Integrated Photonics

Mario Paniccia, Mike Morse, and Michael Salib

Intel Corporation, Santa Clara CA 95054, USA
{mario.paniccia,mike.morse,michael.s.salib}@intel.com

Abstract. This chapter attempts to bring an industrial perspective to the topic of silicon photonics integration. Comparisons between microelectronics and silicon photonics are made. Monolithic and hybrid integration are discussed throughout the chapter with attention to cost, manufacturability, and performance. Potential applications, design issues, and packaging issues are addressed. Special attention is given to processing problems common in silicon-based optical systems.

1 Introduction

The need to transmit signals optically is becoming increasingly apparent as data rates exceed 1 Gb/s. Such high rates are straining the capabilities of electronics signaling technology. High-speed electronic transmission over copper is currently limited to distances around 100 m, and this distance will certainly shrink as data rates rise. If an integrated silicon photonic chip could be developed, it could bring significant, even revolutionary changes, to the next generation communications industry, by radically altering the price, power, and size for photonic components.

This chapter will draw parallels between the manufacturing of photonic devices and transistors in silicon. From the first integrated circuit in 1958 to today's high-end microprocessors that contain over 500 million transistors on a single chip, silicon remains the material of choice for electronics. Decades of industrial research and investment in silicon manufacturing have brought significant economic and technological advances to the semiconductor world, largely driven by the ability to reduce transistor size. These advances enable the creation of devices with increased performance at significantly lower costs. Figure 1 plots the number of integrated transistors per chip as a function of time since 1970. At this rate, there will be over one billion transistors on a single die before 2007. Figure 2 shows the resulting increase in processor performance as measured in Million Instructions Per Second (MIPS) and the corresponding decrease in cost per MIPS since 1985. If silicon photonics were to make similar advances as those made in microelectronics over the past thirty years, then this would create a radically new communications landscape.

Optical devices are limited in size by the wavelength of light routed through them. Due to the high refractive index of silicon one can make significantly smaller devices as compared to those made in a low index medium

L. Pavesi, D. J. Lockwood (Eds.): Silicon Photonics, Topics Appl. Phys. **94**, 51–90 (2004)

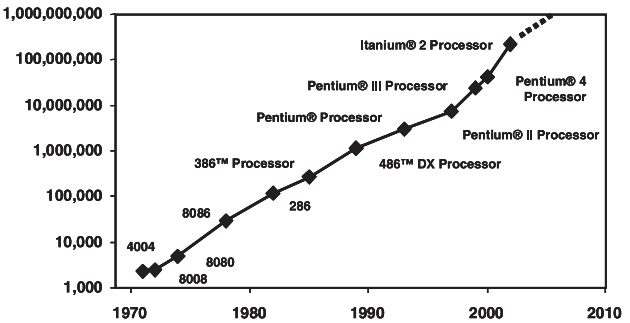

Fig. 1. Number of transistors on a die as a function of time since 1970

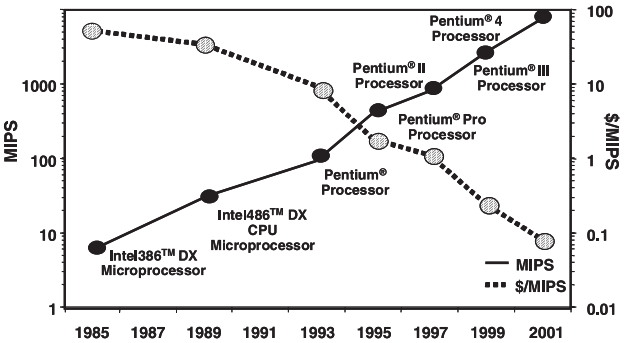

Fig. 2. Plot depicting the net result of Moore's law scaling as measured in $/MIPS (Million Instructions Per Second) as a function of time

(e.g. silica). All of this, however, comes with challenges such as increased scattering losses due to sidewall roughness, increased sensitivity to Polarization Dependant Loss (PDL), and increased coupling losses [1, 2].

Many reviews have been written since the early 1990's describing silicon based optoelectronics devices [3, 4, 5, 6, 7, 8, 9], and although much progress has been made, many technical challenges remain. Most experts see silicon as lacking in many properties needed for high quality optical performance. Three key properties are light emission, fast modulation for data encoding and low coupling loss between silicon waveguides and optical fibers.

Silicon is not the only material in which optical integration is proving difficult. To date, there has not been much commercial success with monolithic integration, even with preferred III–V optical materials. This is because integration for most devices has not demonstrated improved performance or a reduced cost. The reasons for this include complicated epitaxial regrowth steps needed for III–V materials that significantly reduce yield and issues with facet formation. Weak market demand for integrated optics has also been a factor. If, however, silicon based photonic devices could be developed

using processes similar to baseline CMOS processes, one could achieve much higher yielding optical devices at lower cost.

Developing cost effective integrated optical devices in silicon will require an understanding of the issues related to processing, testing, and packaging. The ability to take advantage of existing fabrication infrastructure and knowledge greatly improves the chance that silicon-based optical devices could be produced in high volume at a low cost. This chapter will present an industrial perspective of creating integrated silicon photonic components.

2 Applications for Integrated Silicon Photonics

To overcome the limitations imposed by electronic communication, optics has been introduced at longer distances. Currently, it is used for almost all high-speed (> 1 Gb/s) interconnects running longer than 100 m, and is competing in 1–100 m commercial applications. Figure 3 shows a diagram of the applications for which photonics already exist, or is being considered. At the longest distances, dominated by telecommunications at 1550 nm, photonic components are being used for lasing, amplifying, multiplexing/demultiplexing (MUX/DEMUX), and attenuating functions among others. Ethernet and Fibre Channel protocols are typically used for shorter distances including enterprise and storage area networks (SAN). In general, these systems are less complex than their long-haul telecommunication counterparts, but some are implemented in dense wavelength division multiplexing (DWDM) systems. At intermediate distances of 1–100 m, both copper and fiber solutions coexist for applications such as rack-to-rack connections. Optical solutions here are relatively simple with no amplification or multiplexing involved. At even shorter distances, electrical signaling dominates for technical and economic reasons. Photonic components will have to be tightly integrated with the pre-existing infrastructure and have very low costs to be accepted in this space.

A discussion about the merits of optical interconnects is beyond the scope of this chapter, but others have written on this matter [10, 11, 12]. Instead, the insertion point of optics will be summarized. Figure 4 shows an estimate of the transition from electrical to optical communication over different distances and data rates. There is clearly some uncertainty in the timing because the electrical performance of silicon is constantly improving. Electrical data rates of greater than 8 Gb/s have already been demonstrated with limited equalization [13, 14, 15, 16] and modeling has predicted 10 Gb/s data rates for up to half a meter [17]. Companies like Intel are even more optimistic about extending the performance of copper to at least 20 GHz [18]. Since chip-to-chip interconnects are not predicted to reach such bandwidth until the next decade, optical signaling will not be implemented soon for this application. However, board-to board and rack-to-rack applications will need optical interconnect sooner. Since the requirements for silicon photonics

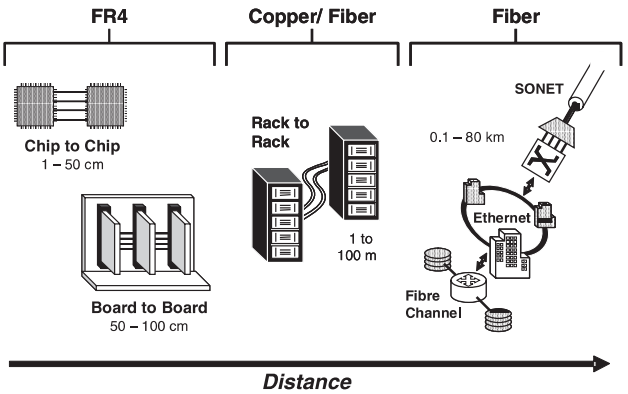

Fig. 3. Applications for communications. Optical technologies are currently only being used for longer links

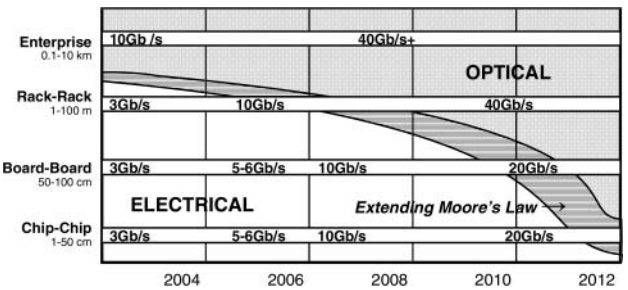

Fig. 4. Timeline for the transition between electrical and optical communication for various distances

are very different for the chip-to-chip application as compared to all other applications, separate sections of this chapter will cover the possible integration routes for each. The majority of the attention will be devoted to the chip-to-chip communications since that application offers the largest volume opportunity for the integration of photonics and microelectronics.

2.1 Chip-to-Chip Applications

Several reviews have discussed the prospects of optical interconnects for chip-to-chip communication [17,19,20,21,22] in depth, and the Chapter by *Gaburro* in this book will more thoroughly cover the material. It is important however to understand the current system design and performance as a starting point to a discussion of system constraints.

Figure 5 shows a typical architecture of an Intel CPU and chipset. The most demanding interface is the front side bus between the microprocessor and the memory controller hub (MCH), a chip that serves as the link to the graphics and memory functions. Below this is the input/output (I/O)

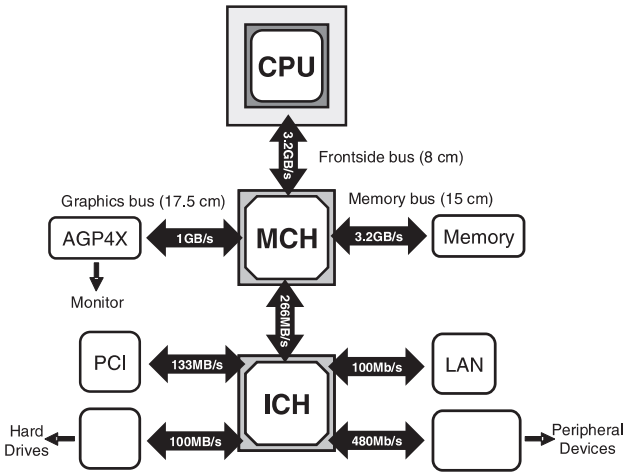

Fig. 5. System architecture of a desktop PC, showing the connectivity of the microprocessor and chipsets with the aggregate data rate and length

controller hub (ICH) that provides a connection to external ports. The speed of the link between the processor and off-chip memory is important because the processor will have to interrupt execution to wait for needed data to be fetched from memory. While microprocessor speeds have been growing exponentially for over twenty years, the front side bus speed has not kept up (Fig. 6). This means that the computer's performance is not scaling with the speed of the microprocessor. Similar problems also exist in multiple processor systems that share a front side bus. To minimize the off-chip latency, cache memory has been put directly on the microprocessor. This trend will continue, with even more cache memory being added, but architectural changes such as moving from a bus to point-to-point architectures will also help. Some of the first results of these concentrated efforts to improve the front side bus speed are seen in Fig. 6, where the most recent processors show a significant jump in performance.

Current proposals for chip-to-chip optical interconnects can be grouped into three camps: monolithic integration of emitters and detectors into silicon [23], hybrid integration of an off-chip light source with an integrated photodetector [24], and a totally hybrid approach where a III–V emitter and detector are placed on top of the silicon die [25]. In order to be competitive with electrical interconnects, any optical solution must have similar bandwidth, cost, power, and latency as the copper solution, while being capable of very high volume manufacture. The implications are that each optical channel must have a data rate of at least 20 Gb/s by 2010, and the entire cost of the link should be less than $10.

Monolithic integration is the most challenging approach due to the limitations of silicon-based light emission. Extremely low optical power output

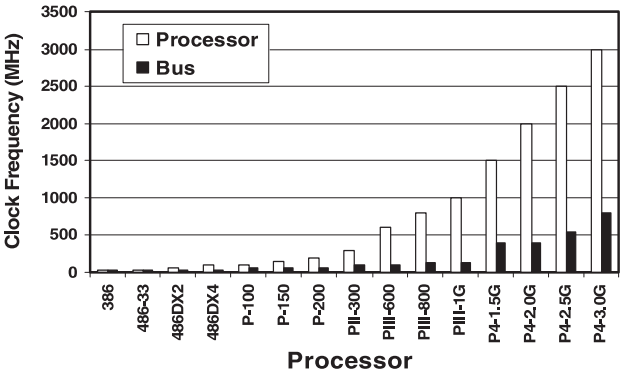

Fig. 6. Frequency gap between the microprocessor and the front side bus over the last 15 years

has plagued silicon devices, such as erbium-doped light emitting diodes, for more than a decade [26, 27]. This is discussed in more detail in the Chapter by *Fauchet* in this book. Another issue is that none of the current research in silicon light emitters promises modulation bandwidths greater than about 1 Mb/s, because of long carrier lifetimes in either silicon or excited states in the rare-earth centers. There is hope that this can be overcome by using a fast silicon-based modulator to encode data on a continuous wave light source. The vast majority of research on these modulators has been focused on waveguide-based devices, which tend to be fabricated either on thick Silicon-on-Insulator (SOI) wafers or on specialty epitaxial wafers. Neither of these two substrate solutions is particularly amenable to integration with leading edge microelectronics, which needs a substrate optimized for electrical performance. It is critical, therefore, to design a compatible solution before significant time and effort are expended on development of the optical devices. The fastest modulator data published to date is based on current injection in silicon waveguides, with a reported bandwidth of 20 MHz [4, 5], so there is a significant technology gap with the requirements of chip-to-chip communications. However, several groups around the world are addressing this issue, and current work is advancing the performance rapidly.

In the case that total monolithic integration of silicon optical and electrical components fails, or indeed as an alternative solution, III–V materials could be used to add optical functionality to silicon microelectronics. GaAs or InP-based lasers could either be bonded to the chip or packaged beside the silicon. Alternatively, an on-chip modulator could be used in conjunction with a III–V laser [20]. These modulators could be silicon, III–V, or polymer based. The integration challenges of all of these approaches will be discussed later.

The selection of the wavelength for optical communication is critical in determining candidate components for integration onto silicon. For wavelengths below 1100 nm, silicon is opaque. This means that silicon waveguides

and planar modulators are very lossy, but that silicon photodetection is efficient. Above 1100 nm the converse happens, with the absorption coefficient of silicon falling to insignificant levels. SiGe alloys could be integrated on silicon to potentially stretch the absorption edge to 1550 nm, but no commercially available, integrated SiGe photodetector has yet been made.

If the goal is to achieve true monolithic integration on silicon it makes much more sense to operate at a wavelength that is more easily detected (1310 nm or less). The design of any silicon-based light emitter needs to take this into account. If planar silicon modulators are also required, the wavelength of operation for the emitter is similarly restricted due to waveguide absorption, resulting in an approximate usable range of 1000–1310 nm.

2.1.1 Temperature Issues

One of the practical constraints for integrated optics on silicon is the performance at elevated temperatures. The worst-case thermal environment occurs when the optical devices are located directly on the microprocessor. Figure 7 shows the modeled junction temperature of a proposed server die with four microprocessor cores and a large memory cache, under two conditions. In Fig. 7a, the top two microprocessor cores are active while the two bottom cores are off. This causes a hot spot of 105 °C to form over the active cores, with the remaining surface temperature ranging from 55–95 °C. In Fig. 7b, all four cores are active, but due to the system constraints on the processor, each is running at lower power. In this case, the hot spot has been reduced to 85 °C. It is expected that devices directly bonded to the top surface would actually experience a slightly hotter temperature by some 5–10 °C. However it should not be assumed that the microprocessor acts as a heat sink for the optical devices. It is much more accurate to describe the microprocessor as a heating plate for the optoelectronics. It is very unlikely that III–V light emitters could operate for long times if they were subjected to temperatures above 80 °C. Future silicon-based light emitters might have fewer catastrophic device failures due to resistance to dark line defects, but the output power might fall so much at elevated temperatures that the optical loss budget is exceeded. In addition to surviving at the maximum temperature, the optical device has to operate over a range of temperatures as the device heats and cools under normal operating conditions for a lifetime of the order of ten years. Many of the proposed optical solutions located at the processor should be re-evaluated in light of this thermal environment [28]. Receivers and modulators are not expected to be as problematic as the light emitters since their failure mechanisms are less temperature sensitive, but their performance may be seriously compromised at high temperatures due to variations in current, refractive index and efficiency.

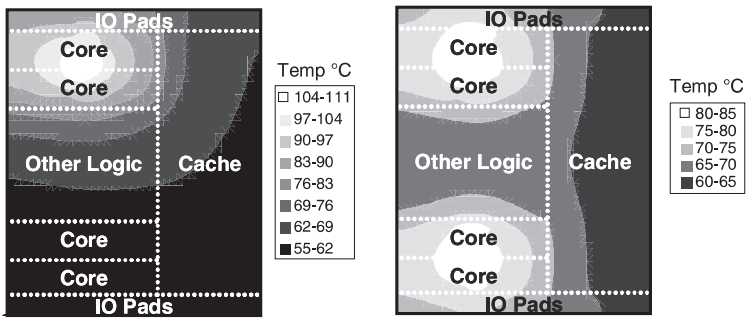

Fig. 7. Modeled thermal maps of the junction temperature for a potential server microprocessor. The die consists of four logic cores and a large cache. In (**a**), the top two cores are active and the bottom two inactive. This is the worst-case scenario for the part. (**b**) shows the case when all four cores are active, but with reduced power

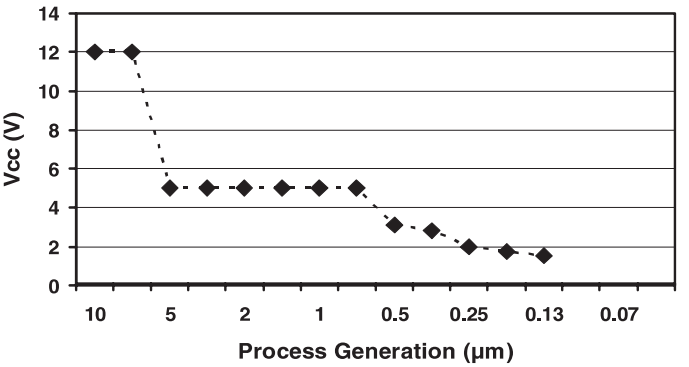

Fig. 8. Voltage scaling of microprocessors

2.1.2 Voltage Issues

The relentless pursuit of Moore's Law necessitates the reduction of voltage for every new CMOS generation. Current microprocessors are running at 1.3 V. Special circuits on the chip, commonly called charge pumps, can generate higher voltages with proper design, some possibly reaching five times the supply voltage, but they are the exception and not the rule. A higher DC source on the circuit board, for example 3.3 V, would be required at the starting point for the charge pumps in order to reach higher voltages. Figure 8 shows the voltage scaling that has been achieved, as function device critical dimension. The key point is that optical devices that operate at voltages lower than 10 V have a less of a barrier to entry into CMOS designs.

2.2 Longer Reach Applications

As mentioned earlier, several applications areas beyond chip-to-chip are viable for integrated silicon photonics, including telecommunications and data communications. The recent collapse of the telecommunications market has dramatically slowed the development and rollout of integrated optical components. Nevertheless, the convergence with data communication at 10 Gb/s is putting significant downward pressure on cost and will force operators to adopt lower cost equipment. A high level of performance must be maintained while power consumption and size are decreased, and ease of management is increased.

The application (and market size) determines which devices should be integrated. In serial integration, consecutive components of a single channel are integrated together to form a larger part of a photonic integrated circuit. An example would be the Externally Modulated Laser (EML) where a laser and modulator are integrated. Parallel integration occurs when multiple copies of a specific device exist across different channels. Banks of photodetectors or attenuators for different wavelengths in a wavelength division multiplexed (WDM) system would be good examples. The reader is referred to the Chapter by *Janz* on silicon DWDM in this book for more details. Components using parallel integration are usually much larger in order to accommodate the physical spacing between channels. Their cost can therefore be prohibitive unless low cost materials or reduced pitch devices are used. Since no semiconductor can compete with silicon in cost and device size, silicon will have an advantage for parallel integration applications that do not require light emission.

Silicon photonics will have a better chance to gain a foothold into data communications than telecommunications for three reasons; some performance can be sacrificed because the distances are smaller than in telecommunications, the market is extremely cost sensitive, and a large part of the data communication is anchored to servers and desktop computers where silicon in ubiquitous. For the shorter distances, fiber loss is not significant so higher losses can be tolerated at the interfaces between components. The large potential volume of the market and the competition with copper will necessitate inexpensive optical solutions.

Figure 9 shows a schematic of an integrated transceiver that could be implemented for data communications, combining an InP laser with silicon-based photodetection and waveguiding. SiGe photodetectors would be necessary if InP is used due to the low absorption coefficient of silicon at wavelengths above 1100 nm. Transceivers like this could be used for board-to-board and rack-to-rack communications as depicted in Fig. 10, as well as for enterprise applications. Additional channels could be integrated together to increase the aggregate data rate if needed (Fig. 11).

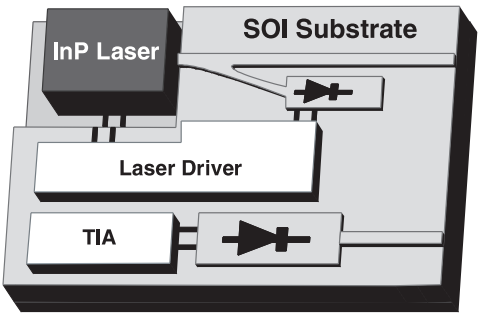

Fig. 9. Example of the integration of silicon and III–V to make a transceiver. Two silicon-based photodetectors are included, one for power monitoring and another for receiving data

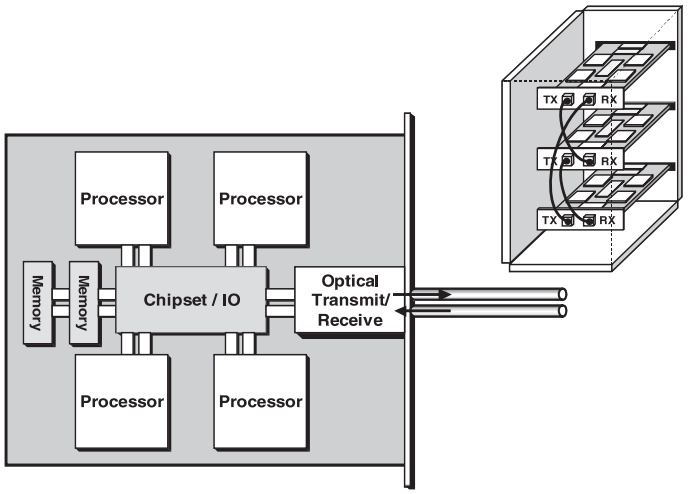

Fig. 10. Example of a potential board-to-board optical interconnet. Aggregated electrical data from several processors is converted to optical data at the edge of the board before being sent to other boards

The environment for these applications is much less harsh than the chip-to-chip case. Expected operating temperatures for the components are less than 60 °C, with available power supplies of greater than 10 V.

3 Integration

There are two primary benefits from integration: increasing performance and decreasing cost. Unlike the case of electronic integration that results in a greater number of identical, higher performance, transistors on a chip,

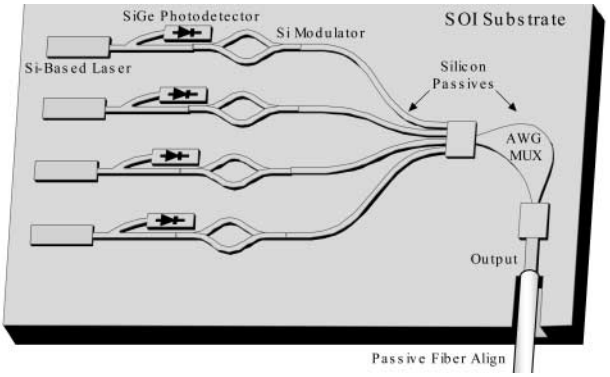

Fig. 11. Example of monolithic integration on silicon

optical integration involves assembling components that either perform different functions, or that are made from dissimilar materials. This has profound implications for the gains in performance and cost. Since the method of integration is critical in determining the quality of the component, it is necessary to first define the different integration approaches before discussing their potential benefits.

3.1 Types of Integration

The most common definition of monolithic integration refers to a process by which all of the components (e.g. electronic circuits, light sources, photodetectors, modulators, waveguides, and multiplexers) are manufactured on the same piece of semiconductor substrate. The example shown in Fig. 11 assumes a Si-based light source exists. A more liberal definition of monolithic integration would allow the inclusion of III–V compounds or polymers (either grown or deposited) processed on a silicon platform. This latter definition is convenient because it allows for different materials to be combined during processing. Such a mixed-material monolithic process provides an opportunity for greater flexibility in the choice of devices to be integrated. This would be an option for integrating a non-silicon-based light source on an otherwise all-silicon chip. To fully realize mixed-material integration, compromises in processing may need to be made which could severely affect both performance and yield. Despite these concerns, researchers continue to work on both single-material and mixed-material monolithically integrated planar optical chips [28, 31, 32, 33, 34].

Hybrid integration is accomplished by assembling disparate parts onto one common platform; the example depicted in Fig. 12 uses SOI as the base material. Common hybrid optical components are those in which III–V compound light sources and detectors are attached onto SOI, silica, or polymer platforms. As can be seen from Fig. 12, the power of hybrid integration lies

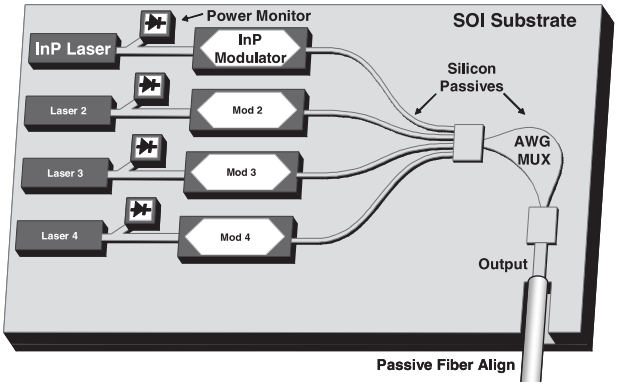

Fig. 12. Example of Hybrid Integration on silicon

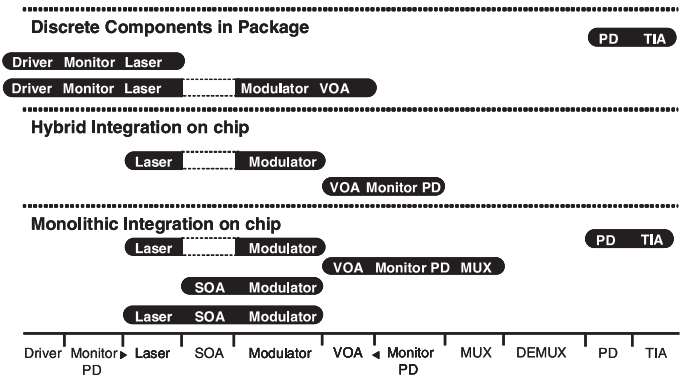

Fig. 13. Level of device integration in commercially available products

in the ability to combine the best performing devices onto one common chip, despite the different technologies used to manufacture each individual device.

Optical components using both types of integration are commercially available. Figure 13 shows which devices have been integrated together on a single chip in comparison to the standard packaging of discrete devices. This latter category is typical for components used in telecommunication applications. The order of devices on the axis follows the order that is found in an optical link. Examples of monolithic integration are almost exclusively made in InP, with the only exception being the integration of a normal incidence photodetector and amplifier in silicon. The integration of optical and electrical devices is the exception rather than the rule however. Combining optical devices is more common because it offers clear performance advantages while maintaining production yields. This will be discussed in depth in the next section.

3.2 Cost

Although integrated optical components had their beginnings over thirty years ago, they have not evolved with the same success, neither in complexity nor in functionality, as integrated electronics [29]. Most of the credit for the technological successes of integrated electronics can be attributed to the advancements in silicon processing. With the opportunity to use these mature manufacturing processes, semiconductor electronics manufacturers are in an exceptional position to drive optoelectronic research, development, and product implementation. An additional benefit of silicon photonics is the opportunity to utilize older technology process tools. Using depreciated factories dramatically reduces the costs of research and development, making investment and production of silicon optical devices more attractive.

Other factors in the cost of integrated optics are the lack of convergence to one material and the absence of device standards, as in the case for volume manufacturing of electronics. With the successful development of additional optical components in silicon, a standardization of market specifications, and focus on a common material platform, silicon-based integrated optics may realize the economy-of-scale benefits that integrated electronics has enjoyed for so long. However, convergence alone will not drive down costs.

In order for silicon to be a viable material for monolithic optical integration, it will need to compete against the high yields associated with hybrid integration. This will require tapping into available processing knowledge and infrastructure to create a stable environment for research and development of new optical components. Monolithic integration will also need to manage production costs, which rise as the complexity and number of processing operations increase. Each process technology is more expensive to implement than the one before it. Modern lithography reticle costs are rising with every technology generation (about \$1 M for a 0.13 μm CMOS reticle set), resulting in a need for high volume production over which to amortize such large initial costs. Fortunately many of today's optical designs can be made using older generation processing equipment.

With the cost of new CMOS manufacturing fabrication facilities in the several billions of dollars, competition for volume silicon processing is becoming more and more cost prohibitive. Though the same number of processing steps are not needed in optical processing, large costs of operation are promoting more 'fabless' start-up companies as the number of optical manufacturing companies dwindles. Figure 14 shows the costs of several optical semiconductor materials (\$/in^2) as well as the costs typically charged by foundries to manufacture some common devices. It can be seen that InP area costs can be almost an order of magnitude larger than either GaAs or SOI. In addition, although GaAs and SOI wafer area costs are almost equal, it is expected that SOI costs will drop as the demand for volume production in the electronics industry increases. Similarly, as the electronics industry matures on 12 in silicon wafers, a reduction in the initial wafer area cost is also expected.

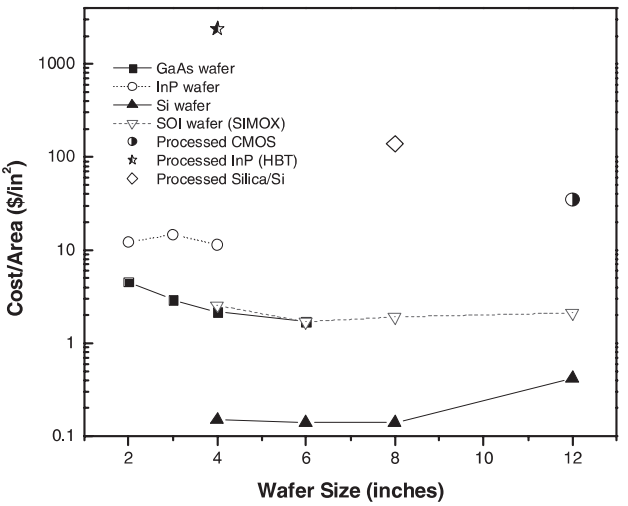

Fig. 14. Start wafer costs and wafer fabrication costs per unit area for several semiconductor materials

Wafer area costs alone do not provide enough insight into the total cost structure of optical devices. Device size and manufacturability are also large factors in determining whether or not there exists an economic incentive to pursue a particular material. As a point of reference, the foundry cost to process a 5-metal layer silicon electronic device (about 25 total mask layers) on a 300 mm wafer using 0.13 μm process technology can be \$5000. The price per square-inch is around \$40, as shown in Fig. 14. The foundry cost of an InP Distributed Feedback (DFB) laser can be over \$2000/in^2 while that of a silica-on-silicon based device is about \$150/in^2. These costs should be considered when designing optical components that require large areas (e.g. waveguides, AWGs).

Although monolithic components in silicon may be simpler to manufacture than those in III–V materials, the potential to add electronics to active silicon-based optical components is an attractive feature to corporations. Unfortunately, the challenges associated with marrying silicon-based planar optical devices and conventional electronics on the same die have not been overcome. Instead, several research groups have combined normal incident silicon photodetectors with receiver circuitry [30]. These efforts have been successful because almost no new processing steps were required to produce the photodetector. The manufacturing cost and yield of the integrated part is similar to the cost of a single discrete part. This is not always the case when two devices are integrated in silicon. If a SiGe process is used for the photodetector, it can increase the number of mask layers and impose thermal constraints required for film stability. In addition to decreasing the total yield, there could be a negative impact on the throughput as thick buffer

layers or slower growth rates are required. To better anticipate some of the potential issues that await silicon-based optoelectronic integration, it may be helpful to briefly review, as a case study, the history of optical integration on InP.

InP is the quintessential optoelectronic material because high performance electrical and optical devices can be made on it. Groups in universities and companies have been doing research on the integration of either all optical devices, or optical and electric devices [31,32]. Examples include the complete integration of WDM optical devices from the lasers through a MUX/DEMUX to photodetectors [33]. Integration of a photodiode and the amplifier has also been shown [34]. However, many companies have given up on commercializing these devices, either due to lack of demand or technical difficulty.

The reason that integration on InP has largely been relegated to research is that it has failed to decrease the cost of production or improve performance. Putting different types of devices on InP usually entails epitaxial regrowth steps that often *significantly* decrease the yield. The number of processing steps such as lithography, etching, and cleaning are also increased, further reducing yield. With these increased number of processing steps come increased wafer handling, which, since InP is such a brittle material, results in higher incidence of wafer breakage. Some devices such as AWGs consume a large fraction of the incredibly valuable substrate surface, even though no device specification is reliant on the materials properties of the III–V material. Added to these cost disincentives is the R & D budget required to make the process functional at all. In order to recoup these costs, there has to be a very large market for the components, and the recent telecommunication collapse has prevented that from happening for all but the EML, which is currently the only commercially successful component.

Since there are so many differences between InP and Si integration, there is reason to hope for a better outcome in silicon. Most of the optical devices that are being considered for silicon are based on either intrinsic silicon or silicon doped with another element, so the need to resort to yield-killing epitaxial regrowth steps can be avoided. If the optical functionality can be added while keeping the process similar to the baseline process for electronic devices, it is reasonable to expect very high yields as compared to yields of 5–15% for InP EML. It is still an open question as to how close the optical fabrication process flow will be to a normal CMOS process. It is also likely that fewer device compromises will be required in silicon than in InP if everything is nominally made from the same material.

Examining only manufacturing costs of devices is not sufficient since the costs of testing and packaging PLCs can be as much as 60–80% of the total cost [35]. Therefore, unless significant improvements are made in these areas, simply reducing substrate manufacturing costs will not provide the necessary economic incentive to pursue highly integrated devices. Although both monolithic and hybrid integrated devices require testing and packaging, the

time and cost is significantly reduced in the monolithic case. Testing and assembly times of hybrid integrated devices increases with the number of parts on the substrate. In contrast, the time to test and assemble monolithic devices can be significantly reduced since there are fewer mechanically aligned interfaces to test, because many of the interfaces are lithographically defined in the substrate. As a result, the costs associated with assembly and testing can be shared by many devices on one chip. Another benefit of monolithic integration is the potential to increase the number of devices with little or no impact to production time. This is not true for hybrid integration since an increase in the number of devices necessitates a proportional increase in the amount of processing time.

3.3 Performance

One of the most contentious issues in integrated optics is determining the impact of integration on the component performance. The fundamental question hinges on the trade-offs that must be made in the integration process, and it is different for the two types of integration described previously.

Monolithic integration on silicon has distinct advantages associated with alignment and interfaces, as compared to hybrid integration. Misalignment between devices in the monolithic circuit is defined by the accuracy of the lithography tool. Integration of components entirely based on a silicon waveguide platform, such as the integration of VOAs and AWGs, are defined in a single step and are not subject to misalignment at all. This minimizes or eliminates the losses between the individual devices. Interfacial losses can also be reduced if the mode is always in a constant refractive index material with constant dimensions. Since losses between discrete components can be 0.1 dB/interface, the savings in an integrated system can quickly accumulate.

The problem with monolithic integration is the poor device performance resulting from constraints placed on material selection. This is the most important for devices that actively manipulate the light, such as emitters, modulators, and detectors. Most of these devices operate near the band edge of the material. Since photodetection only becomes efficient when the wavelength is well above the band-gap, a silicon-based detector with a much smaller band-gap than the silicon-based emitter has to be integrated. Unfortunately, band-gap engineering in silicon-based materials is constrained by problems in lattice matching, so compromises in device performance have to be made. The result is that totally monolithic integration will be best suited for less demanding applications.

Hybrid integration offers more degrees of freedom for the design of each optical component. Each individual component – emitter, detector, etc. – is optimized during fabrication and integrated later. The penalties paid for this are the additional interfaces and misalignment between components. When the interfaces are designed correctly with anti-reflection coatings and mode

matching, the losses can be on the order of 0.1 dB, but this requires an extensive alignment and packaging effort. Issues associated with this process will be discussed later in the chapter.

3.4 Processing

Over the past decade, there have been many reviews [1, 2, 3, 36] on the opportunities associated with optical devices and the potential integration with silicon electronics. In particular, researchers have presented a vision of optical devices monolithically integrated with modern CMOS devices [36, 37]. Though quite realizable, the details of such a vision must be completely compatible with the physical constraints associated with silicon-based processing. Optical devices in silicon have been, to date, made without the monolithic integration of silicon electronic devices. Albeit a promising step towards compatibility, it is an altogether different problem to integrate these research devices with silicon electronics. In this section, we hope to inspire more serious considerations with regard to the feasibility of creating high-volume silicon-compatible optical devices.

Although there are many silicon electronic processes, we will use CMOS processing as a working example for the remainder of this section. Throughout the section, we will try to point out several instances in which devices can utilize existing infrastructure and processing techniques yet remain incompatible with the technology to which they are being integrated, and vice versa. In order to successfully manufacture a monolithic circuit, one must carefully choose the order of each process step to avoid significant physical conflicts. Conflicts such as chemical incompatibility, thermal instabilities, or challenges to lithography and etching processes due to underlying topography, can present stubborn processing roadblocks to successful integration. Realizing CMOS compatible monolithic integration of optical devices will require special attention to processing details. In this section we will discuss several processing challenges.

3.4.1 Thermal Stability

Thermal budget limits are critical when integrating different devices into the same process. In the case of CMOS devices, inserting additional processing steps must be done while observing the original processing limits. Failing to account for thermal budgets could result in the interdiffusion of dopant species, weakening of metal layers, and the introduction of stress due to differing coefficients of thermal expansion (see Table 1).

CMOS processes are carefully tuned to optimize transistor performance. In order to avoid compromising electrical performance, additional steps should either conform to existing thermal budgets or be completed before electrical processing begins. The same holds true for ensuring optical performance is maintained; a Si modulator may operate at slower switching speeds

Table 1. List of refractive index and linear coefficient of thermal expansion for common optical materials at $T = 300\,\mathrm{K}$

Material	n (1550 nm)	CTE ($10^{-6}\,\mathrm{K}^{-1}$)
Si	3.45	2.56
SiO_2	1.5	0.55
Si_3N_4	2.0	2.50
$Si_{1-x}Ge_x$	3.45–4	2.56–5.7
GaAs	3.38	6
InP	3.18	5.5
Polymers	1.6–1.7	10–100

or suffer an increase in optical loss if dopants diffuse from their original implanted regions.

Some of the most common CMOS thermal processing steps are listed below in Table 2.

Table 2. List of common front end thermal steps used in a typical CMOS process

Process Step	Approx. Operating Temp (°C)
TEOS anneal	1000
Gate anneal	1050
Source/drain anneal	1050
Silicide formation	650–1100
Tungsten deposition	500–650
Cobalt RTA	600
Gate oxidation	1050
Activation RTA	950

Table 2 ignores the total thermal budget, that is, the temperature *and* time of heating; we are also not addressing any complications associated with thermal cycling or ramp rates for a particular step. Spike anneals and fast thermal cycles are required to reduce transient enhanced diffusion and dopant deactivation. The effects of transient diffusion are the main reasons why the electronics industry moved from slow furnace batch annealing to single-wafer Rapid Thermal Annealing (RTA) techniques such as spike and flash anneals. One difficulty for optical integration would be the incorporation of an optical process step that requires growing a thick SiO_2, layer. For such a large thermal budget, usually many hours at temperatures above 900 °C, this step would need to be done prior to any backend (metal) or critical ion implantation steps. Thermal budgets for both aluminum and copper back-end processes with oxide-based dielectrics are limited by the Interlayer Dielectric (ILD). Although some low-k dielectrics ($k < 3.9$) are being used (e.g. Carbon-Doped Oxide (CDO)), the thermal stresses and the weaker mechanical properties of CDO will not be able to tolerate temperatures in excess of 450 °C; low-

k solutions such as Spin-on-Dielectrics will drive the thermal budgets even lower. In the case of electro-optical (EO) polymers, the thermal budget can decrease dramatically since typical glass transition temperatures are in the range of 150 °C–250 °C. For all practical purposes, these polymers can only be integrated as the last steps of the process.

In addition to the thermal issues of decomposition and softening mentioned above, the problem of strain relaxation must be considered for processes that use heteroepitaxy. While this could be the case for III–V growth on silicon, it is far more likely to be seen if germanium is introduced to create photodetector production. Germanium concentrations in excess of 50% are desired to achieve acceptable responsivity from the SiGe photodetector at 1310 nm and 1550 nm. At these concentrations, it is very easy to exceed the thermodynamic critical thickness [38] and grow a thermally metastable film. If latter processing is sufficiently hot to allow the formation of misfit dislocations to relax the film, the optical properties of the device can be degraded. This could restrict the processing of thermal oxides and RTA steps.

Another factor in the successful processing of monolithic optical devices is the choice of materials. Silica-on-silicon technology typically requires that thick layers (several microns) of SiO_2 be processed in contact with crystalline silicon. After cooling from the high-temperature (approximately 1000 °C) oxidation step, a great deal of strain is created between the glass and crystalline layers. Interlayer strain can cause undesirable effects including strain in the optical layer, de-lamination of the oxide layer from the silicon substrate, or cracking of the glass layer to name just a few.

The thermal budgets of hybrid integration mainly revolve around the alignment stability of the packaged part and the degree of thermal isolation between devices. Each of the optical components may be secured in place with a different method (e.g. gold eutectic bonding, laser welding, epoxy, etc.), which will impose different thermal constraints on the operating temperatures of the device. Thermal crosstalk can be damaging to the alignment between two optical devices. Several options exist to ensure protection of the hybrid component from thermally-induced misalignment. Some options include keeping operating and environment temperatures at the lowest value common to all sub-components, ensuring thermal isolation between the individual components, or actively controlling individual component temperature to compensate for thermal crosstalk.

3.4.2 Chemical Compatibility

During integration, chemical compatibility between different processes is a cumbersome problem to deal with. Most conflicts can be addressed by incorporating extra de-contamination steps or adding dedicated tools for the problematic operations. As one would expect, these solutions can be extremely expensive. Depressed yields due to chemical contamination have been brought under control in standard silicon processing, but would need to be addressed

if the process demanded the inclusion of elements and compounds not used in standard processing. Integration problems related to chemical contamination are likely to pose practical as opposed to physical limitations.

The term 'contamination' can encompass problems related to incompatible chemical processes in addition to particle defects. Contamination of the wafer surface or of thin-film materials can lead to adhesion and delamination issues. Certain techniques commonly used in university research processing are avoided in silicon fabrication due to their high likelihood of contaminating the production line. An example of this is the case of fast anisotropic wet-etching of silicon with KOH to form V-grooves for fiber alignment. In a commercial environment the risk of alkali contamination in silicon is too high. As a result, defining v-grooves must be done after all electronics processing is complete. Similarly, gold metallization (typical for III–V devices) is avoided in the silicon process until the final step, to prevent diffusion of metal into silicon. Even after its introduction over three years ago, the semiconductor industry is still trying to address issues associated with copper (Cu) contamination. The problem of Cu contamination extends beyond the factory and to the vendors and suppliers who perform cleaning services and other handling steps. Copper protocols can dramatically impact cost by requiring a special cleanroom layout and air flow as well as a duplication of all process equipment, tools used to service semiconductor equipment, wafer carriers, and even the gowns worn by the fab personnel. There are still no universally acceptable levels of Cu contamination, and many questions still remain about how far beyond the fabrication tools and chambers one must de-contaminate. Similar logistical problems may be waiting for the monolithic integration of optical devices.

3.4.3 Lithography

Minimum dimensions of photonic devices are limited by the wavelength of light in the medium which for which they are designed; ($\lambda = \lambda_o/n$), where λ_o is the free-space wavelength and n is the index of refraction of the medium. In the case of silicon ($n_{\mathrm{Si}} = 3.45$), devices designed for a free-space wavelength of 1550 nm have minimum widths of approximately 450 nm for an operating wavelength of 1310 nm, widths are approximately 350 nm. In optical systems, unlike electronic circuits, bends in transmission lines with small radii of curvature can cause unacceptably high radiation losses and transition losses. The large refractive index contrast of SOI ($\Delta n = 2$) provides better confinement than lower index contrast systems, allowing for relatively compact designs. Silica-on-silicon based Arrayed Waveguide Gratings (AWGs), with an index contrast of $\Delta n < 0.01$, have typical design sizes of approximately 12.5 cm^2 for a 1×4 channel multiplexer/demultiplexer. The same device can be designed in SOI at 0.5 cm^2, providing a reduction in device size of 25×.

The ability to faithfully reproduce designed features on a wafer is important to the success of optical devices. Critical Dimension (CD) control and

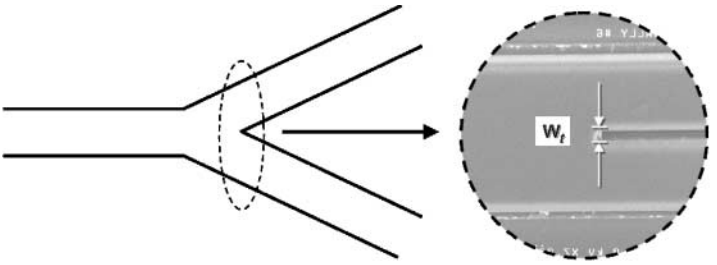

Fig. 15. Picture of a Y-splitter in SOI. Scattering occurs primarily at the junction tip labeled w_t

smooth device surfaces are two of the most important parameters that must be controlled during processing. Producing CDs of approximately 350 nm is hardly a challenge for modern optical lithography processes, whose minimum features have already been demonstrated below 60 nm. This does not, however, reflect the minimum CD in the design. In devices such as Y-splitters, AWGs, etc., the reduction of the splitter region to a point will always benefit from state-of-the-art patterning technology. In the case of a Y-splitter (Fig. 15), rounding and increased tip size in the splitter region increases the scattering loss.

The ideal situation would be one that reduces the tip width (w_t) to zero, resulting in zero scattered light. Since such a design cannot be patterned reproducibly, the minimum size of the splitting region is limited to the minimum resolvable feature size available by photolithography. Thus, the better the lithography and etching capability, the lower the scattering losses associated with the splitting region.

Another example of the benefits high-resolution patterning brings to planar optical circuits can be demonstrated by a Bragg grating. These gratings can be used as wavelength selective filters or laser cavity mirrors. To define Bragg gratings for a specific wavelength λ_b, the relation $\lambda_b = 2n_{\text{eff}}\Lambda$ must be satisfied, where Λ is the grating pitch and n_{eff} is the effective index. To create a first-order Bragg grating in silicon for $\lambda_b = 1550$ nm and $n_{\text{eff}} = 3.45$, a pitch of 225 nm is required. Thus, for a line/space duty cycle of 50%, lithographic patterning must be capable of 110 nm lines or spaces. Such features are common for the 90 nm lithography node. Although it is possible to create higher-order Bragg gratings with lower resolution lithography, the resulting loss is greater.

State-of-the-art lithography resolution may not be required for most conventional waveguide designs, but ring resonators and photonic crystals might require CD values on the order of 100 nm. Resolution and CD control are both important to the success of the devices. In such cases, CD control that accompanies modern patterning techniques (about 1 nm over an 8 in wafer) can be a benefit. In the case of Si-based ring resonators, one of the critical parameters to control is the gap (about 100 nm) between the ring and bus. Since the

device operates through evanescent coupling, the coupling is exponentially dependent on the size of the separating gap. Thus, in order to reliably process high-Q devices, control of a few nm demands CD control readily achieved by modern 130 nm or 90 nm optical lithographic techniques [39, 40, 41]. This level of CD control would also be required for other small coupler designs.

More recently, SOI-based photonic crystals have been receiving considerable attention. Though we leave a more detailed discussion for another chapter, we would like to point out the most dramatic benefits of a high-index contrast system like SOI for novel physical systems such as photonic crystals. Among the many advantages to be discussed later, two are of special note: compatibility with modern lithographic resolution and reduced device form factor. For structures designed to operate at wavelengths of 1550 nm and 1310 nm, photonic crystal lattice structures (holes) are of the same dimensions/design rules as those for 130 nm and 90 nm CMOS transistor designs. This is of particular interest since, to date, almost all research efforts, especially photonic crystals, have been restricted to the use of e-beam lithography for pattern definition.

Although small CDs for optical designs (about 100 nm) can be patterned using modern 130 nm lithography tools, it is not a viable research process. Furthermore, since optical lithography minimum resolution is still many years away from that of today's standard research e-beam writers, e-beam tools are preferred for research and small commercial manufacturers. A few of the greatest advantages of e-beam lithography are the relatively low cost, the ability to pattern < 20 nm features, and their relative ease of use for processing pieces of wafers. In contrast, today's 130 nm optical lithography scanners have a resolution of approximately 100 nm, can process only whole wafers, and are prohibitively expensive. Owing to factors such as long patterning times, relatively large field-stitching errors, and susceptibility to particle contamination, e-beam lithography is a long way from being a competitive commercial alternative to conventional optical microlithography. A state-of-the-art research e-beam system can have patterning times of 30 minutes or more for a 10 μm^2 field with a 50% pattern density of 100 nm minimum feature sizes. Should minimum resolution be desired (20 nm), the patterning time can increase. With exposure times of less than 1 s for a standard scan field, optical lithography offers an incredible advantage in processing time if an optical design requires a larger area than that provided by one optical lithography scan field (approximately 25 mm $\times$ 35 mm), the lithographic stepping error due to stitching two fields together is < 25 nm 3σ. If however the design requires an area less than or equal to one scanning field, modern optical lithography is a better tool since the stitching error of the e-beam field (60 nm 3σ) can result in additional optical losses due to the many discontinuous sections that compose the device. Overall, standard 130 nm optical lithography can deliver equal or better results as compared to most research e-beam systems for today's planar optical designs.

High resolution lithography requires flat surfaces on which to pattern features. Height variations in wafer topography must be less than the lithography tool's usable depth of focus (DOF). Since typical waveguide heights exceed the DOF, process steps must include extra layers that can be mechanically polished. Even the shortest SOI single-mode waveguides must be covered in a planarized layer (e.g. SiO_2, SiON) before the next layer is patterned.

3.4.4 Etching

In addition to lithography is the subsequent etching of the patterned area. There are two common methods used in semiconductor processing: wet chemical etching and reactive ion etching (RIE) or 'dry etching'. Wet etching is a purely chemical etch and can be isotropic or anisotropic, depending on the chemicals used and the materials being etched. In some cases, researchers have exploited the fact that certain crystal planes etch faster than others to achieve anisotropic etch rates, but the anisotropy and precise CD control have not matched those found in dry etching. RIE is a technique that uses a plasma consisting of ions, free electrons, free radicals, and neutral molecules. By controlling the electrical potential of the wafer, positive ions can be guided to the wafer surface in a directional manner. This combination of chemical etching (free radicals reacting with the wafer) and physical etching (ion bombardment) provides a highly anisotropic etch; the vertical etch rate is much greater than the horizontal etch rate. Most optical and electrical designs demand vertical sidewalls, and therefore require the use of highly anisotropic etch techniques. In addition, unlike purely wet chemical etching, RIE provides excellent critical dimension control and the ability to etch high aspect-ratio patterns.

Most critical etching is accomplished by controlling the time of the process. Until recently, researchers could not rely on RIE control for accurate patterning of critical dimensions. In the case of ring resonators, researchers have used thin film deposition to manage the separation distance of resonantly coupled features (the space between the bus and the ring). Such a technique requires additional mask layers to accomplish what can now be done with standard lithography and RIE.

Since RIE relies in part on physically etching the wafer surface, the etch selectivity (selectivity = etch rate of material to be etched/etch rate of patterned protective mask layer) must be relatively high in order to achieve high aspect ratio etching. If the selectivity is low, the masking layer (e.g. photoresist) thickness is increased. Unfortunately, such a situation cannot be supported by modern lithography since a consequence of reduced DOF is the need for a reduction in photoresist thickness. The DOF puts a limit on the photoresist thickness, which is rapidly approaching $< 100\,\text{nm}$. This is not an issue for CMOS processing since typical etch depths are on the order of 0.1–0.3 μm for critical layers requiring minimum feature sizes. Typical etch depths for optical designs are between 0.1–6 μm.

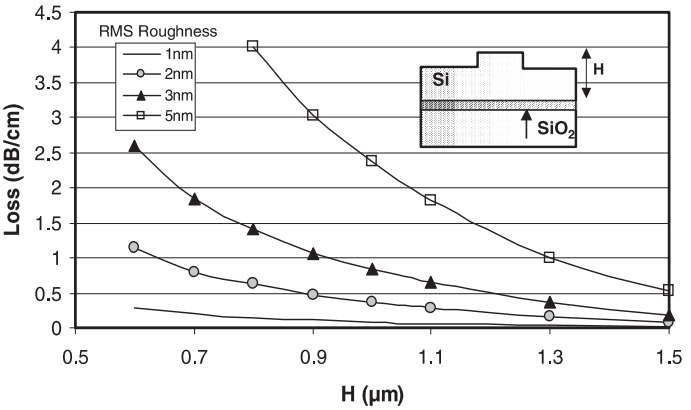

Fig. 16. Modeled scattering loss for SOI rib waveguides. The dimensions of the etched rib height were chosen to satisfy the single-mode condition for a particular total WG height (H)

RIE processing results in roughened sidewall surfaces, which, if significant, can result in additional optical loss. Contributions to waveguide propagation loss are scattering and absorption. If optical transmission losses in waveguide structures are reduced to negligible levels, the scattering associated with sidewall/interface roughness plays the dominant role in the overall loss of undoped crystalline materials. Unlike the scattering associated with defects in the waveguide medium, interface scattering is a direct result of lithography and etching performance. *Tien* derived an approximation to the loss related to interfacial scattering [42, 43, 44]. Today's conventional 180 nm lithography/etch processes can reliably produce sidewall roughness figures on the order of several nm rms. Unfortunately, since the scattering losses vary rapidly with the rms interface roughness as waveguide dimensions are reduced, several nm may not suffice for certain applications. An example of this dependence can be seen in Fig. 16, which shows modeled results of the scattering loss of SOI rib waveguides. The modeling was performed by using a commercial software package called BeamPROP [45], which incorporates advanced finite-difference beam propagation techniques for the simulation. The plot demonstrates the effect of increased roughness and decreased waveguide dimensions on optical loss. In addition to optimizing etch parameters to reduce silicon sidewall roughness, post-etch oxidation of the silicon surface has been used to reduce the roughness and hence the loss [46, 56] relying on the fact that convex silicon features oxidize faster than concave areas near the surface.

3.5 Yield

Today's electronic devices are highly complex and continually shrinking, requiring fabrication methodology to include parametric testing for process control. Electronic devices are inherently easy to probe. Even before they are completely processed, they can be accessed across an entire wafer. It is for this reason that today's typical CMOS fabrication testing plan can include thousands of parameters to be monitored in the hope of identifying potential threats to device performance and reliability.

Over several decades of CMOS development, the automated testing of devices before completion has been instrumental in reducing per-wafer manufacturing costs. This so-called 'in-line' testing provides performance and defect density information essential to predicting and rapidly correcting for process errors and depressed yields. Thus, metrology collected after only a fraction of the total processing is complete can provide excellent insight into the final performance of the device under test.

Optical devices, however, do not readily lend themselves to convenient in-line testing. There are at present no convenient, fast, automated, and non-destructive probing techniques that can be well correlated to final, packaged optical device performance. Much like the electronic test cells used for collecting parametric data, optical device test structures can be designed and dedicated to the collection of relevant inline information. Some efforts have been made to use top-down probing of optical devices but none of these techniques have been introduced into the manufacturing process.

One such technique for top-down optical probing of a waveguide is through the use of a prism [42,47]. In Fig. 17a are depicted two prism couplers, held above the surface of the waveguide at a distance of approximately $\lambda_o/4$. After light enters the prism it is totally reflected at the base. This reflection creates a standing wave in the prism whose field extends below the base and into the waveguide below (evanescent tail). The evanescent tail excites a light wave in the waveguide; coupling light to the waveguide through the air gap is known as optical tunneling. Prisms can be used as both input and output couplers and, since they are not a physical part of the device under test, the distance between them can be adjusted to collect loss measurements as well as transmission data. In addition, it is possible to test the product directly without the need for dedicated test structures that monopolize expensive field area. However, positioning the prisms within one-quarter of a wavelength to the waveguide surface can easily lead to surface damage. This, coupled with the need for prism material with a higher index than the waveguide material, makes prism couplers a difficult practical solution.

Another successful technique is to couple the incident light perpendicular or nearly perpendicular to the plane of the wafer via a grating [48, 49, 50] (Fig. 17b). By appropriately designing the grating period, efficient coupling can be achieved. Modeled efficiencies up to 95% have been reported using this technique, with measured efficiencies of 70% for rectangular gratings [51] and

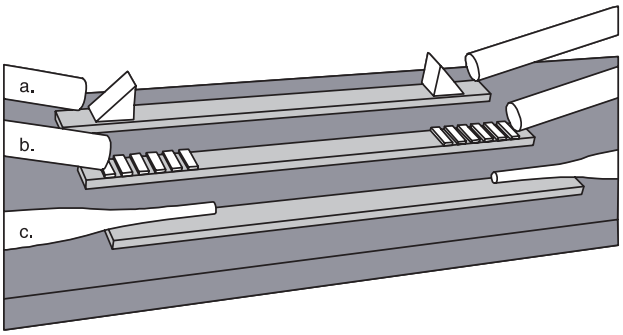

Fig. 17. Waveguide coupling via (**a**) prisms (**b**) gratings, and (**c**) evanescent tapers

84% for blazed gratings [52]. Since a grating coupler must be monolithically integrated into the waveguide under test, a special test structure would need to be created specifically for in-line optical testing. Fortunately, most planar waveguides are only a few microns wide, allowing for test structures to be placed in noncritical scribe area, similar to metal test structures in electrical test chips. The practical difficulty of implementing this technique will be in accurately positioning the waveguide probes since the grating coupler efficiency is highly sensitive to the angle of incidence.

Yet another in-situ test technique relies on evanescent coupling between a tapered optical fiber and a planar waveguide. As the optical fiber is brought into close proximity with the waveguide (Fig. 17c), resonant coupling of the evanescent field transfers the optical power into the waveguide. Modeled coupling efficiencies of approximately 98% have been reported for photonic crystal waveguides [53]. Like the coupling via a prism, this technique can also be used to couple directly to waveguides without the need for dedicated test structures.

In addition to increasing complexity in parametric characterization, reductions in the size of electronic devices demands tight processing control over unwanted defects. Some examples of defects are metal shorts, particle contamination, bumps in thin films (low-k deposition), etch residues left after etch cleaning, post-contact etch cracking due to Inter-Layer Dielectric (ILD) stress, and adhesion loss causing blistering of bond pads, to name a few. These defects can be monitored during processing, and appropriate measures can be taken to correct for them. However, the nature and effect of these defects are understood and relatively easy to detect. Most problems occur at the surface of the wafer after a given process step, allowing for surface inspection (e.g. optical microscope, laser, or pattern recognition/comparison).

Defect detection and characterization are not mature sciences for the production of optical devices. For a defect to impact the operation of an optical device, it need not be detectable from the surface of the wafer. One of the most common material issues in photonic devices is a variation in the re-

fractive index of the guiding/cladding layers. These variations can be due to stress, defects, or nonuniformities in thickness and composition. Effective in-line techniques to inspect optical devices for these variations have yet to be developed.

3.6 Coupling and Packaging

Coupling light into and out of a silicon chip is very challenging. Particularly difficult is the coupling of light from a standard optical fiber to a silicon waveguide due to the large modal mismatch. Furthermore, lacking efficient silicon-based emitters, external light sources must be directly coupled into silicon devices. These challenges require the development of additional processes and structures beyond those of the core device. This section will survey some of the commonly used practices for the assembly of commercial products.

3.6.1 Tapers

Coupling light into and out of a silicon chip is very challenging. Particularly difficult is the coupling of light from a standard optical fiber to a silicon waveguide due to the large modal mismatch. A single mode fiber core ($n = 1.5$) usually has a diameter of 8 μm with a symmetric mode while a silicon waveguide ($n = 3.45$) is typically only a few microns in width with an asymmetric mode. One way to overcome these large differences in effective index, core size, and symmetry is with a waveguide taper. This allows for a reduction in coupling loss through a modal transformation. A taper can also be used to increase the alignment tolerance to other optical devices, such as III–V lasers.

Several methods have been proposed and have demonstrated efficient coupling from a relatively large silicon waveguide into an optical fiber. These include pseudo-vertical tapering and using gray-scale lithography for more gradual horizontal and vertical modal tapering [2]. Figure 18 is a schematic depicting these two mode transform designs. Results using a pseudo-vertical taper (Fig. 18a), tapering from a 12 μm by 12 μm input down to waveguides on order of 4–5 μm have given coupling losses as low as 0.5 dB/facet [54].

There are a few key processing parameters that must be controlled in order for successful processing of either of these mode-transfer devices. Specifically these are the lithography and etching processes.

For pseudo-vertical tapering, a horizontally tapered waveguide is patterned on top of another waveguide and the optical mode is gradually squeezed from the top taper to the smaller, lower waveguide. The key parameters for this transition are the length of the taper (the longer the length the more slowly one can transform the mode resulting in lower loss) and the taper tip width (see Fig. 18a). In order to reduce optical losses associated with the finite size of the tip width, the tip should be designed such that the

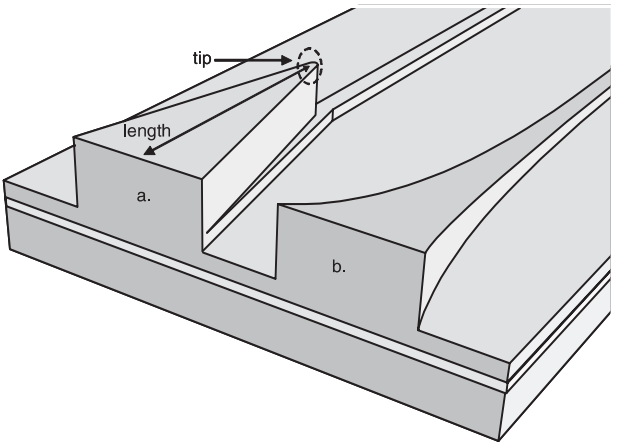

Fig. 18. Two examples of 3-D modal transformers used for coupling light from a fiber into a silicon waveguide: (**a**) pseudo-vertical taper, (**b**) grayscale taper

minimum width is substantially smaller than the wavelength of light transmitted in the waveguide. Optimally, the tip width should be so small that there is almost no optical mode in the tip region. For wavelengths of interest around $\lambda_o \approx 1550\,\text{nm}$ this corresponds to a wavelength in silicon of around $\lambda \approx 450\,\text{nm}$. This final tip dependency is very critical and the effect of tip width variation on taper loss is shown in Fig. 19. The plot shows results of simulation of a pseudo-vertical taper design starting from 10 µm by 10 µm and tapering down to 2.5 µm by 2.3 µm. The modeling was performed by using a commercial software package called BeamPROP [45] which incorporates advanced finite-difference beam propagation techniques for the simulation. In the simulation, the loss due to sidewall roughness was assumed to be zero. The taper length was 1 mm and free space wavelength of 1550 nm was used. One can see from Fig. 19 that for a fixed sidewall angle of 80° the change in the tip point from 2µm down to 0.5µm improves the total taper loss by more than 20 dB. The higher loss for wider tip point is due to the fact that there is a more significant proportion of optical field existing in the tip region that can not propagate along the final waveguide so that it radiates out of the waveguide.

Unfortunately, to fabricatE such a taper, having the lithographic patterning and photo resist capability alone is not enough. In order to complete the process, etching these taper devices invariably leads to some rounding of all sharp edges, thus never allowing one to really achieve the desired sharp point.

In the case of grayscale techniques, appropriate resists and etch recipes must be created and controlled to provide smooth surfaces to slowly vary the height of the taper from the input down to the final waveguide dimension. The technique of grayscale lithography has been known for several years, and

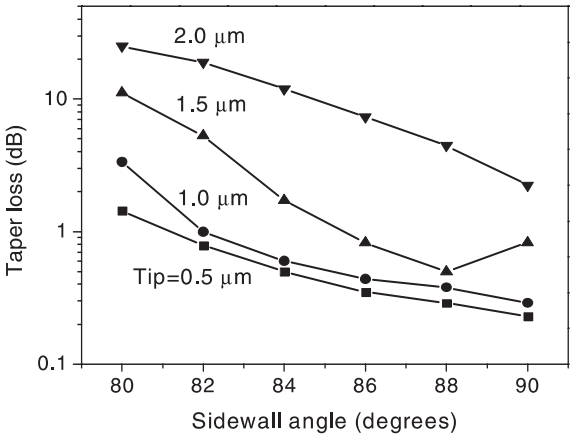

Fig. 19. Plot showing dependency of coupling loss for a silicon taper on both final tip dimension and sidewall angle. The sidewall roughness contribution to loss was assumed to be zero

is widely used in optical Micro Electrical Mechanical Systems (MEMS) to create lenses, prisms, and various other structures [55]. However, grayscale based lithography adds some additional complexity and cost. A grayscale reticle is typically five times more expensive than a standard lithography reticle and process refinement to develop and pattern the resist is not trivial. Although fundamentally tapers based on gray scale should result in much lower loss, these types of tapers have not been as prominent as top tapers.

Another key parameter for processing tapers is that of sidewall roughness and sidewall angle as a result of the silicon etch. Creating taper devices usually relies on significant etching away of silicon material in order to produce the final taper design. Although one can use wet-chemistry for etching, it is often difficult to control and it is not practical, especially if one would like to obtain vertical sidewalls or more exotic profiles, such as a parabolic design. Thus the most common approach is RIE.

RIE also has its drawbacks as it produces surface roughness on the exposed sidewall. The effect of sidewall roughness as a function of waveguide height was discussed previously (Fig. 16). As the optical loss due to the surface roughness is strongly dependent on the waveguide size, the loss due to roughness should be minimal at the beginning of the taper since the taper dimension is large. At the taper end, however, the waveguide becomes smaller and the loss contribution from the roughness becomes larger. Therefore, it is critical to control the surface roughness at the smaller end of the taper. Techniques to reduce loss have been proposed and demonstrated, which involve smoothing waveguides by use of repeated oxidation and stripping steps [46, 56]. However, oxide smoothing also results in rounding of sharp points which could add additional loss due to tip width rounding.

Another approach to reduce the effect due to surface roughness would be to pattern the taper into a thick oxide layer and then grow the taper using an epitaxial growth step. This may result in lower loss due to the reduction of the silicon etch step but adds complexity into the process which may jeopardize the entire process flow as discussed previously.

In addition to surface roughness one must monitor the resulting sidewall angle of the pseudo-vertical taper after silicon etching is complete. Figure 19 shows the effect of sidewall angle on the taper loss. It can been seen that a sidewall angle variation from 90° to 80° can increase the loss by up to 20 dB for a given tip width.

Top tapers have been successful for coupling to waveguides with cross sectional dimensions of 4–5 μm. Modeling has shown that these designs will not work well when the waveguides are smaller than 2 μm. Assuming no loss due to roughness and perfectly vertical sidewalls, modeling [45] shows that tapering from a 13 μm by 13 μm input to a 1 μm by 1 μm waveguide still gives a loss of 2.3 dB/facet. This is true even with a tip width as small as 0.05 μm and a tapering length of 5.2 mm. Thus a different approach must be used for coupling to very small waveguides (on the order of 1 μm or less) to standard optical fibers. Alternative approaches have recently been proposed for coupling to sub-micron waveguides, including the use of grating assisted couplers [1, 2].

3.6.2 Laser Die Alignment

Barring a breakthrough in the power output of silicon light emitters, III–V lasers will be required as on-chip light sources. These laser die are attached either by an active or passive alignment procedure. Active alignment refers to a process by which the laser die is turned on during placement, and the die position is optimized through real-time feedback. The required electrical connections for this process make the alignment difficult. An addition complication is the size of the sub-mount for these electrical connections. The sub-mount must be large enough to be probed electrically and yet small enough to be easily aligned to the waveguide and bonded to the substrate [57].

In contrast, passive alignment of a light emitting die is performed without electrically connecting or probing the III–V device. This removes the requirement for a sub-mount, so the die placement can be performed using a vacuum tip chuck to lift and position the III–V die onto the silicon substrate. In order for the final integrated module to be cost effective, the passive approach should drive as much cost out of the die placement process as possible. The silicon substrate could be used to increase placement accuracy by using mechanical etch stops. Also, the need to place the III–V die with high accuracy could be reduced by adding tapers or other loss reducing devices, to the silicon waveguide (Fig. 20).

The mechanical challenges of the die attachment process center around the issue of alignment accuracy. Depending upon the relative size and shape of the

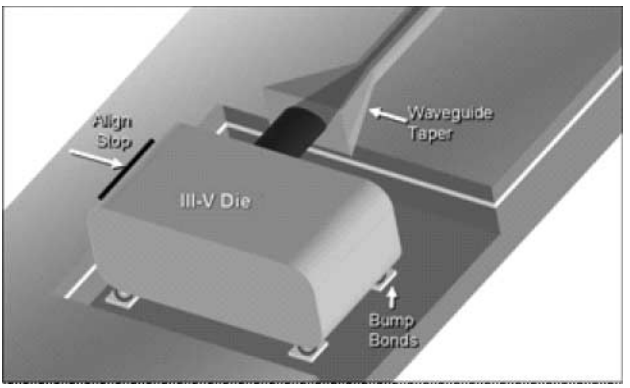

Fig. 20. Simple schematic showing a III–V die placed onto a silicon on insulator platform. The diagram depicts key features that could be used to enable passive alignment

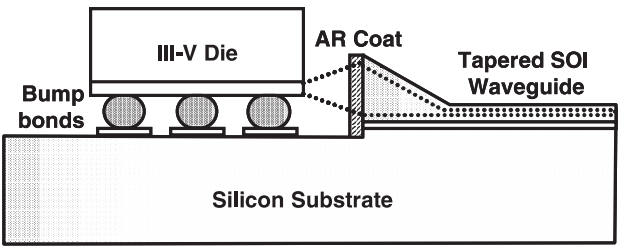

Fig. 21. Simple schematic showing a cross-sectional view of a III–V die placed active side down onto a silicon platform and aligned to silicon waveguide. Shown in figure are key process related activities that needed to be developed

waveguide mode, the mode of the III–V die, and the tolerable system loss, the required placement accuracy could be tighter than 1 μm, with angular tolerances of several degrees. This required placement accuracy imposes tight requirements on the die attachment technique, especially under both broad temperature and long lifetime requirements. In addition very accurate control of the etch depth is required to ensure that the vertical alignment can also meet the comparable alignment tolerance.

Many different types of die attach and alignment methods are possible for a passive alignment scheme. High precision die placement is commonly accomplished using accurate pick-and-place machines outfitted with vision recognition systems [58, 59]. Self-aligned solder reflow techniques with and without mechanical stops to control die placement have been published [60, 61]. Commercial laser die typically have x-y dimensional tolerances on the order of ±5 μm due to the limitations of the cleaving method used to separate individual die. In order to overcome these errors, alignment techniques of-

ten use pillars on the die surface. The position and size of these pillars are formed through photolithography and etching, so the distance between the edge of the pillars and the center of the laser is strictly controlled. High quality pick-and-place tools equipped with machine vision align lithographically defined marks on the laser die to alignment marks on the silicon. However the greater the required alignment accuracy the longer the cycle time needed for die placement and the more expensive the pick-and-place machine. Typical die, pick and place machines with accuracies of around 3 μm are five times lower in cost than a similar die placement machine that has < 1 μm accuracy. In order to reduce the overall cost of assembly both the tool cost and cycle time must be accounted for. A system designed with large tolerances for the manufacturing process (> 1 μm) will drive down the cost of the tool and decrease the cycle time resulting in net lower cost of assembly.

Typical die attachment techniques involve using gold as the bonding metal with either thermo compression or soldering as the die attachment mechanism [62, 63]. As mentioned previously, gold is well known in the electronic industry as an electronic trap and thus deposition and patterning of gold cannot be performed inside a CMOS fabrication facility. As a result all gold processing must be done at the back end of device fabrication and isolated from the remainder of the CMOS fabrication facility. This could present significant challenges with the die attachment and integration process. Increasing the challenge, the metal (see Fig. 21), must be patterned at the bottom of the trench into which the III–V die is placed. The trench is typically 5–15 μm in depth and thus control over the size, height and potentially the volume uniformity of the metal pattern will be affected by how well the lithography process at the bottom of such a trench can be patterned. Additionally solders are often deposited by an electroplating process in order to achieve the correct stoichiometry and the control over the deposited thickness can be challenging to control to better than 5%.

Low loss passive alignment requires that the mode between the light emitting die and the SOI waveguide be well matched. The similar index between silicon and most III–V materials improves the situation as compared to mode matching a III–V light emitting die to say for example a silica (SiO_2) waveguide ($n \approx 1.5$). However mode matching still requires careful design. Microlenses have commonly been used to shape and focus the laser output mode before coupling into an optical fiber, but are not low cost as compared to integrated solutions.

The mode of a typical edge emitting laser diode is elliptical, horizontal in the near field and vertical in the far field and this mode diverges rapidly (within a few wavelengths of the edge of the laser facet). On the silicon side a typical SOI single mode rib waveguide mode is also asymmetric. The amount of this asymmetry is dependant on the actual size and dimensions of the silicon waveguide [2]. In order to achieve better mode overlap between the laser die and the SOI rib waveguide one or both of the optical modes must

be transformed. This can be achieved by either integrating a mode converter into the light emitting die or into the SOI waveguide. Mode shaping at either the III–V dies or waveguide input has the additional benefit of allowing the mode to be expanded, decreasing the required die placement accuracy. Mode transfer devices have not yet been integrated into commercial III–V laser die, requiring tapers to be made on the silicon waveguide.

Another challenge for successful die placement is the formation of high quality etched waveguide facet (typically rms roughness of 10 nm or less is required) in order to reduce the scattering loss at the air-waveguide interface. Due to the large index difference between air and silicon ($\Delta n \approx 2.5$) minimization of back reflections from the waveguide facet requires the deposition of an antireflection coating (ARC) on the etched facet. This ARC needs to be processed at the wafer level and could be combined with an angled facet in either the waveguide or light emitting die to produce net back reflection of $< 10^{-3}$. Integration of an ARC with the process of gold bump patterning while maintaining smooth waveguide facets can be accomplished with significant process development.

3.6.3 Fiber Attachment

The final steps of a processed optical chip include testing and packaging. These steps often demand the alignment and attachment of fibers, which can add significantly to the final cost of the device. Active alignment is inherently more precise due to the real-time feedback in positioning, but can be significantly more expensive due to the time required for alignment and the specialized equipment. A lower cost, passive alignment approach is possible in silicon through the use of conventional lithography. The drawback to this technique is a potential reduction in coupling efficiency due to variations in the in the fiber geometry and in the silicon processing. A photograph of passive alignment using a U-groove is shown in Fig. 22. In this approach, trenches are etched into the SOI substrates which are used to hold and self-align the fiber to the silicon waveguide.

The alignment accuracy of a passively aligned fiber in a U-groove depends not only on the etching tolerances of the groove, but also on the manufacturing tolerances of the fiber. Standard single mode fiber cladding specifications are typically $125\,\mu\text{m} \pm 1\,\mu\text{m}$. Newer fibers with tighter specifications are now available, but come at a higher price. The lateral alignment tolerance of an $8\,\mu\text{m}$ single-mode fiber core to a $13\,\mu\text{m} \times 13\,\mu\text{m}$ square planar waveguide input facet was modeled using FIMMPROP [64]. The modeled results, based on film mode matching and local mode expansion methods, show that the lateral alignment tolerance is $\pm 2\,\mu\text{m}$ for 1 dB excess loss. By increasing the input taper to a sufficiently larger size (e.g. $15\,\mu\text{m}$ by $15\,\mu\text{m}$), it could be possible to relax the alignment tolerance somewhat and help overcome the manufacturing variations of the fiber geometry. Producing such large tapers is extremely challenging. Due to the large index difference between silicon

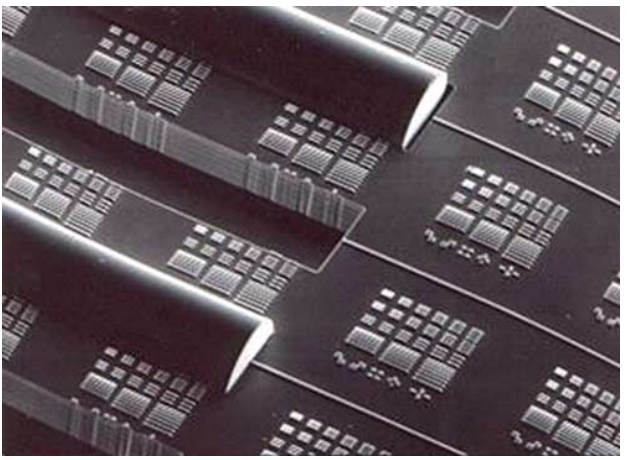

Fig. 22. SEM photograph of passive fiber alignment using a U-groove

and glass fiber, an ARC will also need to be deposited onto the etched facet of the waveguide.

4 Summary

Microelectronics manufacturers, through their years of processing experience, are uniquely positioned to advance integrated silicon photonics. With increased investment in silicon photonics, manufacturers will begin to overcome some of the issues associated with optical and electrical process integration. The fabrication requirements for most optical devices do not currently require state of the art processing capability. By using depreciated infrastructure, companies should anticipate developing low-cost devices for applications that demand high-volume. Unfortunately, cost-effective processing alone will not guarantee success; coupling and packaging will remain as two of the most stubborn obstacles to achieving low-cost optical solutions.

Several things must happen before integrated silicon photonics becomes commercially sustainable. First and foremost, there must be an application for integrated photonics where either an electrical or discrete optical solution will not work. Electrical signaling of 20 Gb/s over meter distances will occur in the near future, so photonics will not be used for chip-to-chip communication until the next decade. Longer distance applications are suitable candidates for silicon photonics over the next 10 years, but there must be a large enough market to justify the cost of integration. Standardization of component designs will go a long way to create demand. The development of a silicon-based light emitter, modulated directly or indirectly at data rates > 100 Mb/s, would also greatly expand the areas for which silicon photonics could have large impact. The monolithic integration of other silicon photonic

components such as detectors, couplers, and multiplexers will also be useful if processing can be kept similar to that of baseline semiconductor electronics, thereby reducing cost.

The integration of photonics and electronics involves many constraints that need to be considered. There will always be compromises between transistor performance and optical performance. High speed electronics require expensive, cutting-edge processing. Monolithic integration of electronics and optical devices, which do not require such a level of processing sophistication, could be cost prohibitive. Therefore, the choice between hybrid and monolithic integration will be largely driven by economic forces.

The electronics industry has seen revolutionary advances in the last thirty years. Increased research in silicon photonics is just beginning, and one or two major advances in this area could spark a revolution in photonics. Something similar occurred fifty years ago during the transition from vacuum tubes to transistors. In discussions at Bell Laboratories then about the value of transistors, an observation was made that "the significance of the transistor is not that it can replace the tube but that it can do things that the vacuum tube could never do" [65]. The same may hold true for silicon photonics.

Acknowledgements

The authors would like to acknowledge Sean Koehl his help with figures and editing, Ansheng Liu for his assistance with simulations and plots, Drew Alduino for discussions on passive alignment. In addition the authors would like to thank David Ayer, Jerry Bautista, Shekar Bokhar, Randy Mooney, and Valluri Rao for many useful discussions. Finally, the authors would like to thank Prof. Graham Reed of the University of Surrey, UK, for advice on the content and presentation of this text.

Legal Notice

References

1. G. T. Reed, C. E. Png, G. Masanovic, S. T. Lim, S. P. Chan, W. Headley, T. W. Ang, A. Liu, M. Paniccia: *Silicon nanophotonic circuits* (Invited Paper, XLVII ETRAN Conf. 813, Herceg Novi, Serbia and Montenegro 2003)
2. G. T. Reed, A. P Knights: *Silicon Photonics: An Introduction* (Wiley Interscience, New York 2004)
3. R. A. Soref: Proc. IEEE **81**, 1687–1706 (1993)
4. C. K. Tang, G. T. Reed: Electron. Lett. **31**, 451–452 (1995)

5. P. Dainesi, A. Kung, M. Chabloz, A. Lagos, P. Fluckiger, A. Ionescu, P. Fazan, M. Declerq, P. Renaud, P. Robert: IEE Photon. Technol. Lett. **12**, 660–662 (2000)
6. G. Franzò, S. Coffa, F. Priolo, C. Spinella: J. Appl. Phys. **81**, 2784–2793 (1997)
7. L. T. Canham: Appl. Phys. Lett. **57**, 1046–1048 (1990)
8. L. Pavesi, L. Dal Negro, G. Mazzoleni, G. Franzò, F. Priolo: Nature **408**, 440–444 (2000)
9. W. L. Ng, M. A. Lorenço, R. M. Gwilliam, S. Ledain, G. Shao, K. P. Homewood: Nature **410**, 192–194, (2001)
10. J. Trezza, H. Hamster, J. Iamartino, H. Bagheri, C. D. Cusatis: IEEE OCS **1**, 1 (February 2003), S36–S42, suppl. to IEEE Comm. Mag. **41** 2 (2003)
11. Q. Tan, G. De Pestel, G. Willems, J. Cannaerts, W. Rehm, N. Kaiser, K. Vandeputte, A. Van Hove, J. Van Koetsem: Proc. 47$^{\text{th}}$ Electron. Comp. Techn. Conf. (ECTC), San Jose, USA, May 1997, pp. 234–239
12. E. Griese: IEEE Trans. Adv. Packaging **24**, 375–383 (2001)
13. M.-J. Lee, W. J. Dally, P. Chiang: IEEE J. Solid-State Circ. **35**, 1591–1599, (2000)
14. R. Mooney: *Scaling methods for electrical interconnects to meet the performance requirements of microprocessor platforms*, Workshop Notes, 14$^{\text{th}}$ Annual Workshop on Interconnections Within High Speed Digital Systems, Santa Fe, NM, USA, May 2003
15. J. Jaussi, G. Balamurugan, D. Johnson, B. Casper, A. Martin, J. Kennedy, R. Mooney, N. Shanbhag: *An 8 Gb/s source-synchronous I/O link with adaptive receiver equalization, offset cancellation, and clock deskew*, to be presented at 2004 IEEE Int. Solid-State Circuits Conf. (ISSCC), San Francisco, CA, USA, February 2004
16. B. Casper, A. Martin, J. Jaussi, J. Kennedy, R. Mooney: *8 Gb/s differential simultaneous bidirectional link with 4 mv 9 ps waveform capture diagnostic capability*, Tech. Dig. 2003 IEEE Int. Solid-State Circuits Conf. (ISSCC), San Francisco, CA, USA, 2003, pp. 78–79
17. C. Svensson: IEEE Trans. VLSI **10**, 777–788 (2002)
18. N. Savage: IEEE Spectrum, 32–36 (2002)
19. D. A. B. Miller, H. Ozaktas: J. Parallel Dist. Comp. **41**, 42–52 (1997)
20. A. Krishnamoorthy, D. A. B. Miller: IEEE J. Sel. Top. Quant. **2**, 55 (1996)
21. P. Wijetunga, A. F. J. Levi: Proc. Electrochem. Soc. **2002-4**, 381–397 (2002)
22. V. Rao: Proc. UCSB Solid State Technol. Rev. (UCSB SSTR '02), University of California, Santa Barbara, 2002 `http://www.ece.ucsb.edu/OTC/SSTR02/`
23. R. Soref: Proc. IEEE **81**, 1687–1706 (1993)
24. B. Schuppert, J. Schmidtchen, A. Splett, U. Fischer, T. Zinke, R. Moosburger, K. Petermann: J. Lightwave Tech. **14**, 2311 (1996)
25. Y. Liu: IEE Trans. Adv. Packaging **25**, 43–49 (2002)
26. M. Markmann, A. Sticht, F. Bobe, G. Zandler, K. Brunner, G. Abstreiter, E. Muller: J. Appl. Phys. **91**, 9764–9771 (2002)
27. M. Castagna, S. Coffa, M. Monaco, L. Caristia, A. Messina, R. Mangano, C. Bongiorno: Physica E **16**, 547–553 (2003)
28. D. Fehly, A. Schlachetzhi, A. Bakin, A. Guttzeit, G. Wehmann: IEEE J. Quant. Electron **37**, 1246–1252 (2001)
29. H. Kogelnik: IEEE Trans. Microwave Theory Tech. **MTT-23**, 2–16 (1975)

30. S. M. Csutak, J. D. Schaub, S. Wang, J. Mogab, J. C. Campbell: IEEE Proc. Optoelectron. **150**, 3, 235–238 (2003)
31. M. K. Smit: Advanced Components for WDM, Proc. ECOC99, Nice, France, Sept. 1999, pp. 98–99
32. M. Masanovic, L. Val, J. S. Burton, E. J. Skogen, L. A. Coldren, D. J. Blumenthal: IEEE Photo. Tech. Lett. **15**, 1117–1119 (2003)
33. White paper from III–V Photonics, See `www.35ph.com` for details
34. G. Rondeau, S. Biblemont, J. Decobert, G. Post: Compound Semicond. **6**, 83–83 (2000)
35. J. Mills: Photonic Integration Combats Network Complexity, fibers.org (2002)
36. H. Zimmermann: *Integrated Silicon Optoelectronics* (Springer, Berlin, Heidelberg 2000)
37. K. W. Goossen: Integration of GaAs modulators to silicon IC's by monolithic and hybrid techniques, Sarnoff Symposium 1994 (IEEE Princeton/Central Jersey 1994), 100–107
38. D. C. Houghton: J. Appl. Phys. **70**, 2136–1327 (1991)
39. B. E. Little et al.: IEEE Photonics Tech. Lett. **10**, 549–551 (1998)
40. B. E. Little et al.: J. Lightwave Tech. **15**, 998–1005 (1997)
41. D. V. Tishinin et al.: IEEE Photonics Tech. Lett. **11**, 1003–1005 (1999)
42. P. K. Tien: Appl. Opt. **10**, 2395–2419 (1971)
43. D. Marcuse: *Light Transmission Optics*, (Van Nostrand Reinhold, New York 1972), Chap. 9
44. F. Ladouceur, J. D. Love, T. J. Senden: IEEE Proc. Optoelectron. **141**, 242–248 (1994)
45. `www.rsoftdesign.com`
46. K. Lee et al.: Opt. Lett. **26**, 1888–1890 (2001)
47. W. A. Pasmooij et al.: J. Lightwave Tech. **7**, 175–180 (1989)
48. T. Tamir, S. T. Peng: Appl. Phys. **14**, 235 (1977)
49. T. W. Ang, G. T. Reed: A. Vonsovici: Appl. Phys. Lett. **77**, 4214 (2000)
50. D. Taillaert et al.: J. Quant. Electron **38**, 949–955 (2002)
51. T. W. Ang et al.: Proc. SPIE **3620**, 79–86 (1999)
52. T. W. Ang et al.: Appl. Phys. Lett. **77**, 4214–4216 (2006)
53. P.E. Barclay et al.: Conf. Lasers Electro-Optics (CLEO), Baltimore, MD, USA, May 2003
54. I. Day et al.: Tapered silicon waveguides for low insertion loss highly efficient high speed electronic variable attenuators, IEEE Opt. Fiber Comm. Conf. (OFC 2003), Atlanta, GA, USA, March 2003
55. M. B. Frish et al.: Proc. 2002 IEEE Int. SOI Conf. (IEEE, 2002), pp. 165–166
56. L. Lai, E. A. Irene: J. Appl. Phys. **86**, 1729-1735 (1999)
57. W. Kowbel et al.: Aerospace Conf. Proc. **5**, 2423–2428 (2002)
58. G. Lecarpentier, J. Mottet, J. Dumas, K. Cooper: High accuracy machine automated assembly for opto electronics, 50^{th} Electron. Comp. Technol. Conf. (ECTC), Las Vegas, USA, May 2000, pp. 1–3
59. T. Hashimoto, Y. Nakasuga, Y. Yamada, H. Terui, M. Yanagisawa, K. Moriwaki, y. Suzaki, Y. Tohmori, Y. Sakai, H. Okamoto: IEEE Photon. Technol. Lett. **8**, 1504–1507 (1996)
60. K. P. Jackson, E. B. Flint, M. F. Cina, D. Lacey, Y. Kwark, J. M. Trewhella, T. Caulfield, P. Buchmann, C. Harder, P. Vettiger: J. Lightwave Technol. **12**, 1185–1191 (1994)

61. R. A. Nordin, D. B. Buchholz, R. F. Huisman, N. R. Basavanhally, A. F. J. Levi: IEEE Transactions on Components, Hybrids, and Manufacturing Technology **16**, 783–788 (1993)
62. Q. Tan, Y. C. Lee: Soldering technology for optoelectronic packaging, Proc. 46th Electron. Comp. Technol. Conf. (ECTC), Orlando, FL, USA, May 1996, pp. 26–36
63. A. Ambrosy, H. Richter, J. Hehmann, D. Ferling: IEEE Transancations of Components, Packaging and Manufacturing Technology, Part A **19**, 34–40 (1996)
64. `www.photond.com`
65. H. Huff: J. Electrochem. Soc. **149**, 35–58 (2002)

Index