INP AND III-V COMPOUNDS

Contents

Situation (Infrastructure) Analysis 2
Materials 2
Manufacturing processes 2
Epitaxial growth and regrowth *3*
Lithography *3*
Etching *4*
Deposition and annealing *4*
Lapping and polishing *4*
Cleaving and coating *4*
Manufacturing Equipment 4
Quality/Reliability 5
Environmental Technology 5
Material availability *5*
Health issues *5*
Test, Inspection, Measurement (TIM) 6
Manufacturing Cost 7
Contributors 9

SITUATION (INFRASTRUCTURE) ANALYSIS

MATERIALS

The group of III-V semiconductor materials have similar growth properties to silicon, which is well established in microelectronics. III-V semiconductor materials are epitaxially grown on mono-crystalline semiconductor substrates. A main difference is in the opto-electronic properties, where most III-V semiconductors have a direct bandgap, which is a prerequisite for making efficient lasers and optical amplifiers, a property silicon is missing. Additionally, several III-V semiconductors, such as GaAs and InP have better electronic properties than silicon, which makes them suited for high-end RF-applications.

A key difference between the various III-V semiconductors is the wavelength range in which they support optical functions like generation, amplification, transmission, and detection of light. For GaAs, which was the first III-V material applied in semiconductor lasers, the operation window ranges from 800-1100 nm, which makes it suitable for short-range communication. GaAs vertical-cavity surface-emitting lasers (VCSELs) are the dominant light source for short distance (< a few hundreds of meters) communications. For InP and its quaternary compounds InGaAsP and InGaAlAs, which can be grown on an InP substrate, the operation window ranges from 1200-1700 nm, which covers the most important wavelengths for high-speed communication over longer distances (O-band, C-band, and L-band). It is, therefore, the material of choice for high speed communication over long and medium distances.

An additional advantage of InP and its compounds InGaAsP and InGaAlAs is that their optical properties (gain, transparency, absorption and detection, and electro-optic modulation efficiency) can be engineered locally within the wafer while retaining the possibilities for optimizing performance over a wide wavelength range. This makes it the material of choice for use in complex PICs where a wide range of functionalities has to be integrated into a single chip. Examples are coherent transmitters and receivers, and more generally, any circuit where lasers and optical amplifiers need to be integrated with efficient modulators and detectors, as well as low-loss passive-optical elements (e.g. optical filters).

Dielectric materials for passivation and isolation are very similar to those used for silicon microelectronics. Metals for electrical inter-connections are different. Gold is frequently used for III-V semiconductors because of its good electrical and mechanical properties, whereas it is not applied on silicon because of the risk of diffusing into the silicon, where it is very harmful. On the other hand, aluminium and copper are seldom used for III-V materials. In particular copper impurities degrade electrical and optical properties in III-V materials.

Wafers commercially available for III-V materials are smaller than for silicon. For GaAs 4”, 6” and 8” diameters are commercially available. InP wafers with 2”, 3” and 4” diameter are commercially available with good quality. Larger 6” wafers are commercially available for R&D purposes, with a slightly larger Etch Pit Density (EPD), which will be improved when the demand increases.

MANUFACTURING PROCESSES

The most important group of processes for the fabrication of III-V photonic components and integrated circuits are:

- epitaxial growth,
- lithography,
- etching of semiconductor material and dielectrics,
- deposition of dielectrics and metals for passivation and metallization,
- grinding and polishing, and
- cleaving and coating.

We will briefly describe them.

Epitaxial growth and regrowth

The first step in the fabrication of most III-V components or integrated circuits is the growth of the epitaxial layer stack, which usually includes a number (up to a few tens) of layers with different compositions and doping types and levels, including Quantum Well or Quantum Dot layers. InP based materials (including the substrate) can be made semi-insulating by doping with Fe atoms, thus enabling efficient electrical isolation of the individual integrated devices and facilitating very-high-frequency operation.

InP-based materials facilitate access to a wide range of bandgaps, which is required to monolithically integrate low-loss passive and high-performance active functions with precise wavelength control and detuning between laser and modulator. This requirement increases complexity in epitaxy and material characterization as well as mask design. Several integration technologies have been commercialized (impurity induced layer disordering, butt-joint regrowth and selective area regrowth) each with trade-offs in manufacturing cost, yield and performance. We will briefly discuss the most commonly used approaches for high performance monolithic integration.

In many PICs several different layer stacks are monolithically combined, by using selective butt-joint regrowth: the first grown layer stack is removed everywhere where it is not needed using lithography and etching, after which a second layer with a different layer stack is locally grown. If more than two different layer stacks are needed this process can be repeated. In this way we can get optimal layer stacks for different components (e.g. lasers, modulators, detectors and transparent waveguides) at the regions where those components are needed.

An alternative approach is selective area regrowth (SAG) which is a special integration process relying on local growth rate change induced by proximity to a dielectric mask. In a Quantum Well layer stack this change in the growth rate leads to a shift of the band edge, which can be beyond 100 nm, thus allowing fabrication of lasers, detectors, modulators and transparent waveguide devices with a single growth step.

The quality and control of the epitaxial layer stack is of key importance not only for the performance of active components like lasers and optical amplifiers, but also for modulators and detectors. In this respect III-V technology differs from silicon technology, where epitaxial growth is applied for special cases such as Ge detector layers and SiGe Quantum Wells but is not used in the mainstream technology.

The most frequently used epitaxial growth technique is Metal-Organic Chemical Vapor Deposition (MOCVD), also known as Metal Organic Vapour Phase Epitaxy (MOVPE). For components which require very large dopant gradients, such as Avalanche Photodiodes (APDs), Molecular Beam Epitaxy (MBE) is sometimes used, which operates at a lower temperature. MBE reactors require a high vacuum and have lower deposition rates, which makes them less suitable for very high-volume production. However, they have proven competitive for a number of dedicated applications. MOCVD reactors are provided by a number of manufacturers, with a focus on the large volumes required for fabrication of LEDs. In comparison with the LED market, the PIC market is very small hindering many manufacturers from making the large investments necessary to develop automated high-performance equipment tailored to PIC manufacturing.

Lithography

The most frequently used lithography is I-line stepper lithography, having a resolution of about 250 nm, though contact lithography may be used for less critical steps. For higher resolution, as required in fabricating gratings, E-beam lithography is typically used, which is a direct-write technique with lower throughput. High resolution tools with high throughput, such as the 193 nm DUV scanners used in microelectronics, are not yet generally available for InP because the machines were not designed for exposing wafers smaller than 6”. It has already been demonstrated that the tools can be adapted for 3” and 4”. Optimizations on both the process and the required wafer properties (such as flatness) are to be intensified.

Etching

Removing part of the epitaxial layer stack by etching is an important step in any PIC processing. The most frequently used processes are Reactive Ion Etching (RIE) and Inductively Coupled Plasma (ICP) etching. The used chemistries are commonly CH_4/H_2 and Cl and Br etchants. Etching control and uniformity are usually of the order of a few percent, which is sufficient for many applications, but not for complex high-performance PICs. Edge roughness is extremely important in high-confinement waveguides in order to keep propagation losses low.

The edge roughness is caused by a combination of roughness generating mechanisms in both the lithography and the etching. Etching of dielectric layers is performed with wet chemistry or with RIE or ICP dry etching. If the dielectric layer is used as a hard mask, the etching requirements on edge roughness are very tight. Due to the use of platinum and gold, which are difficult to etch, metal patterns are usually fabricated with lift-off lithography.

Deposition and annealing

Deposition of dielectric layers for passivation or for use as a hard mask (SiO_x or Si_3N_4) is usually done with Plasma Enhanced Chemical Vapor Deposition (PE-CVD), which can be performed at moderate substrate temperatures. Layer thickness control and uniformity are usually of the order of a few percent, which is sufficient for many applications, but not for all. Stress in deposited films is also an important parameter to be controlled.

For deposition of metals usually E-gun evaporation is used, which has good properties for use as selective deposition in conjunction with lift-off lithography. If better adhesion or better step coverage is required, RF-diode or magnetron sputtering is the preferred alternative.

Usually an annealing step is required for getting good passivation properties at the interface between the semiconductor and dielectric layer materials and for getting a low resistance at the interface between the semiconductor and contact metals. Here the most frequently used process is Rapid Thermal Annealing.

Lapping and polishing

Prior to singulating wafers into individual chips, the PIC wafers are thinned to improve the quality of cleaved facets. For these two steps, the wafers are mounted on a plate using wax or a film to protect the process side. The mounted wafers are mechanically lapped by placing between two counter-rotating cast iron plates or a grinding wheel with chemically abrasive slurries. Subsequent polishing or chemical mechanical planarization (CMP) is required to remove scratches and damage from the lapping process.

Cleaving and coating

For singulating PICs from a processed wafer, the most frequently used process is cleaving: small scratches (scribes) are made at the edge of the wafer after which controlled pressure is exercised in order to cleave the wafer along a crystal plane. On proper cleaving, the chip has an atomically flat end facet. For proper cleaving, thinning the wafer down to 100-200 μm is an essential and demanding process, particularly for larger wafer sizes. After thinning the wafer is first cleaved into bars, which contain a series of PICs. After cleaving, the facets are coated to achieve a defined reflectivity, mostly anti-reflection (AR) or high-reflection (HR) coatings are used. This process is performed shortly after cleaving to keep a clean interface. The next step is cleaving or dicing the bars into single PICs. PICs are often tested on bar level.

MANUFACTURING EQUIPMENT

In most cases, InP-fabs use equipment developed to process other materials, such as GaAs and silicon, e.g. equipment for wafer handling, epitaxy (with requirements on control of layer thickness and composition), lithography (high resolution and exposure in the order of 10 units per hour), etching (low-damage, accurate depth control, deep etching) and metrology with 3D measurements and high aspect ratio (HAR) capabilities. However, InP-fabs demand special requirements such as the capability to process smaller wafers (3”and 4”) and handling of

fragile InP wafers without damage. InP devices have challenging lithography requirements with respect to line-edge roughness, large depth of focus and extreme critical dimension (CD) control, while additionally the cost of use should match the market size. This requires existing equipment to be adapted to enable the processing of InP.

Currently, most equipment for InP volume manufacturing is partially automated, however, many steps are still performed manually. Full automation will increase process reliability and reproducibility and reduce manufacturing costs. So far, the market need has not been large enough to motivate the investments to develop fully automated high-performance equipment tailored for high volumes.

QUALITY/RELIABILITY

InP lasers and GaAs VCSELs produced in large numbers for long, medium and short-range communication links demonstrate that III-V components, if properly manufactured and packaged, show very good reliability with product lifetimes of at least 20 years for telecom parts. Also, for PICs integrating tens of components such as complex tunable lasers, good yields are reported, although no detailed information about yield numbers is given by the manufacturers. The monolithic encapsulation of active elements within passive circuits does offer a strong yield-driven motivation for tighter integration.

It should be noted that yield is strongly dependent on performance specifications. Yield is influenced by the number of killer defects, but this is usually low. In the same way that silicon electronics is not limited by fundamental density-related yield mechanism, there is no evidence of such a limit for III-V PICs today. Depending on the level of automation and scaling, wafer yield is more likely to be dominated by manual handling, (tight) processing windows, tool stability, and assembly technologies. Experiments with large PICs suggest that with adaption of high-performance manufacturing equipment yields can be high, also for high-performance PICs. Furthermore, it should be noted that yield reductions can occur in several consecutive stages of the device fabrication. For example, assembly and packaging are other manufacturing steps which can introduce significant yield reductions.

ENVIRONMENTAL TECHNOLOGY

Material availability

There have been discussions about the availability of InP for a long time. Indium is available in large amounts in ores, but strongly diluted in other materials. It is won as a by-product of other metals (mostly zinc). Prices may increase significantly if the demand increases beyond the level that is supported by the production of other metals. Indium prices showed a strong increase at the beginning of the century (up to 1000 $/kg) but have since stabilized at roughly half that price. Most indium is used to make indium-tin oxide (ITO), which is an important constituent of touch screens, flat-screen TVs and solar panels. It is also used in microelectronics, and as a special coating for glasses and bearings.

If the demand for PICs increases as expected and we move to 6" wafers which are significantly thicker than 3" wafers, the demand for Indium Phosphide substrates will increase, but this will have a relatively small overall impact. As the bill of materials for PICs is only a small part of the total costs, the effects of an increase in materials costs on the PIC costs are marginal.

Health issues

The use of III-V materials, and in particular InP, has been the subject of health risk analysis. There are indications that InP-based compounds and precursors used in the manufacturing of PICs are toxic on exposure to high concentrations, and in Europe procedures are running to classify InP as toxic material. InP dust can be generated during polishing or thinning of wafers or singulation of PICs, the latter especially when wafers are diced instead of cleaved. Over the years, the PIC and III-V manufacturing industry has implemented stringent safety and health regulations in their fabrication process, containing dust and other by-products. In an adequate cleanroom environment, the risks for human health are considered to be very low.

TEST, INSPECTION, MEASUREMENT (TIM)

PIC testing contributes significantly to the cost of a module. In Figure 1 the typical stages in the PIC supply chain are presented in the top row, while testing aspects relevant to those stages are shown in the lower row.

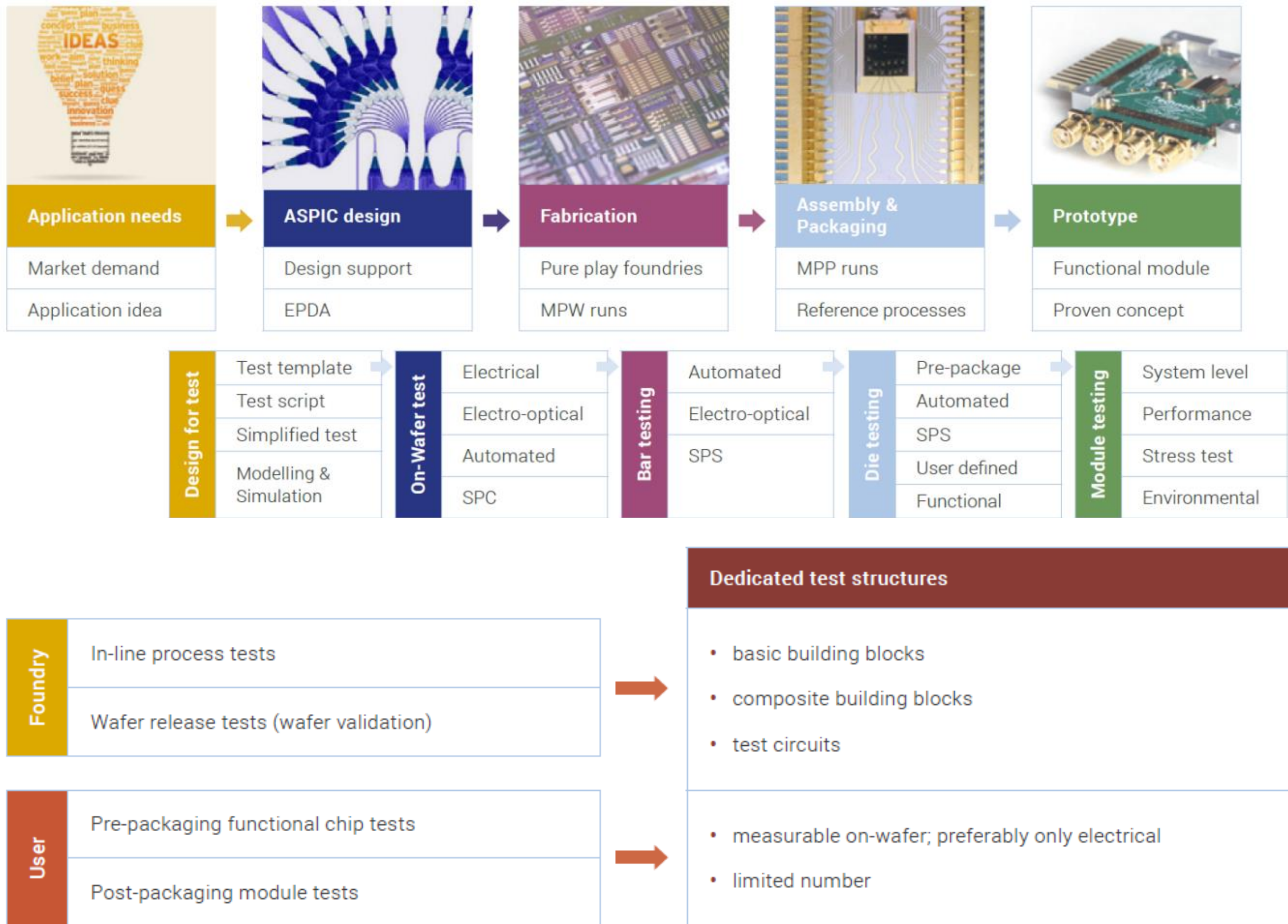

Figure 1 Testing across value/supply chain of Application Specific Photonic Integrated Circuits. (source: JePPIX roadmap 2018)

Significant R&D effort is required to introduce and improve testing at all stages of the PIC process and supply chain. This will allow manufacturers to optimize and accelerate the whole production process and enable early identification of Known-Good-Dies (KGD). In order to facilitate fast testing procedures optical parameters should be measured in an electrical fashion wherever possible. Dedicated test structures relevant for test requirements of foundries and users need to be developed, as depicted in Figure 1.

Smart testing throughout the production process is required to ensure earlier testing to reduce process spread, optimize the PIC manufacturing process windows and maximize yield. An increased level of automation across the full supply chain will result in a reduction of time required for testing and KGD identification. For wafer verification, on-wafer measurements in both the electrical and the optical domain are desirable to allow for testing at various sites across the wafer prior to cleaving. To this end vertical optical out-coupling structures should be integrated. Viable implementation options could be turning mirrors and grating couplers integrated into appropriate waveguide sections. However, as of today, in contrast to Si based PICs, vertical grating couplers are not routinely used in InP-PICs due to their low coupling efficiency and their large real estate.

MANUFACTURING COST

Scaling laws for InP are similar to other thin film fabrication technologies used in CMOS electronics or silicon photonics. Costs are primarily dominated by the amortization costs for the fab, the complexity of the process (number of process steps) and by the loading of the fab. Material costs usually make up only a small part of the total PIC costs. The cost of InP, silicon photonics (SiPh) or polymer PICs, for example, is not primarily determined by the costs of the material, rather by the cost of the process, which is to a high degree determined by the number and the complexity of the processing steps and by the production volumes.

Figure 2 illustrates the cost dependence of InP-PICs on the aggregate annual load of the fab and the size of the wafers run in the fab. It is based on a simple model as described at the end of this sub-section. The load of the fab is expressed in the total number of chips/year for an average chip size of 10 mm^2. For smaller chips the curves will shift to the right, for larger chips to the left.

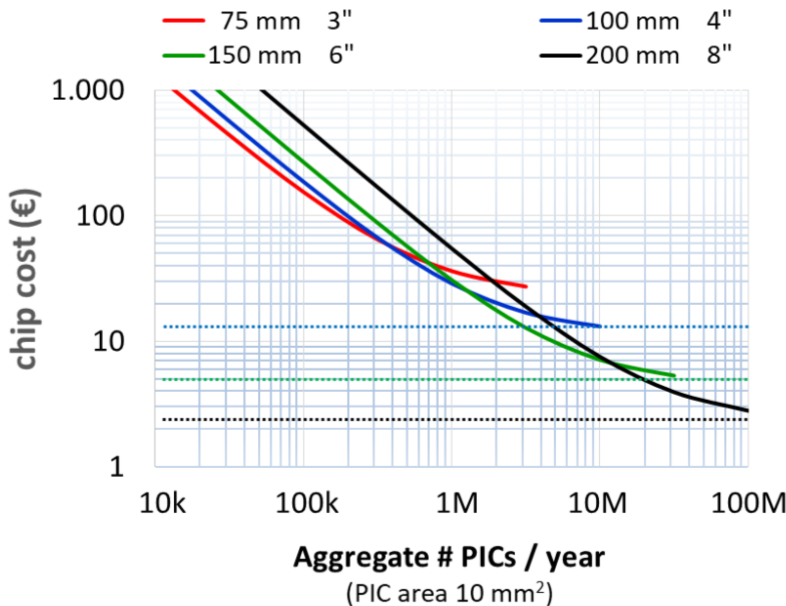

Figure 2. Chip cost as a function of yearly fab load for different fab scenarios, calibrated for an average chip size of 10 mm^2.(source: JePPIX roadmap 2018)

For volumes much smaller than the fab capacity, the chip cost is dominated by the investments in the fab, which are high for larger fabs. So, for smaller aggregate volumes, large fabs are more expensive than small fabs. For volumes approaching the fab capacity the chip cost is mainly determined by the processing cost, which is only weakly dependent on the wafer size; a large wafer in an expensive fab is not much more expensive than a small wafer in a small fab. This is the main reason that large fabs are more cost effective at high volumes.

The solid lines indicate the dependence of the PIC costs on the aggregate annual volumes. PIC costs are minimal if the fab is fully loaded. The costs are then dominated by the marginal costs of processing a batch of identical wafers. In a fully loaded fab, smaller volumes using the same process can be fabricated at costs which are close to those of large volumes, as indicated by the horizontal dotted lines.

The line labeled 200 mm is indicative of a small silicon photonics fab. If the PICs are fabricated in a CMOS process in a fab which is fully loaded by electronic ICs, they can be produced at low cost, even when the aggregate volume of photonic ICs is much smaller than the fab capacity, because the aggregate volume is determined by the electronic ICs. For larger wafer sizes of 300 mm operating at higher throughput, the dotted line will be even lower.

If the PICs are fabricated in a dedicated photonic process in a CMOS fab the cost curve will be somewhere between the solid and the dotted curve: the costs of the equipment are shared with electronic IC production, but the costs of the process not. For advanced silicon photonics processes on large wafers, for example, the costs of the masks are

extremely high, which increases the costs of a photonics run significantly at lower volumes, thus rendering silicon photonics MPWs in advanced processes more expensive than InP MPWs in smaller fabs.

The mask costs for Silicon Photonics are higher because of the higher node, a typical SiPh process is run at 45nm node. The requirements for SiPh are stricter as the waveguide dimensions are much smaller. As an example, the losses in a waveguide scale with $roughness^2/width^4$ so in InP this requirement is 10x less; similar scalings apply for phase errors.

The graphs are indicative for the dependence of chip cost on aggregate volume and wafer size, but actual costs may differ significantly from the costs shown in the graph. If an old fab is used, which is already largely depreciated, the costs at lower volumes can be significantly lower. Investment costs are also dependent on the wafer capacity of the fab and the degree of automation, which will increase the investment cost and reduce the marginal wafer cost (the horizontal dotted line). Further, yield is an important factor which is strongly dependent on user requirements: if the requirements are well within the building block specifications and the design rules, it will be high. But if they are close to the process window limits, it can be significantly lower. With these boundary conditions in mind we can draw a few conclusions from the graph:

- Because the square millimeter costs are strongly dependent on the total volume, all users of a fab running an open-access generic process can get their chips at a price corresponding to the aggregate yearly chip volume, even though their own chip volume may be much smaller. This will also make the costs for small users significantly lower.
- For a square millimetre price below 10 €/mm^2, volumes well over 100,000 chips per year are required for a 10mm^2 die area.
- For a square millimetre price below 1 €/mm^2, volumes well over 1 million chips per year are required. This cost reduction should be achieved by tool automation.

We anticipate that square millimeter costs of InP PICs will generally remain higher than for SiPh PICs. However, the costs of an (InP) light source and its assembly (or heterogeneous integration) need to be added for SiPh-PICs. Additionally, active InP building blocks such as phase-modulators and lasers can be significantly smaller. When the costs of advanced InP-PICs is reduced to a few Euros/Dollars, InP-PICs are expected to be very competitive for medium or even larger volumes where high performance is required, and also where complex functionalities (a number of lasers and/or optical amplifiers on board) are required.

Explanation of the Model of Figure 2.
The curves in Figure 2 are based on a very simple model that is useful for a qualitative analysis of the effects of scaling on chip cost. For quantitative purposes it is not accurate enough, although with an appropriate choice of input parameters it can be indicative for cost levels.
The model calculates the cost of a chip from the annual depreciation of the fab cost and the running cost divided by the number of chips that the fab produces per year. And it adds to this the marginal costs of processing a wafer if the fab is (almost) fully loaded. The analysis is done for 4 different fab types with wafer diameters of 3", 4", 6" and 8".

		Fab 1	Fab 2	Fab 3	Fab 4
1	Wafer diameter (mm)	75	100	150	200
2	Fab cost (M€)	75	100	150	300
3	Fab capacity (wafer/yr)	10000	20000	50000	500000
4	Useful wafer area (%)	70%	70%	70%	70%
5	Useful wafer area (mm2)	3,091	5,495	12,364	21,980
6	Reference PIC size (mm2)	10	10	10	10
7	Fab Capacity (PIC/yr)	2.16E+06	8.79E+06	5.56E+07	1.10E+09
8	Depreciation / yr	10%	10%	10%	10%
9	Running cost / yr	4%	4%	4%	4%
10	Yearly Fab Cost (M€)	10.5	14	21	42
11	Marg cost / wafer (k€)	4	4	4	4
12	Yield (%)	70%	80%	90%	100%
13	Profit margin	25%	25%	25%	25%

Row 1 in the table shows the wafer size. The first three fabs could be InP fabs, the latter one a dedicated silicon fab or a fab that processes PICs in a photonic InP membrane on top of an 8" silicon wafer, as described in section 6.1. Row 2 presents the investment in building the fab (in millions of euros). It is assumed that fabs for large wafers are more automated and have, therefore, a larger wafer capacity (row 3). Row 4 gives the useful wafer area in % and row 5 in mm^2. Row 6 specifies the average PIC size for which the analysis is done, in this example 10 mm^2. From rows 3-6 the fab capacity in PICs/year (row 7) is calculated. The yearly exploitation cost of the fab (row10) is calculated as the sum of the yearly depreciation (row 8) and the yearly running cost (row 9). Row 11 shows the marginal costs for running a wafer in a fully loaded fab (material and operator cost). The curves in Figure 2 show the yearly fab costs (row 10) divided by the total number of PICs that the fab produces per year, increased by the marginal cost per PIC (row 11 divided by # PICs/wafer), corrected for yield (row 12).

CONTRIBUTORS

Co-chair	Mike Wale	Eindhoven University of Technology, NL
Co-chair	Gloria Hoefler	Infinera, US
Past chair	Meint Smit	Eindhoven University of Technology, NL

TWG Members:

Mohand Achouche	3-5 Lab, FR
Huub Ambrosius	Eindhoven University of Technology, NL
Luc Augustin	SMART Photonics, NL
Ton Backx	Eindhoven University of Technology, NL
Victor Calzadilla	Eindhoven University of Technology, NL
Jean-Louis Gentner	Almae Technologies, FR
Giovanni Gilardi	Finisar, US
Thomas Collins	Huawei, BE
Norbert Grote	Fraunhofer HHI, GE
Robert Harper	Compound Semiconductor Centre (CSC), UK
Hiroshi Ishikawa	Hikari Path Communications, JP
Tim Koene	EFFECT Photonics, NL
Jinkwan Kwoen	University of Tokyo, JP

Martin Moehrle	Fraunhofer HHI, GE
Yoshi Nakano	University of Tokyo, JP
Shu Namiki	AIST, JP
Yasutomo Ota	University of Tokyo, JP
Kelvin Prosyk	Ciena, US
Sylvain Raymond	NRC-CNRC, CA
Frank Shepherd	NRC-CNRC, CA
Toshi Umezawa	NICT, JP
Katsuyuki Watanabe	University of Tokyo, JP
Kevin Williams	Eindhoven University of Technology, NL
Hideki Yagi	SEI, JP
Weiming Yao	Eindhoven University of Technology, NL
Shaoxian Zhang	VA-Photonics, NL

Fair use disclaimer: this report contains images from different public sources and is a compilation of the opinion of many experts in the field for the advancement of photonics, the use of which has not always been explicitly authorized by the copyright owner. We are making such material available in our efforts to advance understanding of integrated photonics through the research conducted by the contributors of this report. We believe this constitutes a fair use of any such copyrighted material as provided for in section 107 of the US Copyright Law. In accordance with Title 17 U.S.C. Section 107, the material in this report is distributed without profit to those who have expressed a prior interest in receiving the included information for research and educational purposes.