Silicon Photonics/Optoelectronics for Interconnects for high bandwidth computing

A literature review of current Silicon Photonics/Optoelectronics technology and its application to computer interconnects.

Name: Andrew Cooke

Student Number: 117379906

Coláiste na hOllscoile Corcaigh, Éire
University College Cork, Ireland

Introduction:

High performance computing has an extensive range of applications in industry and has a significant impact on our lives. High performance computing has use in weather forecasting, modelling geophysical systems (e.g. simulating seismic events) and medicine (simulating viruses such as Hepatitis C and simulating the brain to better understand how diseases such as Alzheimer's work)[1], to name but a few.

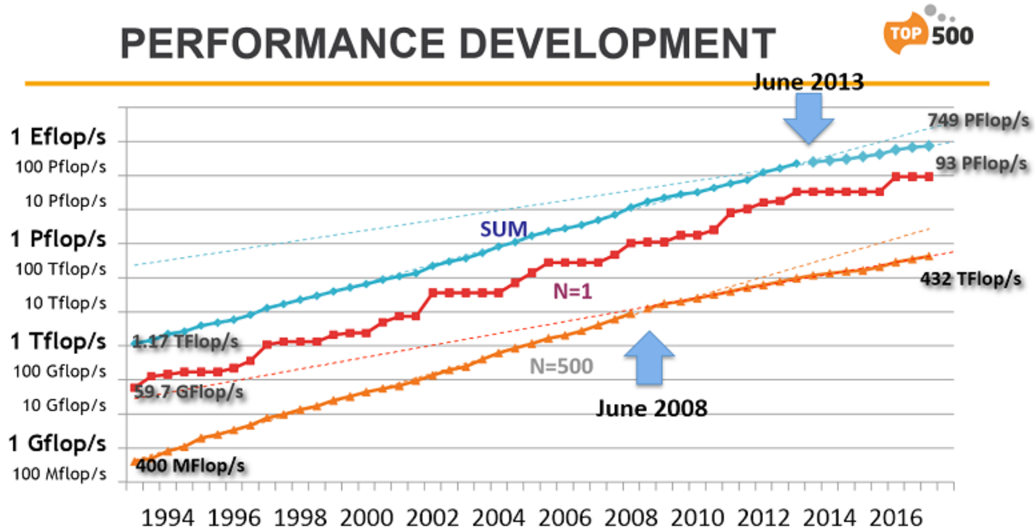

Figure 1: The historical performance of supercomputers, showing the growth of the top ranked computers in the Top500 list, the bottom ranked, and the performance sum of all the computers making up the list[2].

The performance metric for supercomputers is "FLOPS", which is the number of floating point operations performed per second. Historically, the growth in supercomputer performance (shown in Figure 1) has been driven by the ever shrinking size of transistors and improvements in chip design, these two reasons also fuelled the growth predicted by Moore's Law[3], which states that the number of transistors on an integrated circuit doubles every two years. However, the decrease in transistor size can't go on forever, as they're physically limited by atomic size. Performance enhancements must be applied to other parts of the computer architecture in order to continue this growth in performance.

Electrical interconnects (e.g. on-chip traces) are a major bottleneck for computer performance for a number of reasons. As the interconnects consist of metal traces, they have a parasitic resistance and capacitance. As well as this, the metal traces have other problems that occur during the transmission of high frequency signals, such as skin effect[4], where the current density has a non-uniform distribution which is concentrated at the outer surface of the conductor (i.e. the skin). This results in frequency dependent losses due to greater parasitic resistance (cross sectional area is reduced). Other high frequency effects include cross talk[4], where the signal in one conductor has a disruptive effect on the signal in an adjacent conductor. Therefore, improving the performance of supercomputers by increasing the bandwidth has its limits with electrical interconnects.

A solution to this problem relies on using optical interconnects in high performance computers, thereby increasing the bandwidth and overall performance of the computer, as the speed of light is much greater than the drift velocity of an electron. Optical interconnects do not suffer from the same limitations as electrical interconnects- due to the nature of the optical signal, the impedance to the electromagnetic wave is much lower than the associated resistance in the electrical interconnects. Optical interconnects at this scale are frequency and distance independent, i.e. the interconnect for a 500MHz signal may well be used to transmit a 500GHz signal, and the distance of the interconnect along which the signal travels will not affect signal degradation[4]. Optical interconnects also solve problems with physical density. Due to technologies such as Wavelength-Division-Multiplexing (WDM), where multiple optical signals are combined (or multiplexed) onto a single mode waveguide, the number of interconnects required (on-chip and off-chip) can be greatly reduced as the bandwidth density of each interconnect is much larger.

Clearly, integrating photonics with computer architecture to allow greater bandwidth computing is the next step in high performance computing. There are currently two integration strategies being used; monolithic integration and heterogeneous integration. Monolithic integration involves combining multiple optical and electronic devices on a single chip[5-7], whereas heterogeneous integration involves using chip to chip interconnects to connect electronic and optical devices on different chips[8-10]. Monolithic integration would be the ideal integration strategy for optoelectronics as it would allow greater device density and therefore greater processing power (larger number of devices available in a smaller area). Monolithic integration of silicon optoelectronics may one day be commercially viable, however there are currently a number of barriers that must be overcome before a complete suite of optoelectronic devices is available for integration. The primary problem is producing an efficient, electrically pumped silicon laser, which will be covered later on in the report. In the past, off chip electrically pumped lasers were used as a light source for optoelectronics, however this does not represent true monolithic integration.

A number of devices are required for optical interconnects in computers- a light source (usually a laser), a waveguide medium for the light to travel down, a modulator to encode data in the optical signal, a (de)multiplexer to reduce the number of input/output ports required and an optical coupler to couple light from the single mode fibre to the waveguide medium. The challenge is to fabricate all of these devices from silicon in order to monolithically integrate them onto one chip.

Current Technological Position and Associated Issues:

The waveguide medium is an essential feature of the optical interconnect, allowing the light to travel from one part of the chip to another. The waveguide is characterised by its two main components, the core medium along which the light travels and the outer cladding surrounding the core. These two components must have a high index contrast- the inner core should have a relatively high index of refraction, and the cladding should have a lower index of refraction than the core[11]. Traditionally, fibre optic communications occur at a wavelength of 1.55μm. Fortunately, silicon is transparent at this wavelength, making it a great material for the waveguide medium. As well as this, it has a high index of refraction, in the range of 3.42 – 3.48 , whereas silicon oxide (and silicon dioxide) have an index of refraction in the range of 1.4 – 1.55. Therefore silicon and its oxides would make an ideal material for the core and cladding respectively. It's no surprise that the majority of approaches to monolithic integration of silicon optoelectronics utilise some custom Silicon on Insulator (SOI) waveguide. A thick buried-oxide (BOX) is usually used to prevent the optical signal from coupling into the silicon substrate that forms the basis for CMOS computers.

The fabrication of an all-silicon laser for monolithic integration with silicon optoelectronics has represented a huge engineering challenge. As silicon is an indirect bandgap material, it is difficult to get it to emit light due to low rates of radiative recombination (in indirect bandgap materials the electron and hole have different crystal momentums, therefore a phonon with momentum equal to that of the difference between the electron and hole is usually required).

However, attempts to make an all-silicon laser have been made. A continuous wave, low threshold silicon laser has been experimentally demonstrated[12], based on the principle of stimulated Raman scattering. This type of laser requires optical pumping. The photons provided by the optical pump laser are inelastically scattered by the silicon material (incident photon loses kinetic energy), and are re-emitted at a lower frequency (stimulated Raman scattering). The experimental setup used a silicon rib waveguide with two heavily doped silicon slabs (p^+ and n^+) on either side, forming a p-i-n junction. The optical cavity was produced by coating one silicon waveguide facet with a high bandwidth reflective coating. The external laser source was a continuous wave, 1.536µm laser which was amplitude modulated to give a high frequency (10kHz) square wave output. An Er-doped amplifier is used to increase the optical pump beam power to 2W. The optical pump beam travels through a Wavelength De-Multiplexer (WDM) and is coupled to the waveguide by a lensed fibre. The reflected pump beam and Raman laser output from the silicon waveguide facet are recoupled and separated by a WDM. A reverse bias voltage of 25V is applied to the p-i-n junction in order to sweep the electron-hole pairs (produced by two-photon absorption) out of the silicon waveguide, as they incur significant optical losses due to the free carrier plasma dispersion effect. This experiment resulted in a laser threshold of 0.4mW, a slope efficiency of 10% and a Raman/Stokes wavelength of around 1.67µm. Rong et al[12] noted that a micro ring resonator could be used instead of the reflective coating, which was later experimentally demonstrated[13].

This solution to making a silicon laser is not ideal, as it still requires another off-chip laser for optical pumping. The most promising solution would be an electrically pumped silicon laser that could be fabricated on the silicon chip.

Another issue arises in coupling the optical fibre carrying the light from the laser source with the silicon waveguide. The diameter of optical fibre (around 85µm^2[14]) is traditionally much greater than that of the silicon waveguide (around 0.1µm^2[14]), which gives rise to difficulties in focusing the light from the fibre onto the waveguide for transmission.

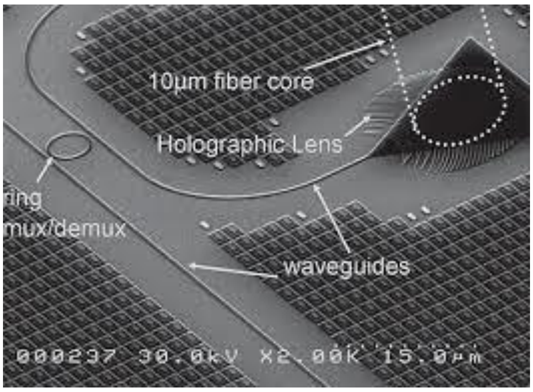

Figure 2: The Holographic lens developed by Luxtera, which takes light incident at 90 degrees to the die and focuses it on the silicon waveguide[11].

Luxtera solved this problem by developing a holographic lens[11], shown in Figure 2, which is a structure etched into the silicon whose function is to take light normal to the plane of the die and turn it 90 degrees to be focused on the waveguide. Luxtera fabricates this structure from transistor gate polysilicon, depositing it directly onto the silicon die where the necessary etching is performed. The fact that Luxtera utilises gate polysilicon makes this design compatible with monolithic integration.

In order to achieve denser integration of optical interconnects, it's necessary to utilise a technology called Wavelength Division Multiplexing (WDM). WDM involves combining (or multiplexing) multiple optical signals to be transmitted along a single waveguide medium or optical fibre by using different optical wavelengths. The utilisation of WDM can reduce the number of input/output ports to an optoelectronic IC by greatly expanding the bandwidth density of the interconnect, as it allows multiple data streams to use the same waveguide[15]. Sun et al[15] describe a WDM system using optical ring resonators.

An optical ring resonator functions on the principles of total internal reflection, resonance, optical coupling and interference. Light travelling along an adjacent waveguide medium can be coupled into the ring resonator, however this is dependent on a number of factors; the refractive index of the waveguide medium, the refractive index of the ring resonator material and how close the waveguide medium is to the ring resonator. Once the light is coupled with the ring resonator, the design of the resonator must ensure that total internal reflection occurs and the light can't escape, which mainly boils down to the material it's constructed with. Fortunately, this paper is focused on utilising silicon photonics. The disparity between silicon's index of refraction (3.42 – 3.48) and its oxide (silica) (1.4 – 1.55) make it an ideal material for this application. Resonance in the optical ring resonator is governed by equation (1), where m is an integer, λ is the resonant wavelength of the light in the ring, $\eta(eff)$ is the refractive index of the ring and R is its radius[16].

$$m\lambda = \eta(eff)2\pi R \qquad (1)$$

From this equation we can see that resonance will occur when an integer number of wavelengths fit inside the length of the ring. Constructive interference occurs once resonance is achieved in the ring[16]. This new wave then leaves the ring by optical coupling. It's important to note that light will only be coupled into the resonator if the light is of the resonant wavelength, which is affected by the effective refractive index as shown in the equation above.

The system described by Sun et Al[15] involves using continuous wave laser light with n wavelengths (λ) which is coupled onto a single-mode waveguide using a vertical grating coupler. An array of n microring resonators is then used to modulate each wavelength of light λ with a distinct bit stream. A form of amplitude modulation called On-Off Keying is used, where logical 1 is represented by the presence of an optical signal, and logical 0 is represented by the absence of an optical signal. A simple transceiver circuit based on this principle is shown in Figure 3. Electro-optical microring modulation can be achieved by shifting the resonant wavelength λ_0 of the ring resonator. This is possible via changing the effective refractive index of the microring resonator, η(eff), which can be done by varying the free carrier concentration[17]. In forward bias, carrier-injection modulators inject carriers into the intrinsic region of the p-i-n junction, resulting in a smaller resonant wavelength λ_0. In reverse bias, the carrier-depletion modulators deplete the carriers from the p-n junction, thereby increasing the resonant wavelength λ_0[15]. Either of these shifts in the resonant wavelength λ_0 will result in resonance not occurring due to the disparity with the wavelength of the optical signal λ, therefore transmitting a logical 0. These modulated optical signals with wavelength λ then leave the array of ring resonators and are coupled by another vertical grating coupler onto a single mode waveguide where they are transmitted to the receiver circuit.

Figure 3: Diagram showing a possible transceiver circuit using optical ring resonators[15].

The receiver circuit operates on much the same principle. An array of micro ring resonators are used to demultiplex the optical signal into each of the n wavelengths, where they're detected by a photodiode and serialized into a different data format, to be stored or transmitted again.

Whereas active electro-optical ring modulators were used by Sun et al[15], other applications involve using ring resonators with a large free spectral range as a notch filter. Other modulator designs do exist, such as the Mach-Zehnder modulator[18]. The Mach-Zehnder modulator splits the optical signal into two separate waveguide mediums, where a phase shift is induced in each arm. The design used by Barwicz et al[18] utilised a silicon rib waveguide with two heavily doped silicon slabs (p^+ and n^+) on either side, forming a p-i-n junction. The phase shift is induced by applying a voltage to the p-i-n junction, which would alter the free carrier concentration, thereby changing the refractive index and inducing a phase shift. The phase shift induced depends on whether a logical 1 or 0 is desired. If a logical 0 is desired, the phase shifts induced would be such that the two optical signals interfere destructively, and vice versa for a logical 1.

Clearly a number of designs are available for an optoelectronic modulator to be monolithically integrated on a silicon die.

Finally, in order to convert the optical signal back into an electrical signal that can be used to perform calculations, it's necessary to use a photodiode that has a very high responsivity and speed at the 1.55μm wavelength. Unfortunately the property of silicon that makes it a great waveguide medium at the 1.55μm wavelength means it's a terrible material to make a photodiode with (at this wavelength). Thankfully, Ge-on-Si photodiodes have been shown to possess these characteristics. The responsivity of an integrated germanium film on silicon has been demonstrated to be 0.45 A/W[18]. However, these photodiodes may not perform as well at extremely high frequencies (around the 100s of GHz range). This is due to the disparity between the rise and fall times demonstrated with an impulse response, a rise time of 100ns was recorded with a corresponding fall time of 700ns[18]. At high frequencies, these rise and fall times may begin to overlap, disrupting the electric current produced in the diode. While Ge-on-Si photodiodes serve a purpose here, they do not represent total monolithic integration of silicon optoelectronics, obviously due to the presence of the germanium film described by Barwicz et al[18]. However, an all-silicon monolithically integrated photodiode may not be possible, again due to the fact that it's transparent to light at the 1.55μm wavelength, which again is the typical wavelength used in optical communications.

Current Research Areas:

It's clear that much of the technology to develop monolithically integrated optoelectronic devices is available, and they have a clear application in improving the performance of computers by increasing their bandwidth. It's no surprise that developing the first monolithically integrated optoelectronic chip is a very hot research area right now.

In late 2015, researchers at MIT produced a microprocessor that uses on-chip optical interconnects for high bandwidth computing[19]. The chip was fabricated using CMOS Silicon on Insulator processes, as a 'zero-change' approach was taken to integrate the optical and electronic devices on one chip in order to demonstrate that optoelectronic chips can be produced using existing high-volume manufacturing processes. A thick buried oxide layer (BOX, as mentioned above) was used to separate the silicon waveguide from the silicon-handle wafer. However, this resulted in high optical losses in the waveguide as the BOX layer was less than 200nm thick resulting in the light from the waveguide coupling with the silicon-handle wafer. The silicon-handle wafer was etched from underneath the areas with optical devices as a result. A 1.18μm continuous wave laser source was located off-chip. The transmitting circuit utilised a microring resonator to modulate the transmission signal (as described above) and an SiGe photodiode was used in the receiver circuit to convert the optical signal to an electric current. A memory bandwidth of 5Gbps was demonstrated, however this was only utilising a single wavelength of light. In future, using multiple wavelengths of light in conjunction with WDM could provide a memory bandwidth of up to 55Gbps without the need for additional fibres.

Ayar Labs, an MIT spin off company, is just one company developing optoelectronic integrated circuits utilising optical interconnects for high bandwidth computing[20]. The TeraPHY silicon chip is their flagship product. The chip will be available in two versions; a 1.6Tbps and 3.2Tbps bandwidth version. The chip has four or eight (depending on the version) input/output single mode fibres, each with a bandwidth of 400Gbps. The chip design includes Wavelength Division Multiplexing, modulators and photodetectors, as discussed above. The TeraPHY chip does not represent total monolithic integration of the optoelectronic components. The laser light source is located off-chip on the 'SuperNova' photonic integrated circuit, which can generate up to sixteen wavelengths of light which are then multiplexed. The power is then split and amplified to four output ports connected to the TeraPHY chip. Very Short Reach links operating at 50 Gbps are used for the electrical interface on the TeraPHY chip. The TeraPHY chip is fabricated such that it can use either on board or in-package electrical traces to interface with the customer's chip. Unfortunately, system reference designs to demonstrate the functionality of the chip are currently unavailable.

Conclusion:

In conclusion, optical interconnects offer great prospects for high bandwidth computing in order to drive the next generation of high performance computers. This review focused primarily on the problems associated with integrating these optoelectronic devices with the current CMOS platform, and also their possible solutions. It's clear that there are solutions to many of the issues such as modulation of the optical signal, reducing the density of the devices (utilisation of WDM), and a suitable waveguide medium. However, there are currently a number of issues that still need to be addressed, such as an electrically pumped all-silicon laser and a silicon photodetector. It's worth noting that there may not be a physical solution to these problems, for example if 1.55µm is used as the standard wavelength in optoelectronics for computing it'd be difficult to manufacture a silicon photodiode due to its transparency at this wavelength. The fact that Ayar Labs is already working on a commercial optoelectronic chip means the outlook is bright for high bandwidth computing using optoelectronics for interconnects.

References:

[1] E. C. Joseph, C. Dekate, and S. Conway, "Real-world examples of supercomputers used for economic and societal benefits: A prelude to what the exascale era can provide," *International Data Corporation, May,* vol. 20, 2014.

[2] M. Feldman. (2017). *TOP500 Meanderings: Sluggish Performance Growth May Portend Slowing HPC Market*. https://www.top500.org/news/top500-meanderings-sluggish-performance-growth-may-portend-slowing-hpc-market/

[3] G. E. Moore, "Cramming more components onto integrated circuits," ed: McGraw-Hill New York, NY, USA:, 1965.

[4] D. A. Miller, "Rationale and challenges for optical interconnects to electronic chips," *Proceedings of the IEEE,* vol. 88, no. 6, pp. 728-749, 2000.

[5] S. Assefa *et al.*, "A 90nm CMOS integrated nano-photonics technology for 25Gbps WDM optical communications applications," in *2012 International Electron Devices Meeting*, 2012, pp. 33.8. 1-33.8. 3: IEEE.

[6] J. Li, G. Li, X. Zheng, K. Raj, A. V. Krishnamoorthy, and J. F. Buckwalter, "A 25-Gb/s monolithic optical transmitter with micro-ring modulator in 130-nm SoI CMOS," *IEEE Photonics Technology Letters,* vol. 25, no. 19, pp. 1901-1903, 2013.

[7] M. Georgas *et al.*, "A monolithically-integrated optical transmitter and receiver in a zero-change 45nm SOI process," in *2014 Symposium on VLSI Circuits Digest of Technical Papers*, 2014, pp. 1-2: IEEE.

[8] X. Zheng *et al.*, "2-pJ/bit (on-chip) 10-Gb/s digital CMOS silicon photonic link," *IEEE Photonics Technology Letters,* vol. 24, no. 14, pp. 1260-1262, 2012.

[9] C. Li *et al.*, "Silicon photonic transceiver circuits with microring resonator bias-based wavelength stabilization in 65 nm CMOS," *IEEE journal of solid-state circuits,* vol. 49, no. 6, pp. 1419-1436, 2014.

[10] M. H. Nazari and A. Emami-Neyestanak, "A 24-Gb/s double-sampling receiver for ultra-low-power optical communication," *IEEE Journal of Solid-State Circuits,* vol. 48, no. 2, pp. 344-357, 2013.

[11] C. Gunn, "CMOS photonics/spl trade/-SOI learns a new trick," in *2005 IEEE International SOI Conference Proceedings*, 2005, pp. 7-13: IEEE.

[12] H. Rong *et al.*, "A continuous-wave Raman silicon laser," *Nature,* vol. 433, no. 7027, p. 725, 2005.

[13] H. Rong *et al.*, "Low-threshold continuous-wave Raman silicon laser," *Nature Photonics,* vol. 1, no. 4, p. 232, 2007.

[14] Q. Xu, B. Schmidt, S. Pradhan, and M. Lipson, "Micrometre-scale silicon electro-optic modulator," *nature,* vol. 435, no. 7040, p. 325, 2005.

[15] C. Sun *et al.*, "A monolithically-integrated chip-to-chip optical link in bulk CMOS," *IEEE Journal of Solid-State Circuits,* vol. 50, no. 4, pp. 828-844, 2015.

[16] W. Bogaerts *et al.*, "Silicon microring resonators," *Laser & Photonics Reviews,* vol. 6, no. 1, pp. 47-73, 2012.

[17] R. Soref and B. Bennett, "Electrooptical effects in silicon," *IEEE journal of quantum electronics,* vol. 23, no. 1, pp. 123-129, 1987.

[18] T. Barwicz *et al.*, "Silicon photonics for compact, energy-efficient interconnects [Invited]," *Journal of Optical Networking,* vol. 6, no. 1, pp. 63-73, 2007.

[19] C. Sun *et al.*, "Single-chip microprocessor that communicates directly using light," *Nature,* vol. 528, no. 7583, p. 534, 2015.

[20] R. Meade *et al.*, "TeraPHY: A High-density Electronic-Photonic Chiplet for Optical I/O from a Multi-Chip Module," in *Optical Fiber Communication Conference (OFC) 2019*, San Diego, California, 2019, p. M4D.7: Optical Society of America.