School of Electrical, Electronic and Communications Engineering, UCD

EEEN20070 Solid-State Electronics (1)

AN INTRODUCTION TO SEMICONDUCTOR PROCESSING

Semiconductor processing refers to the technology required to fabricate and manufacture semiconductor devices. The semiconductor industry has historically sought to produce improved electrical performance at continually lower and lower cost, and it has succeeded dramatically in this objective. A key factor in this success has been the increased scale of the batch production processes used, that is, large numbers of components are fabricated at the same time. A relatively small proportion of total production is geared towards realising so-called discrete semiconductor devices (diodes, transistors…). The great majority of current semiconductor production is devoted towards Integrated Circuits (IC's), consisting perhaps of many millions of interconnected transistors and other components realised on the surface of a piece of silicon about 1cm^2 in area with surface features defined down to ~20nm (1nm= 1 nanometre =0.001μm = 10^{-9} meter) in 2014.

*The progressive reduction of this minimum feature size over time is the main technical 'driver' of the industry. Over the past 45 years, the minimum size has reduced by about 70% every 18 months, leading to a doubling of the potential electronic complexity in a given area. This famous observation of effectively exponential growth in complexity of a given chip over time was first made by the founder of Intel, Gordon Moore in 1965 and is known as '**Moore's Law**'. It is astonishing that it has proven valid for such a long time and it is expected to continue to be valid up towards at least the end of the this decade. The impact on society of the technological revolution encapsulated by Moore's Law has been profound and immense – without it we would have no personal computers, mobile phones, Internet etc. The large Intel facility at Leixlip, Co. Kildare employing some 4000 people, is at the forefront of the world state-of-the-art in IC production.*

It is worth emphasising at the outset that the subject of semiconductor processing, although of enormous commercial and economic importance, is largely concerned with technology which changes quite rapidly, and we will therefore not concentrate on the details here. A basic awareness of the main processes used is important, however, for any electronic or electrical engineer. Students should fill out the following with personal reading, (e.g. Chapters 1 and 9 of Streetman and Banerjee's book "Solid State Electronic Devices" (6th ed.) or by accessing web resources (e.g.

http://jas.eng.buffalo.edu/education/fab/pn/diodeframe.html
http://www.youtube.com/watch?v=4Q_n4vdyZzc&feature=related
http://www.pcworld.com/article/227142/intel_to_bring_3d_transistors_to_nextgeneration_chips.html etc.)

1. Bulk Crystal Growth:

The starting point for most semiconductor manufacture is the preparation of a large (approximately-cylindrical) "boule" of high-quality single-crystal semiconductor material. This is then sliced using a diamond saw into a large number of thin "wafers" of a few hundred μm thickness. Many wafers are processed at the same time, and on the surface of each there may be hundreds or even thousands of identical discrete devices or IC's. Hence, the very high processing costs of each batch may be reduced by being spread over a very large number of components during manufacture.

An increase in the scale of the batch production process can be directly obtained by an increased wafer diameter (D). In fact, the number of semiconductor components obtainable from a wafer increases as D^2. In the case of Si, the maximum wafer size is about 300mm. On the other hand, the rather limited D_{max} for GaAs (up to about 100mm) is a major factor in the higher cost of devices using this material (see Fig. 1).

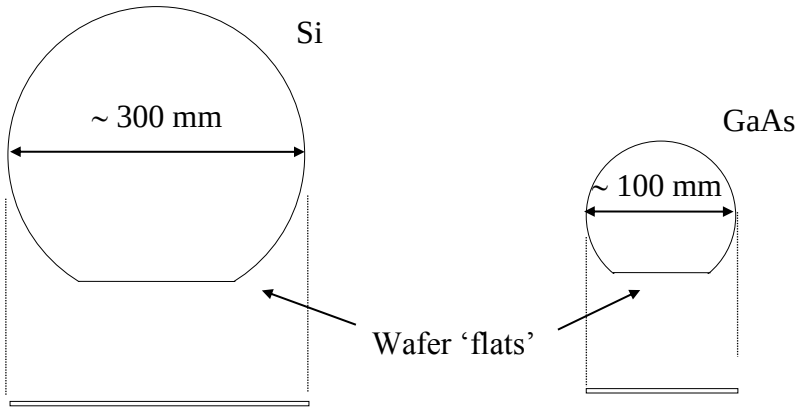

Fig. 1 Comparative Size of Maximum Production Wafer Diameters in Si and GaAs

We first consider the process by which the boule is created (concentrating mainly on silicon). The most common technique for single crystal growth is the **Czochralski "Growth from the Melt" process**.

The "Melt" is formed by RF heating of polycrystalline rods of relatively pure Si, containing about 1 in 10^9 residual impurities (mainly C or O_2), leading to an intrinsic resistivity $\rho_i \approx 200\ \Omega$-cm. Very often controlled amount of impurities are added to the melt, to create *extrinsic* P-type or N-type material. The polycrystalline rods are formed in turn by reduction of silica or quartz (SiO_2) with carbon. Silicon is one of the most abundant elements in the earth's crust (~ 22%) mainly in the form of SiO_2. Indeed, ordinary beach sand is mostly SiO_2. The reaction for silica reduction is:

$$SiO_2 + 2C \rightarrow \mathbf{Si} + 2CO\uparrow$$

The Czochralski growth process starts with a small high-quality piece of "seed" crystal, and is performed in a chamber filled with inert gas (Fig. 2). Rotation is performed during the growth to ensure uniformity of distribution.

The crystallographic orientation of the entire crystal is determined by the orientation of the seed (e.g. <100>, <111> etc.).

For GaAs growth a similar technique may be used except that As is volatile, and the whole surface must be covered by a dense layer of molten Ba_2O_3. This leads to LEC growth of GaAs (Liquid-Encapsulated Czochralski).

A float-zone technique (FZ) can be used to produce very high purity Si using "zone refining", however D tends to be reduced (see the references mentioned earlier for more details).

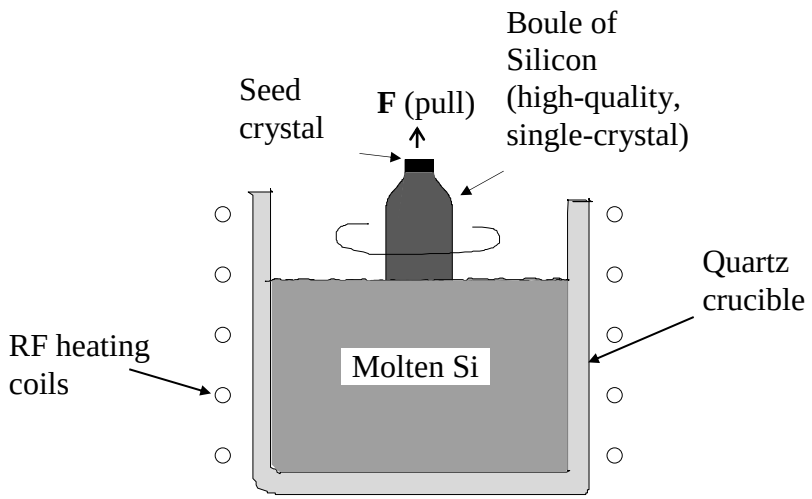

Fig. 2. Czochralski 'Growth from the Melt' Process

2. Epitaxial Growth:

The wafers discussed above are rarely used directly for devices. Rather, they provide a "substrate" or "bulk" region to support later processing operations, which are concentrated on a thin layer (< ~10 μm) on one side of the wafer. One of the most important such processes is *epitaxial growth*. Epitaxy (or 'epi', from the Greek for "arranged upon") involves the growth of a very high quality single crystal layer on the substrate, usually with different dopant characteristics to the underlying substrate, and with a well-defined thickness.

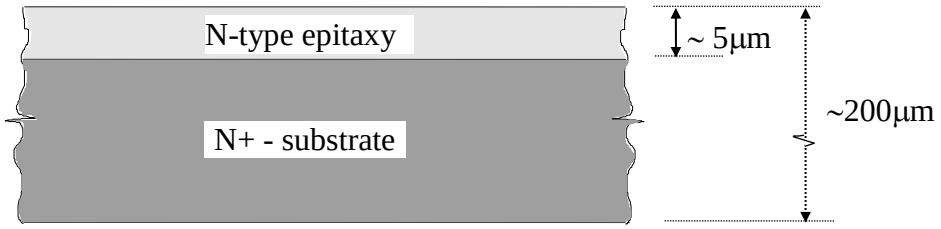

Fig. 3.: Example of Cross-Section Through Wafer Showing Epitaxial Layer

There are many ways of performing epitaxial growth. In all cases, the epi layer takes on the same crystal structure as the substrate. They are:

- **Vapour Phase Epitaxy (VPE):** this is very widely used in Si ICs and for GaAs. One method involves bubbling hydrogen through silicon tetrachloride, and exposing the heated substrate to the gas flow in a reaction vessel:

$$SiCl_4 + 2H_2 \rightarrow \mathbf{Si} + 4HCl\uparrow \quad [1000°C +]$$

  Other (impurity) gases can be easily added to dope the resulting epitaxial layer.

- **Liquid Phase Epitaxy (LPE)**: this process is based on the fact that a mixture of the semiconductor and a second element may melt at a lower temperature than the semiconductor itself. Thus, the molten mixture may be placed on top of the (solid) semiconductor. As it is cooled, a single crystal "skin" forms on the semiconductor and this forms the basis of LPE. This growth technique is often used for III-V devices including lasers and LED's.

- **Molecular Beam Epitaxy (MBE):** the control of thickness and doping properties achievable with VPE and LPE is limited, and recent years have seen an enormous increase in the use of techniques such as MBE. The equipment is expensive but the control is exceptionally good, down to ~ atomic monolayers. In MBE the substrate is positioned in an ultra-high vacuum chamber, and heated to a moderate temperature (~600°C typically). Individual element sources are contained within heated cells, with controllable shutters. The elements are evaporated to form collimated beams which can be precisely controlled to produce high-quality crystal growth on the substrate (e.g. growth rates ~ 1 μm/hour). The process is slow and expensive for production quantities however.

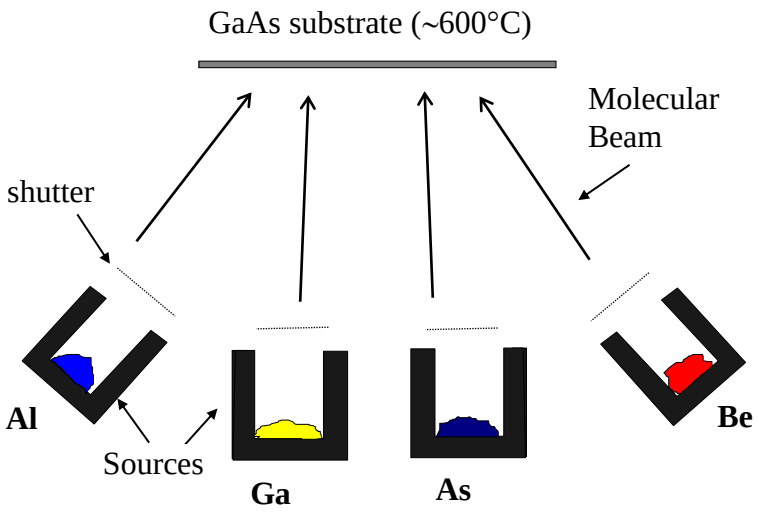

Fig. 4. Example of Molecular Beam Epitaxy Growth System

The system shown in Fig. 4 could be used to grow Beryllium-doped (P-type) AlGaAs on GaAs.

- **Metal-Organic Chemical Vapour Deposition (MOCVD):** This technique is based on a chemical reaction between two compounds, one of which is an organo-metallic compound. For example, trimethylgallium may be reacted with arsine (AsH_3) to form GaAs and methane:

$$(CH_3)_3Ga + AsH_3 \rightarrow \mathbf{GaAs} + 3CH_4\uparrow$$

This technique provides very high quality epitaxy (though not as good as MBE). It is a faster growth process, better suited to production and is widely used for the production of III-V devices for microwaves and optoelectronics. In Ireland, the Tyndall Institute in Cork operates machines of this kind, mainly for optoelectronics applications.

3. Lattice-Matching and Heterostructures:

As we have already seen, epitaxial layers take on the crystal structure of the underlying substrate. Normally, one semiconductor has a different "lattice constant" (dimension of unit cell) to another, and therefore one material cannot be grown successfully on the other. If, however, a binary compound semiconductor is grown on an elementary semiconductor (e.g. SiGe on Si) or a ternary semiconductor on a binary (e.g. AlGaAs on GaAs) we can use the extra degree of freedom produced in material composition to achieve lattice matching, and a stable epitaxial structure

results. Such a structure, consisting of two semiconductors with dissimilar electrical properties but mutually lattice-matched, is called a heterostructure.

Figure 5 shows an example of possible heterostructure combinations in the case of some III-V semiconductors. The graph shows, for example, that AlGaAs and GaAs are reasonably well lattice-matched over the entire range of compositional parameter 'x', while $In_xGa_{1-x}As$ may be grown lattice-matched to InP for a value of x ~ 53%.

Heterostructures have been intensively developed in recent years, especially using techniques such as MBE and MOCVD. Furthermore, such techniques allow growth of extremely thin layers (~1-10 nm) of materials which are not strictly lattice matched to each other - the thin grown layer takes on the lattice constant of the starting layer, and is thus under mechanical strain (tensile or compressive) - such a layer is said to be "pseudomorphic".

The band gap energy (Eg) is a critical semiconductor parameter, affecting all aspects of behaviour, especially optical properties. Clearly, heterostructures allow substantial changes of Eg across a single crystalline structure and open up entirely new possibilities for electrical and optical behaviour. This area of study is sometimes referred to as "Bandgap Engineering".

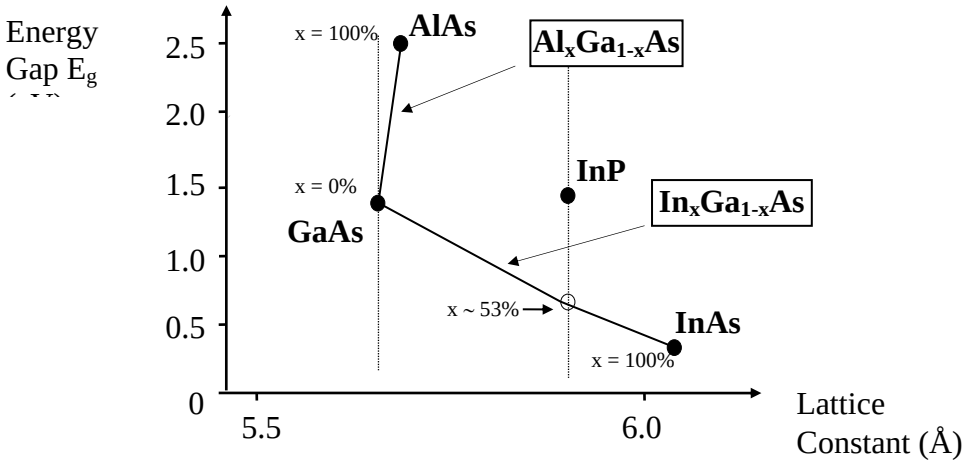

Fig. 5: Illustration of III-V Heterostructures and Lattice Matching

4. Oxidation:

In silicon processing, the growth of a layer of Silicon Dioxide (SiO_2) on the surface of a wafer is a common procedure. Usually this is done by "thermal oxidation" whereby the Si is heated to a very high temperature and exposed to steam in a quartz tube.

The chemical reaction is very simple, but note that some silicon is consumed in the process of creating the oxide:

$$Si + O_2 \rightarrow \mathbf{SiO_2}$$

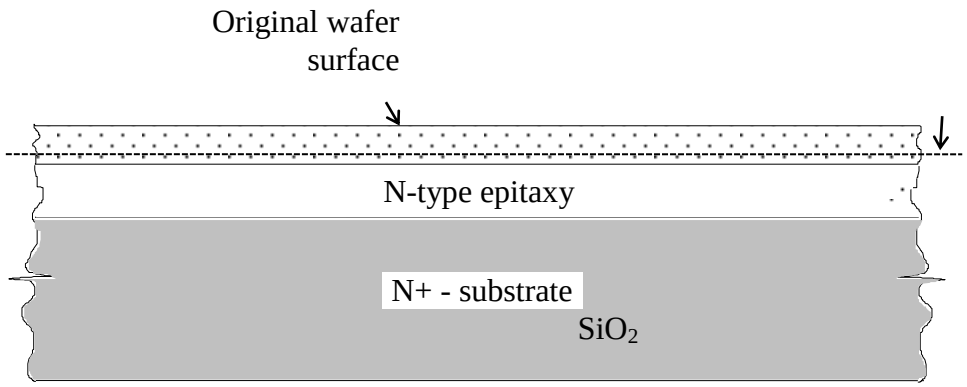

Fig. 6 Wafer Surface following Growth of Layer of Silicon Dioxide

SiO_2 is an excellent stable insulator or dielectric. It can be used as the dielectric in a capacitor or just to protect the surface of a device or IC from outside impurities etc. ("passivation"). Another widely-used passivation layer is silicon nitride (Si_3N_3) which is usually grown from chemical vapour deposition (CVD) and is even more impervious to impurities than SiO_2. Si_3N_3 can also be used as a dielectric/passive layer in GaAs IC's, as GaAs has no useful native oxide.

5. Lithography:

Up to now, the entire wafer surface has been implicitly involved in each of the processes described. The two basic processes of *lithography*, namely *(1) masking* and *(2) etching*, permit selective areas of the wafer surface to be acted upon in subsequent processing steps. Indeed, this is a critical procedure in the realisation of complete electronic circuits (i.e devices, interconnections and passive elements) within a single piece of silicon to create integrated circuits or IC's. However, similar operations are involved in producing a large number of discrete semiconductor devices simultaneously on a wafer. In a complex IC, the lithography cycle may be repeated 10 to 15 times (the "planar process"). Any lithography step (at least in optical lithography) requires the existence of a *mask* for that step. Masks often consist of glass plates with a metallic layer on one side (see Fig. 7).

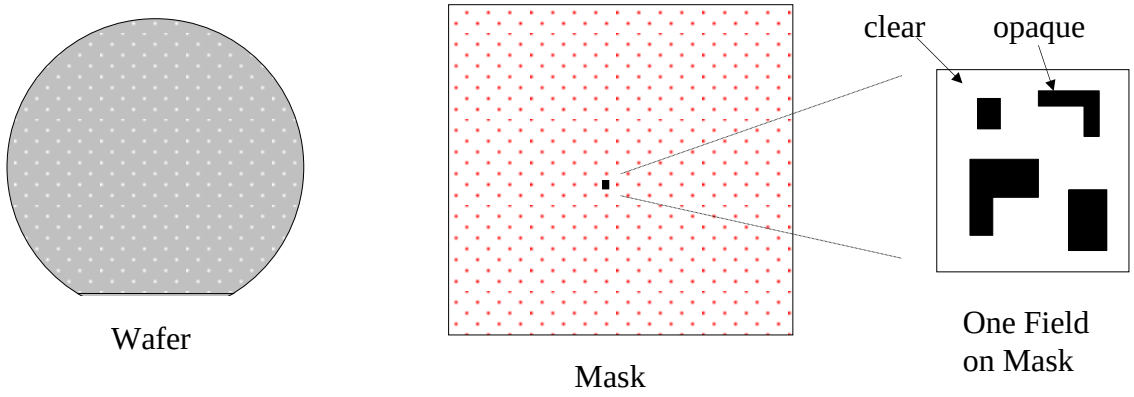

Fig. 7 Mask used in Optical Lithography to Transfer Repeated Patterns onto Wafer

A repetitive sequence of opaque metal rectangular patterns may be defined photographically on a mask, so that each mask consists of a large number of identical rectangular regions or fields (each such region corresponding eventually to one IC or one discrete device). Within each region there are polygonal sub-regions which are opaque while the remaining surface consists of clear glass. The process of "mask-making" is quite specialised, and is often carried out by companies who just do this job. The information for the making of the mask is a set of co-ordinates for each vertex in the pattern. This *layout* information is defined by the designer, working from a workstation and it must be emphasised that computer technology plays a critical role in this whole procedure (involving graphics editors, automatic placement and routing, design rule checking, magnetic tape production, etc.)

The individual regions are repeated over and over on the mask using a "step-and-repeat" camera or "stepper" (to do this a 10X master is required called a "reticle").

The purpose of lithography is to transfer the pattern of shapes on the mask onto the wafer. Suppose we just want to process a large number of individual simple polygonal regions on the wafer. The relevant field on the mask could look as follows:

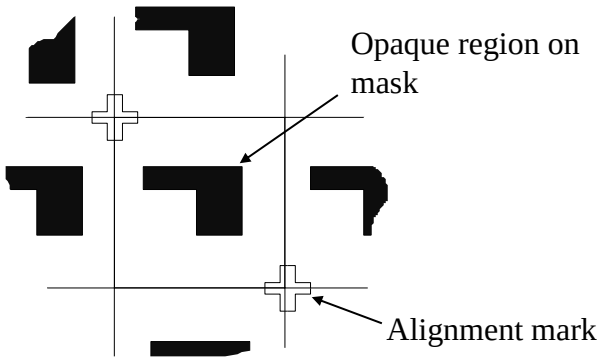

Fig. 8 Example of Simple Mask Pattern Repeated over Mask Surface

The procedure is as follows: grow an oxide layer over the entire wafer surface and then coat the oxide with a thin layer of photosensitive organic material called "photo resist" or PR. (This is done in a machine called a "spinner"). Position the mask close to the wafer surface and illuminate with UV light (Fig. 9).

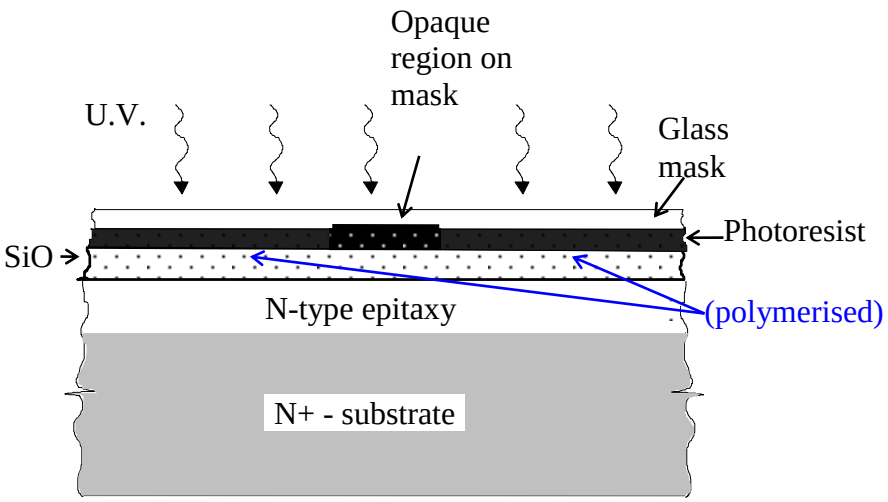

Fig. 9.: Example of Cross-Section Through Wafer Showing Epitaxial Layer

Suppose that the photoresist used is a so-called "negative resist" which has the property that it hardens by polymerising (i.e. forming long molecular chains) when exposed to UV. After exposure, when the mask is removed, an organic solvent may be used to dissolve away the unexposed resist beneath the opaque regions. The top of the wafer then looks as shown in Fig. 10. Positive resist would work in reverse.

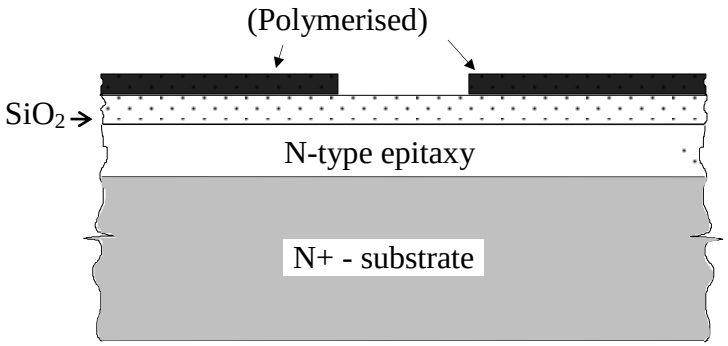

Fig. 10: Wafer Following Etching Away of Unexposed (Unpolymerised) Photoresist

The next step is to etch through the oxide using a wet etch (HCl) or a dry etch (plasma etching or RIE - reactive ion etching), the latter being the more modern approach. In either case, the polymerised resist protects the oxide beneath it. When the exposed SiO_2 is etched away, the resist is chemically stripped off and the wafer is cleaned. The result would be as shown in Fig. 11, for the case considered up to now:

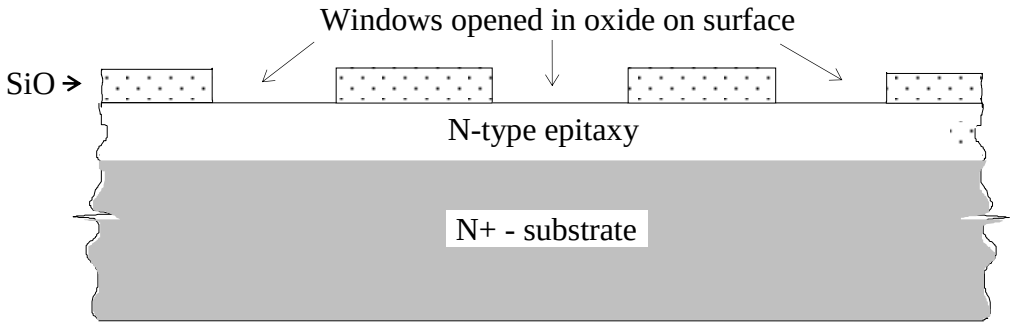

Fig. 11 Wafer Surface following Etching of Exposed Oxide and Photoresist Removal

The above describes a basic photolithography process but there are many variations. As dimensions in IC's are reduced it becomes necessary to reduce the wavelength of the UV (leading to so-called "deep UV" lithography, $\lambda < 0.3$ μm). For the finest dimensions (~0.1 μm) direct-write Electron Beam (or E-beam) systems are used to expose the resist directly under computer control. This is expensive and slow, however. Note, also, that the "photographic" nature of lithography processes require that the ambient light be closely controlled - typically filtered yellow light is used in clean rooms.

Indeed, the importance of *cleanliness* cannot be over-stated in semiconductor processing. This applies to the purity of the chemicals gases glassware and other materials used, which are required at state-of-the-art level for advanced

semiconductor processing. Also the environment used for the processing (especially the lithography) must be within very stringent standards of cleanliness: a vast infrastructure of air handling equipment, etc. is needed to service the clean rooms, where personnel (increasingly kept to the minimum necessary nowadays) move around in protective masks and clothing. One speck of dust on the surface of a wafer can ruin the masking operations for an entire circuit, thereby reducing the "yield", or percentage of working chips at the end of processing.

6. [Solid-State] Diffusion:

In discussing carrier transport in semiconductors, it is shown that free charge carriers diffuse if their distributions are non-uniform in space. At very high temperatures (~1000°C) it is possible for whole *atoms* in the crystal lattice to diffuse also. Specifically, this provides a method of introducing impurity atoms from the surface into an epitaxial layer. Often this is done as a "compensation" process, whereby the doping type of the epitaxial layer is reversed by the diffusion process over a certain region close to the surface. A diffusion furnace is usually used for this purpose, which consists of long cylindrical quartz tube surround by RF heaters with the temperature within the furnace being very closely controlled and uniform. The impurity atoms are introduced from a solid or liquid source on a carrier gas such as N_2 and passed over the hot Si wafers. The process is often performed in two sages:

(i) *pre-deposition* : a strong surface sheet of impurity atoms is created under constant source conditions;
(ii) *drive-in* : longer, high-temperature diffusion with the source switched off.

The mathematics of this process are covered in many text-books, and are based on solving a diffusion equation (Fick's Law). In fact, the drive-in phase above is similar to the Haynes-Shockley experiment for semiconductor charge-carriers with a delta-function initial condition, except there is no drift or recombination and diffusion is one-sided. Nevertheless, the resulting doping profile behaves similarly and is approximately Gaussian-shaped with distance variation beneath the surface.

Let us consider a very important example of a diffusion process applied to the structure discussed up to now - this process involves diffusing a high level of acceptor type impurity (P^+) into the N-type epitaxy. The result would be an example of a simple but very important semiconductor device, namely the PN JUNCTON. To be more precise, this case would correspond to a **P^+N junction**:

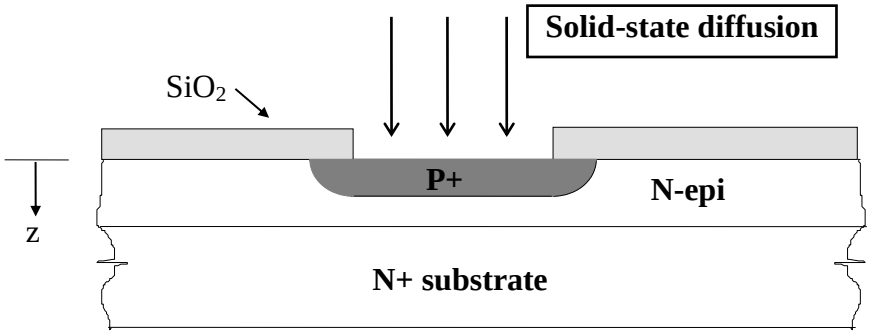

Fig. 12 Diffusion of P-type impurities through a window on N-type epi

The "doping profile" directly under the aperture might vary with z as shown in Fig. 13. It can be seen that the P+ diffusion 'compensates' the background N-type doping to create a P-type region near the surface. The location of the metallurgical junction in the PN structure is indicated by the vertical dashed line in Fig. 13.

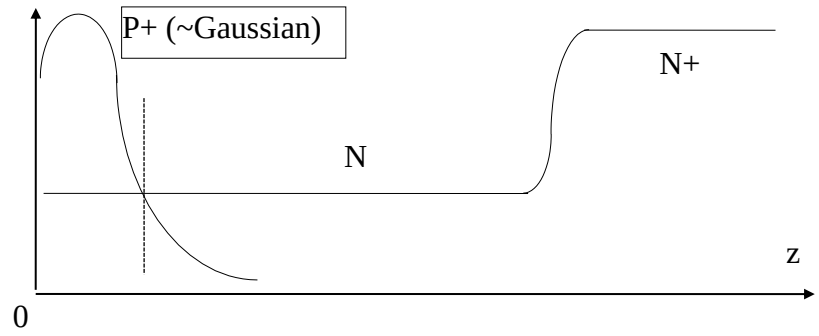

Fig. 13: Variation of Doping Concentration with Depth Below

Of course, if we are just using the wafer to produce PN junctions (as discrete devices), many thousands can be made simultaneously using the above processing steps.

Using different masking operations, several diffusions can be carried out. For example, Fig. 14 shows the use of a second 'deep diffusion', to produce a low-resistance path between the highly-doped substrate and the wafer surface. This would require a complete new cycle of masking and etching in the fabrication process.

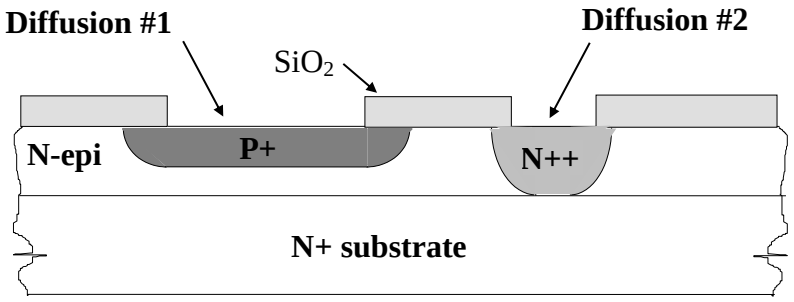

Fig. 14: Deep diffusion (#2) used to create surface contact to substrate

7. Ion Implantation:

Ion implantation (or 'I.I.') is an alternative or complementary technique to solid-state diffusion, which may be used to introduce impurity atoms into selected regions of the surface of a wafer. The starting point is similar to that shown in Fig. 10 and the principle of I.I. is then quite simple: accelerate impurity ions to high energy and fire them at the exposed surface of the wafer (ions may be blocked by PR). The ions lose kinetic energy through collisions with the fixed lattice sites and the electron cloud of the material until they come to rest at an average depth called the "projected range" up to ~1μm max. Ions distribute around this average location in a roughly Gaussian profile, although "channelling" through the crystal structure can be a problem. Ion implantation causes quite severe damage to the crystal near the surface, which is removed by heating the sample for some time to a high temperature ~600°C ("annealing"). The fraction of implanted ions that come active as dopants is called the "activation efficiency". Ion implantation equipment is large and expensive, but on the other hand, the technique offers a great deal of control over total implanted "dose" as well as good control of average depth. Furthermore, uniformity across the wafer is also very good.

8. Metallisation:

The final stages of device on IC fabrication involve one or more metallisation steps. One purpose of a metal layer is to make electrical contact to the semiconductor surface at specified locations, as depicted in Fig. 15 (usually these are required to be low-resistance or "ohmic contacts", although sometimes a metal semiconductor junction can be formed to act as a rectifying contact or Schottky diode). The second major use of metallisation layers is to provide electrical interconnection between parts of the circuit, by defining metal "tracks" running over the oxide layers. Indeed, in modern IC technology, a great deal of the effort goes into creating advanced multilevel metallisation structures, separated by oxide with connecting "vias". Up to about 8 separate metal layers are used for interconnect in the more advanced processes. The most common metal used in Si processing is Aluminium, as it is

stable, a good conductor and has good adhesion to Si. Recently copper interconnect has become more widely used. Sometimes more complex multi-layer metals are used e.g. Ti-Pl-Au, etc. for superior performance. There are two common metallistion techniques:

- **Evaporation:** is performed in a vacuum, and involves heating a sample of the metal near the wafer. Atoms of the metal vaporise and settle on the exposed semiconductor surface;

- **Spluttering:** an electrical (RF+DC) breakdown is produced in a low-pressure invert gas environment (typically Ar). One electrode is made of the metal, the other of the wafer. As the Ar ions strike the metal (cathode), metal atoms near the surface vaporise and travel to the wafer where they deposit. Sputtering gives lower contamination levels than evaporation, but is a slower process.

It should be noted that in practice metallisation is deposited on selected areas of the wafer, requiring further cycles of masking/etching.

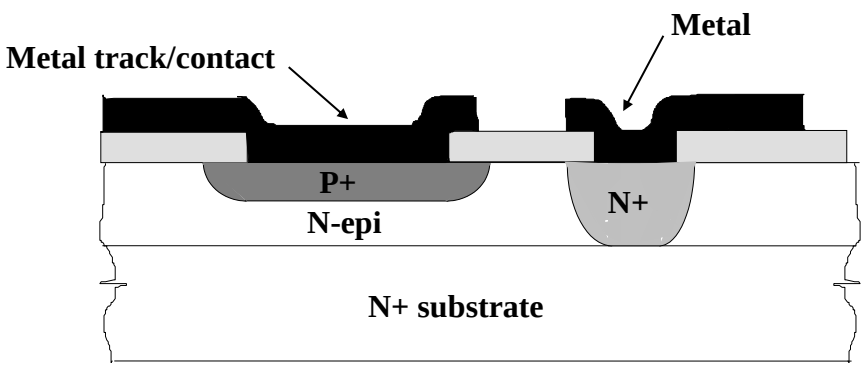

Fig. 15 Complete P^+N Junction device with (Ohmic) metal contacts

9. Conclusions:

We conclude with a few final comments on the last stages of IC fabrication (sometimes called BEOL or Back End of the Line, as opposed to the Front End of the Line processes that we have been considering so far). These are generally seen within the industry as non-critical operations and are commonly performed far away from the original processing location, often in developing countries where labour costs are low.

When the wafer has been fully processed, it looks as shown in Fig. 16. The wafer is sliced up into chips using a diamond saw. The individual chips are tested and marked with ink if defective. The development of good test procedures is a major problem in design. As noted earlier, the percentage of (working chips) over (total chips fabricated) is a vital parameter called the Yield. Using automatic equipment, the working chips are mounted on a metal carrier within a package. The package provides mechanical support and protection from the environment. At high

frequencies, however (such as used in mobile telephones), packages can have a seriously adverse effect on electrical behaviour. A further important role provided by the package is to assist in heat-sinking the chip. The package (nowadays mostly "surface-mount") has a series of metal pins or leads around the periphery. Automatic bonding machines attach Au wires from the pins to the bond pads on the chip. The package is then lidded and marked, before being distributed. Finally the finished and packaged chip finds its way into some piece of electronic equipment and makes its own small contribution to the vast €2200B-per-annum global electrical/electronics industry.

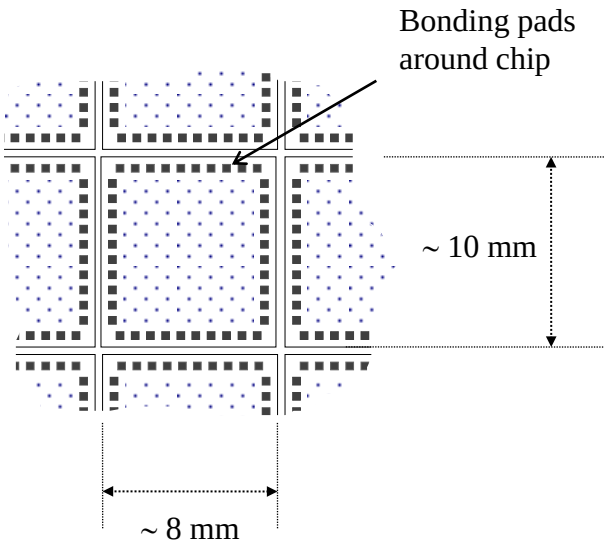

Fig.16 Sample of finished wafer surface showing individual Integrated Circuits