect this generation to be mature for massive production in 2018. By thinking to this generation, associated with an improved electronics (CMOS 28 nm), we could foresee an aggregate of 800 Gbps with energy per bit below 3pJ/bit. Thanks to these breakthroughs the third generation could dramatically decrease both the cost and the form factor size. Among the features offered by the Silicon Photonics the low latency will also play a key role by allowing that generation to offer big advantages in the chip to chip applications. Here we can think to a Silicon Photonics interposer capable to scale down the multi layers electrical board size of almost 5 times. If this will happen we could also expect a consistent volume growth of the Silicon Photonics market (10MU/Y, 20kWs/Y). Based on the recent R&D demos, we could forecast that another key breakthrough will allow to integrate the multi-wavelength micro-ring-lasers by allowing the usage of the dense wavelength division multiplexing. Two other 3D technologies, as the wafer to wafer molecular bonding and the fluidic cooling or black diamond, could be the enablers for the Silicon Photonics fourth generation. This generation,
that we estimate ready for massive production in 2021, could enable the intra-chip interconnections with the characteristics shown in the table of fig. 15. When this will happen, the traditional semiconductors silicon volumes will finally occur and the intra-chip communications, today based on copper, will be achieved by the multi-λ integrated nano-photonics.

![Image of Silicon photonics road map](image)

**Figure 15 – Silicon photonics road map**

### V. CONCLUSIONS

This paper, starting from an analysis of a mature Silicon Photonics technology and the most advanced R&D demonstrators and by introducing a new class of micro and nano photonics devices, provides a vision of a possible road map, up to 2021. Though this road map is based on assumptions and challenges that must be further verified and validated, the Silicon Photonics technology is showing features capable to provide an answer to the increasing demand of short distance, wide bandwidth, low power and low cost wired communications. Volumes are expected to increase while the distances are becoming shorter from tens of meters to millimeters. More the distances will scale and more the Silicon Photonics technology will become efficient and competitive versus copper and hybrid photonics. Finally we have shown how the technology road map is also consistent with the required miniaturization for the efficient communications at intra-chip level.

### VI. ACKNOWLEDGMENTS

This paper has benefited from the cooperation and technical discussion with the members of my team: G. Chiaretti, A. Fincato, A. Pallotta and E. Temporiti, as well as many people inside and outside STMicroelectronics. A particular thank to P. De Dobbeleare [Luxtera] and S. Rotolo for proof-reading this article and providing comments to improve the final version.

**References**


