INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE - CENTRATOM TRIESTE

SMR/443 - 6

ICTP - INFN COURSE IN
"BASIC VLSI DESIGN TECHNIQUES"
6 November - 1 December 1989

Brief Resume of Aspects of

VLSI FABRICATION TECHNOLOGY AND SOME TRENDS TO THE FUTURE

Lecture No. 3

Mino GREEN
Department of Electrical Engineering
Imperial College of Science and Technology
Prince Consort Road
London SW7 2BZ
UK

Brief Resume of Aspects of

VLSI FABRICATION TECHNOLOGY AND SOME TRENDS TO THE FUTURE

Mino Green

Lecture No. 3

In this lecture we cover aspects of LAYERING and ETCHING.

III.1 GATE OXIDE

We start by considering one of the miracles of the silicon story, the fact that a carefully oxidised single crystal silicon surface is virtually devoid of interface states. Silicon when heated in dry oxygen or an oxygen/hydrogen mixture obtained from hot steam (pyrogenic steam) oxidizes to SiO_2

$$Si + O_2 \rightarrow SiO_2$$

The resultant oxide must be free of interfacial electronic trapping states (interface states, also called surface states) and have high dielectric strength. Only in this way can we have field effect transistors (MOSFETs).

The oxide thickness versus time is shown on Figures III.1 and III.2. For gate oxide we are concerned with thin oxide, less than 1000Å and almost invariably the {100} set of planes, this is because the electronic configuration is of the best quality as shown in Fig. III.3. Now the question must be asked, "what is the relation between interface state electronic configuration (density and energy) and atomic configuration?" In truth we do not, even now, really know the full answer, but we know a great deal that is useful.

All three low index planes ({100}, {110} and {111}) are terminated with dangling silicon bonds. These bonds give rise to donor/acceptor states and tend to give a p-type surface. On oxidation these dangling bonds are mostly replaced by silicon-oxygen bonds which push the electronic energy levels far out of the silicon energy gap. Thus it is reasonable to conclude that any dangling silicon bonds which were not eliminated in the oxidation to SiO_2 will remain as interface states. So that the more perfect the match between Si and SiO_2 the lower should be the interface state density. This has been the accepted wisdom for a long time, but there was always a bit of difficulty in conceiving a good atomic fit of amorphous - SiO_2 (a-SiO_2) to the silicon crystal surface. However, recent studies of Ourmazd and Reutschler (Appl Phys Lett 53 (1988) 743) have shown that the first 7Å of SiO_2 in contact with {100} Si are a perfect crystalline fit, as shown in Fig. III.4 (from their paper). Thus the oxide is a structure going as: c-Si, c-SiO_2,

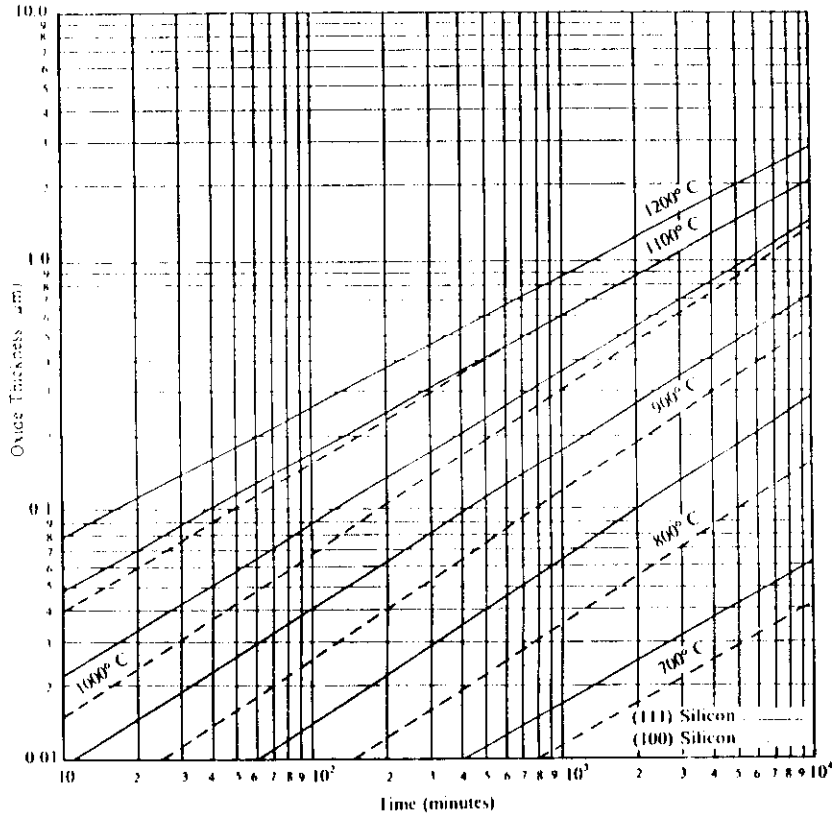


FIG III.1. Oxide thickness versus oxidation time for silicon in dry O_2.

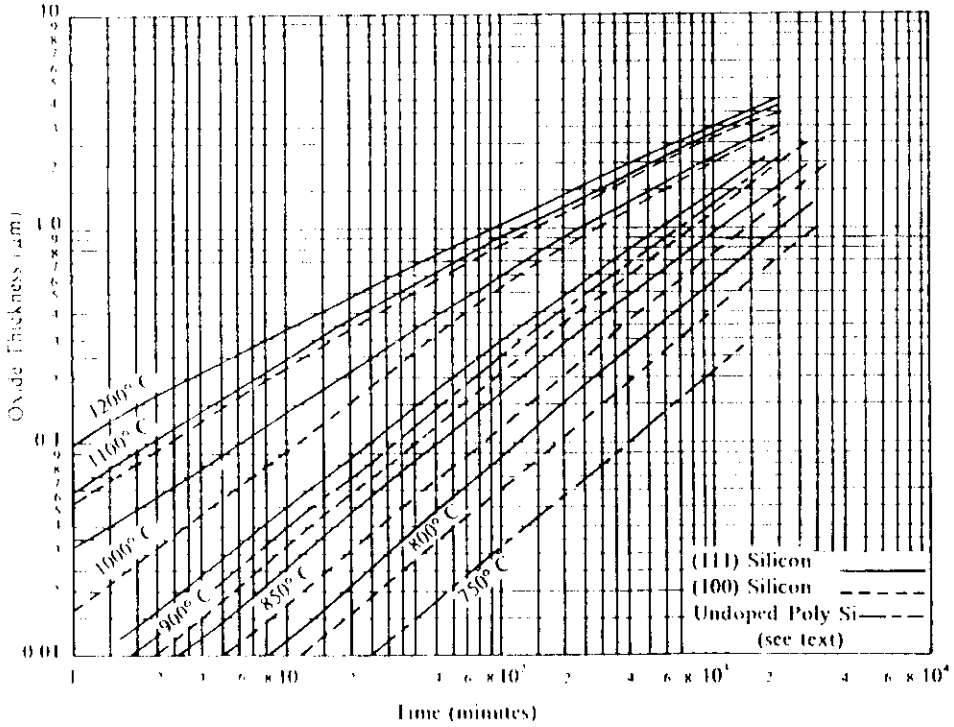


FIG. III 2 Oxide thickness versus oxidation time for silicon in pyrogenic steam (640 Torr).

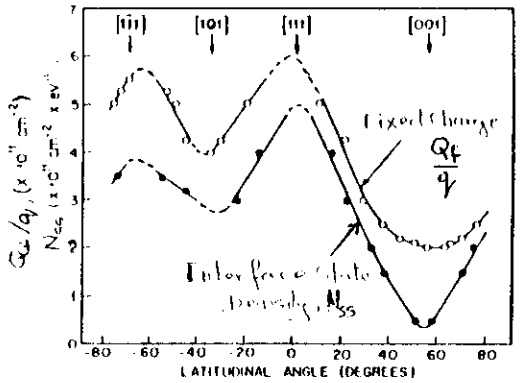


FIG III 3 Oxide fixed charge density and interface trap density as a function of the silicon orientation.

FIG III 4. SiO_2 on Si Specs

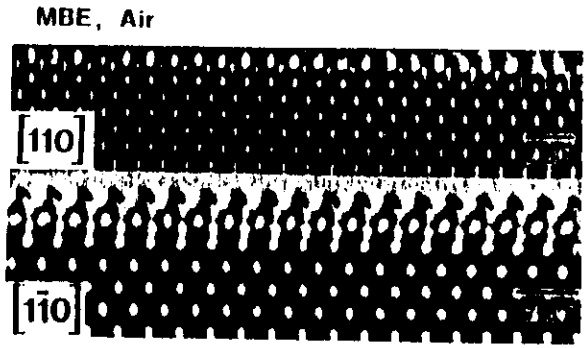


FIG. 1. Lattice images of the (001)Si/native oxide interface in the [110] and [1$\bar{1}$0] projections. Note the atomically flat interface and the interfacial c-SiO_2 phase with 3.8 Å periodicity.

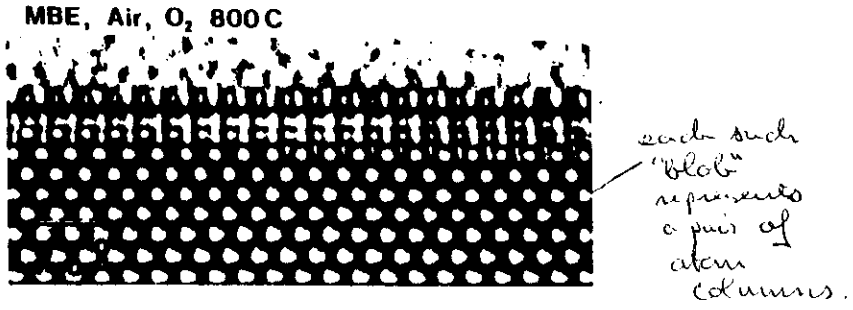


FIG. 2. [1$\bar{1}$0] lattice image of the Si/native oxide sample annealed in O_2 at 800 °C.

a-SiO_2.

It is quite clear that in making the gate oxide considerable care must be exercised. This presents certain problems for fast throughput VLSI fabrication. Thus the oxidation is carried out as batch processes in tube furnaces. The wafers are mounted on silica glass carriers and contained in silica glass tubes. The oxide thickness is measured by ellipsometry. The single most important parameter produced by oxidation is the gate oxide thickness and it must be monitored with precision. There is, however, a tendency, as wafer diameters are forever increasing, to favour single wafer processing, rather than batch processing. But oxidation is, in general, a slow process and is not given to this concept. In order to speed up the oxidation process (at least for the thicker oxides mentioned next) research workers are investigating the process of plasma enhanced oxidation, in which oxygen is made into a plasma in a radio frequency discharge. High pressure oxidation is also being investigated. The future trend does, however, still appear to be with batch furnace processing with the new direction being into fully (robotic) automatic handling in order to avoid particulate contamination associated with humans. This leads us on to consider chemical vapour deposition, CVD.

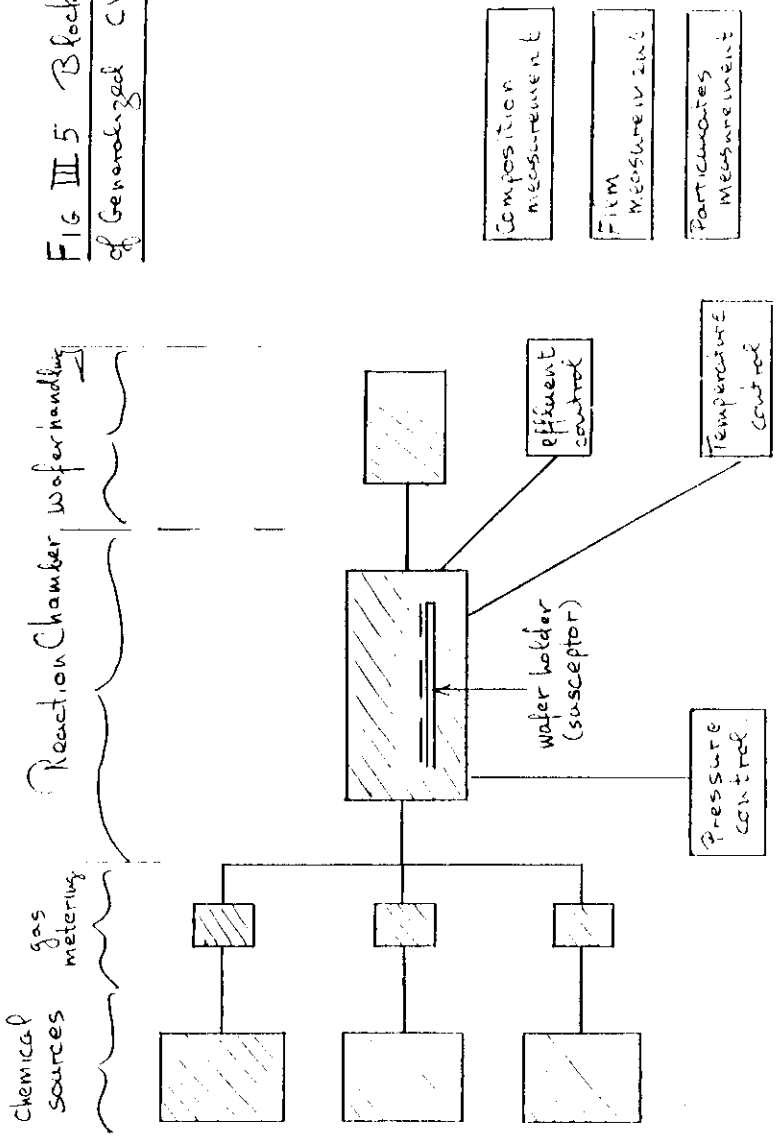
III.2 CVD

Chemical vapour deposition is a process where one or more gaseous species react upon a solid surface and one of the products of that reaction remains on the surface as a solid product. CVD is extensively used in VLSI fabrication technology.

The basic physical layout of the CVD equipment is shown in Fig. III.5. There is a gas handling system, either at atmospheric pressure or low pressure, in which various gases, each under mass flow control, can be fed into the reaction zone. There is a reaction chamber in which the CVD reaction takes place. There are arrangements for introducing the energy flow which will cause the chemical reaction. These may involve heating the chamber walls (hot wall reactor), or heating the wafer only (susceptor heating, cold wall reactor), or sticking a plasma, or introducing light. There must be control of pressure, flow, composition, purity, particulates ad measurement of CVD film properties, e.g. thickness. The spent gases, from the CVD reactor, must be disposed of safely. Finally, it must be possible to load and unload wafers without contaminating or harming them in any way.

The reaction kinetics are specific to each kind of chemical film, but it is true in general that the reaction will either be controlled by the reaction at the surface of the solid, in which case temperature control is of paramount importance, or the process will be controlled by diffusional transport through the gas phase of reactants to the wafer surface. In the latter case, considerations based upon aspects of fluid mechanics (gas dynamics) are of importance; approximate models are considered below. In particular, we have the concept of a boundary layer. When a fluid (gas or liquid) flows adjacent to a

FIG III.5 Block diagram of Generalized CVD system



solid surface, the velocity of the fluid at the surface must be zero. The region in which the fluid velocity changes from its normal value to the zero value on the surface is referred to as a boundary layer (see Figure III.6). For high velocity flows, the thickness of this transition region, δ, can be quite small and is approximately proportional to the inverse of the square root of the Reynold's number, viz,

$$\delta = \sqrt{\frac{x}{R_e}} \qquad (1)$$

where: $R_e = \frac{\rho u x}{\mu}$, ρ = mass density, u = flow velocity, x = distance in flow direction, μ = viscosity.

The quantity μ/ρ is called the kinematic viscosity and is often a source of confusion. It is found that when $R_e \geq 1200$ the flow is laminar and when $R_e \geq 2200$ the flow is turbulent, and between these quantities is ill defined.

Typical flow in a CVD reactor at atmospheric pressure (AP), i.e. in an APCVD system, is at Reynolds numbers of 100's, while low pressure (LP) reactors, LPCVD, function at Reynolds numbers typically 1-10: the mass throughput of the higher pressure system is considerably larger. In both classes of reactor the flow is laminar. If gas is entering the reaction chamber from a narrow pipe, it will require some distance from the entrance point, over which to settle down to stable flow, which is a parabolic velocity profile. Such a characteristic length to achieve "developed flow" is given by

$$L_e = 0.03\, dR_e \qquad (2)$$

where d is the diameter of the reaction tube.

Thus in the case of diffusion controlled CVD reactions the boundary through which we are diffusing is δ thick and the flow, j, to the surface will be given, for iso-thermal conditions by Fick's first law, namely:

$$j = D(c_G - c_S)/\delta \qquad (3)$$

Here D is the diffusion coefficient, typical diffusion coefficients at standard temperature and pressure are 0.1 to 1 cm^2/sec. Mostly c_S, the surface concentration is effectively zero.

An example of a typical cold wall reactor is shown in Figure III.7 and low pressure hot wall reactor in Figure III.8.

Layers which are put down by CVD in the VLSI programme are:

1. Interlayer insulating SiO_2 (CVD oxide) made in low, medium and high temperature processes. Thus

   300 - 450°C:-
   $SiH_4 + O_2 = SiO_2 + 2H_2$
   650 - 750°C:-
   $Si(OEt)_4 = SiO_2 + 2CO + H_2 + 3C_2H_6$
   Ca. 900°C:-
   $SiCl_2H_2 + 2N_2O = SiO_2 + 2N_2 + 2HCl$

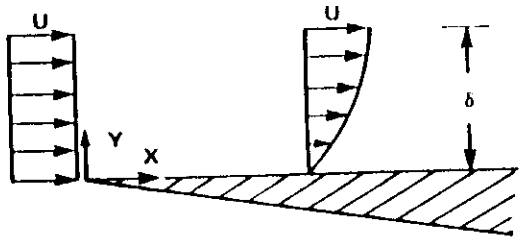


Fig III.6: Fluid dynamic boundary layer.

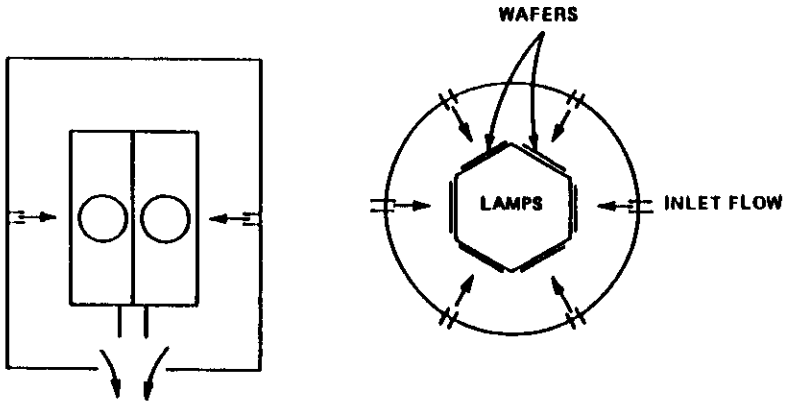


Fig III 7: Bell jar reactor, barrel susceptor, radial flow.

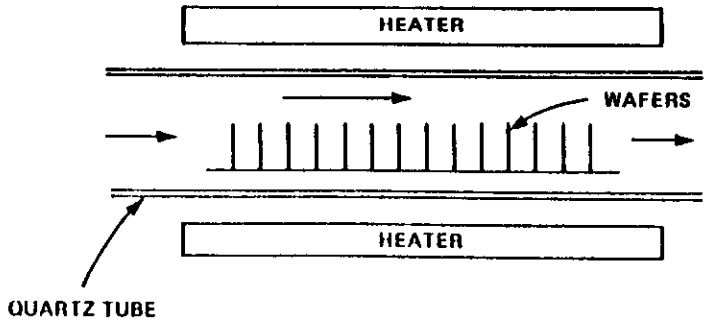


Fig III 8 : Low pressure hot wall CVD reactor.

2. Glasses, for encapsulation capable of high temperature flow in order to have flat surfaces. Also usable as an interlayer dielectric.
   ca 450°C
   SiH_4 : BCl_3 : PH_3 : O_2 to give borophospho-silicate glasses (BPSG)

3. Silicon Nitride, used as the oxidation coating during the formation of field oxide
   700-800°C
   $3\ SiCl_2H_2 + 4NH_3 = Si_3N_4 + 6HCl + 6H_2$

4. Episilicon. Vital step in CMOS, several possible reactions. Typical: about 1200°C and atmospheric pressure of a 1000:1, H_2:$SiCl_4$ mixture one gets, single crystal material
   $SiCl_4 + 2H_2 = Si + 4HCl$

5. Polysilicon. This is obtained by the above reaction at typically 800°C to 900°C.

6. Silicides, e.g. WSi_2, see later.

III.3 METALLIZATION

Metallization (a term universally used) means patterned metal thin film lines of electrical interconnection: Figs. I.3 and I.14 help to illustrate the term. On I.3 is shown the metal rail and contact cut (vias), there are connections from n^+ source and drain regions to other devices and ultimately to the outside through bonding pads. Metal contacts are also made to the gate poly as shown on Fig. I.4. As described, the nMOS process has one metallization step, but in fact the demands of VLSI have resulted in the need for several levels of metallization. Thus one level of metal patterning and interconnection is made and then insulated from the next level of metal pattern and interconnection by CVD silicon dioxide. It is only in this way that a high device packing density can be achieved.

The metal most widely used in metallization has been, and is, aluminium or a dilute aluminium alloy. The electrical conductivity of aluminium thin films approach to within 90 to 95% of the bulk value, which is 3.7×10^5 $(ohm\ cm)^{-1}$. The alloy conductivity values are about 30% higher. Aluminium (or alloy) forms ohmic contacts to silicon (discussed in the next subsection). The adhesion of aluminium, which is always tracking over SiO_2 or BPSG, is excellent because on heat treatment there is a limited exchange of silicon and aluminium via the reaction,

$$3SiO_2 + 4Al = 2Al_2O_3 + 3Si \text{ (dissolved in Al)}$$

Thus aluminium satisfies the need for high conductivity and good adhesion, it is the other requirements of uniform deposition, current carrying capacity and corrosion resistance which are considered next.

Aluminium is usually deposited by thermal evaporation in a vacuum chamber of several hundred litres volume at a base pressure of about 10^{-6} torr. Alternatively the

metal can be deposited by a sputtering technique which uses positively charged energetic argon ions to knock aluminium atoms into the gas phase (this is discussed in section V). Thin film physical deposition is often labelled PVD, in contrast to CVD. In the case of thermal evaporation the material for the films produced comes from a small area source, in line-of-sight to the wafer. Since, at room temperature, the Al is not mobile over the surface, the coverage is not conformal. This is an important problem since at this stage in the process (Mask 7 in nMOS) the topography of the processed wafer presents many steep steps. A solution to this problem is to have multiple sources. Another solution is to use sputter deposition. A technique much favoured is to evaporate onto the hot wafer which is held at 300°C, when the Al is mobile over the surface: the worst effects of non-conformal behaviour are thereby avoided. This is a major problem as vias become smaller.

Aluminium has at least three physical failure mechanisms, two of which, thermomigration and electromigration, become more serious with decreasing feature size. A major drawback to aluminium and, to a lesser extent, its alloys, is the ease of hillock formation. The difference in thermal expansion coefficient between Al (25×10^{-6}/°C) and the substrate Si (2.6×10^{-6}/°C) at least ten times smaller, can lead to compressive stress when the films are heated after deposition; say on deposition of an interlayer dielectric (450°C). The stress tends to be relieved by the formation of hillocks of aluminium, as well as regions voided of aluminium. Hillocks, which can be as high as 2μm, can cause interlayer shorts, and variations in the Al can give rise to difficulties with the lithography. It has been reported that the addition of 1 atomic per cent copper to aluminium films adequately suppresses hillock formation for multilevel interconnections. Also the resistivity of the alloy is 3.3μohm cm, which is well within range of that of pure Al.

An increase in current density leads to a rise in line temperature caused by Joule heating. Imperfections in the line may give rise to hot spots. Since Al migrates very readily at elevated temperatures, the possibility of line rupture by this mechanism is ever present.

The main failure mechanism for aluminium lines is the process of electromigration. This is the transport of aluminium under the influence of a d.c. electric current of typical density 10^6 Acm^{-2}. The transport eventually leads to line rupture. The process is suppressed by alloying the aluminium and producing a microstructure where grain size and line width are comparable. The mechanism of electromigration is still somewhat obscure, but it is widely taken that atoms are displaced as a direct result of electron momentum transfer (the atoms move in the same direction as the electrons).

Consider the positively charged metallic ions sitting in lattice positions. In metals, of course, the conduction electron density is quite high. Current within the conductor is carried predominantly by these electrons, and these electrons collide frequently with the lattice ions. Though the metallic ion finds itself in an energy minimum as it resides in its lattice site, the ion may spontaneously gain enough energy from either the thermal energy of the lattice or the electron collisions to vacate the lattice site and occupy an intersititial position. Electron-ion scattering increases greatly when the ion occupies such an interstitial position. This scattering transfers momentum from the electrons to the ion and thus exerts a force on the ion in the direction of electron flow. Thus, there develops a net transport of positively charged ions in the direction of the most positive potential. Certainly there exists an opposing force due to the conductor electric field on the positive ion. This Couloumb force is generally small as one might infer from the observation that an ideal conductor would exhibit zero resistance and thus zero electric field while the "electron wind" force would presist. Electron screening further reduces the effective Coulomb force. Ironically, then, one will find net transport of the metallic ions in the same direction as the electron flow in metals in which conduction electrons dominate current flow.

Electromigration leads to integrated circuit failure. An ion leaving a lattice site leaves behind a vacancy. A neighbouring ion "upwind" of the vacancy may jump to the vacated lattice site with far greater probability than a neighbouring ion "downwind" of the vacancy. The result is a net transport of ions in the direction of electron flow and an equivalent transport of vacancies in the opposite direction. The excess ions tend to coalesce to form hillocks and extrusions, as do the vacancies. The vacancy coalescence simply leads to macroscopic gaps in the metal film, and if uninterrupted, this gap will grow to the entire cross section of the film and thus "open" the metal line. The net effect of electromigration, then, is to redistribute the metallization into regions of excess metal (hillocks and extrusions) and depleted metal (voids). The excess metal may also lead to circuit failure. Extrusions will grow and bridge to another metallization line. This unintended short circuit renders proper circuit function impossible. Clearly, the short-circuit lifetime will depend on the closest separation of neighbouring metallization lines as well as the physical dimensions and current density of the primary line. The extrusion length does not scale! On the other hand, one may also note that the critical size a macroscopic void must attain for the open-circuit failure mode decreases as circuit dimensions decrease. Thus, one should expect electromigration degradation more quickly than indicated merely by increased conductor current density.

III.4 SILICIDES AND TUNGSTEN

The scaling down of features to the micron and sub-micron level means that polysilicon, with its sheet resistivity of 15-30 ohms per square, is too resistive for device purposes; that the current densities (ca 10^6Acm^{-2}) in aluminium interconnects give cause for concern, because of electromigration effects; and that the filling of small vias is presenting fabricational problems. In addition to these scaling-induced difficulties, there is the need to provide for multilevel metallization (interconnects) so that high device packing density can be achieved, ie long interconnect lines are avoided. The high conductivity and high temperature properties of various metal silicides and refractory metals (see Table III.1) has

offered a way to overcome these problems and some possibilities are discussed here.

Table III.1 **Properties of Some Refractory Metals and their Silicides**

Substance	As deposited	Annealed	Deposition Method
Mo	7-20	-	CVD $MoCl_5$
	22	11	DC Magnatron sputtering
W	8-16	10-12	CVD WF_6
	29	21	DC Magnatron sputtering
$TiSi_2$	350	15-20	Plasma enhanced CVD: $TiCl_4/SiH_4$
	610	25	Co-sputtered
$MoSi_2$	1200	100	Co-sputtered
$TaSi_2$	> 125	48	CVD : $TiCl_5/SiH_4$
	190	50-55	Co-sputtered
WSi_2	600-900	35-60	CVD : WF_6/SiH_4
	460	70	Co-sputtered

We will discuss the gate structure known as the POLYCIDE structure. This structure, shown on Fig. III.9, retains the satisfactory and reliable poly-Si/SiO_2 interface, while at the same time making use of the low silicide sheet resistance of WSi_2. The manufacturing process is as follows below:

a) Poly is laid down (about 2000Å) by CVD (~600°C) annealed at 1000°C, and doped (by ion-implantation or chemical methods).

b) Silicon rich WSi_{2+y} is deposited on the poly in LPCVD cold-wall reactor using SiH_4 - WF_6 - He gas mixtures. The substrates are rear heated using infrared lamps. The reaction temperature is 350-450°C. Total gas pressure 50-300 mTorr. Flow rates are 1-20 sccm. The silicide growth rate is 500-1000Å/min and is dependent on the WF_6 pressure and flow rate. The deposited film is annealed (900 to 1000°C) giving a resistivity of 50-60μ ohm cm (poly is at best 2 x 10^{-3} ohm cm).

c) The gate structure is delineated - mask 5 in the nMOS process. The S/D regions are made by ion-implantation using the self alignment of the gate.

d) The gate and exposed active area are thermally oxidized. The WSi_{2+y} oxidizes as does the silicon: this passivates the gate. The process is a high temperature, 850-1000°C, steam oxidation, and during processing it is possible that P will be lost from the poly, and that B or As, which may have been used in S/D doping (CMOS process) will also be lost from the gate poly.

The resulting devices are as good as pure poly gate devices and better by virtue of the increased gate conductivity.

III.5 REACTIVE ION ETCHING

Etching is an essential part of the patterning process, as well as being an important treatment in wafer preparation. Here we consider reactive ion etching RIE which is extensively used in VLSI processing.

There is an entire field of physics called gas discharges and it is most certainly not the intention here to summarise that subject. Our concern is with radio frequency, rf, discharges in gaseous systems. Here a plasma is a low pressure electrically neutral ionized gas, containing an appreciable concentration of free electrons and ions. The plasma in steady state requires a continuous energy input, since there is continuous energy loss due to electron-ion recombination as well as other decay processes going on.

Suppose we have a chamber containing a gas at low pressure, say 0.1 torr, and a pair of parallel plate electrodes in that chamber across which is an rf field. There will be some ionized species present in the gas (a passing cosmic ray may do the job) which will be accelerated in the field and which will cause ionization of other neutral species. A broad range of chemical and physical processes will occur and a steady state (rf) plasma

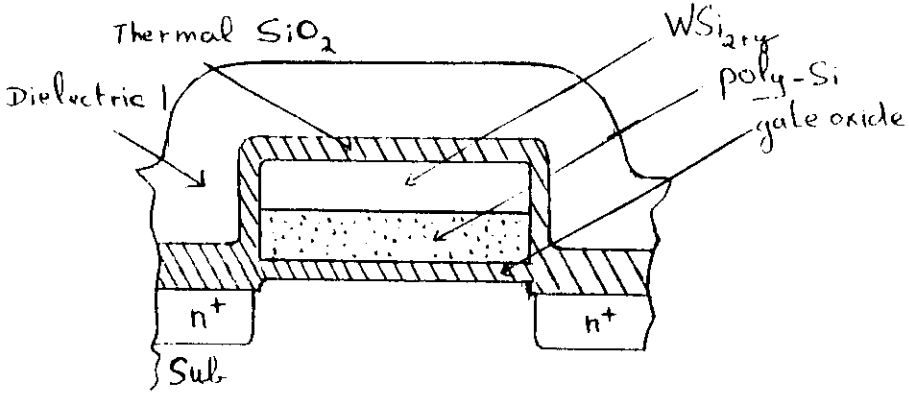


FIG III.9. POLYCIDE STRUCTURE

will be built up. The rf plasma (which will emit visible light and is called a glow discharge) consists of electrons and ground state and excited molecular and atomic species which might be neutral or charged. The 0.1 torr gas would contain 3×10^{15} molecules per cc. The density of charged atomic and molecular species would be about 10^9 per cc. The concentration of reactive neutrals (radicals will be considerably higher, e.g 10_{13} to 10_{14} per cc. The average electron energy will be 1–10eV, which corresponds to temperatures of 12,000K and upwards. The atoms and molecules on the other hand have an energy distribution which corresponds to room temperature. Table III.2 lists some RIE schemes.

The rf discharge in the parallel plate configuration depicted in Fig. III.10 has an interesting potential distribution. It is assumed that the rf generator produces across the plates a sinusoidal voltage output of amplitude V_o that is sufficiently large to sustain the gas discharge. The mobilities of electrons and ions are vastly different so that electrodes became negatively charged with respect to the plasma in order to equalize the fluxes of species to the electrode surfaces. The result of decreasing the electron flux is a decrease in electron concentration near the surface and hence a decrease in the photoemission reactions, which leads to a dark space in front of the electrode. A dark space is called a plasma sheath. In a capacitively driven plasma the major fraction of the potential difference falls across the sheath and only a minor fraction across the glowing, highly conducting plasma. **The plasma potential is always positive with respect to the electrodes.**

When the areas of the two electrodes are not equal then these voltage distributions become unsymmetrical and are determined by the relative capacitances of the electrodes. A situation with unsymmetrical electrodes and an added capacitor in the powered electrode is shown on Fig. III.10.

This unsymmetrical configuration confers a very considerable effective d.c. bias on the powered electrode, making it into a configuration known as a Reactive Ion Etcher (RIE). The self d.c. bias varies up to 200–300 volts.

The acceleration of ions across the sheath in a direction normal to the wafer surface makes the etching process highly anisotropic. Ion directionality means that surfaces parallel to the wafer surface get etched faster than surfaces which are inclined to the ion direction. This directinality is important in retaining feature size control. The mechanism of etching is basically a twofold process. Firstly, positive ions from the plasma, they can be any nuclei since chemistry is not invoked here. Strike the exposed layer on the wafer surface and cause considerable disruption to the lattice for a depth of several inter atomic distances. These disrupted atoms have been made chemically reactive – broken bonds, displaced ions, wrong neighbours etc. – they then react with the specific free radicals present in the plasma. Here chemistry is everything, the radicals in the gas phase must be reactive towards the layer material and inactive towards the resist. Also the reaction products must be volatile and in no way inhibit the etching process.

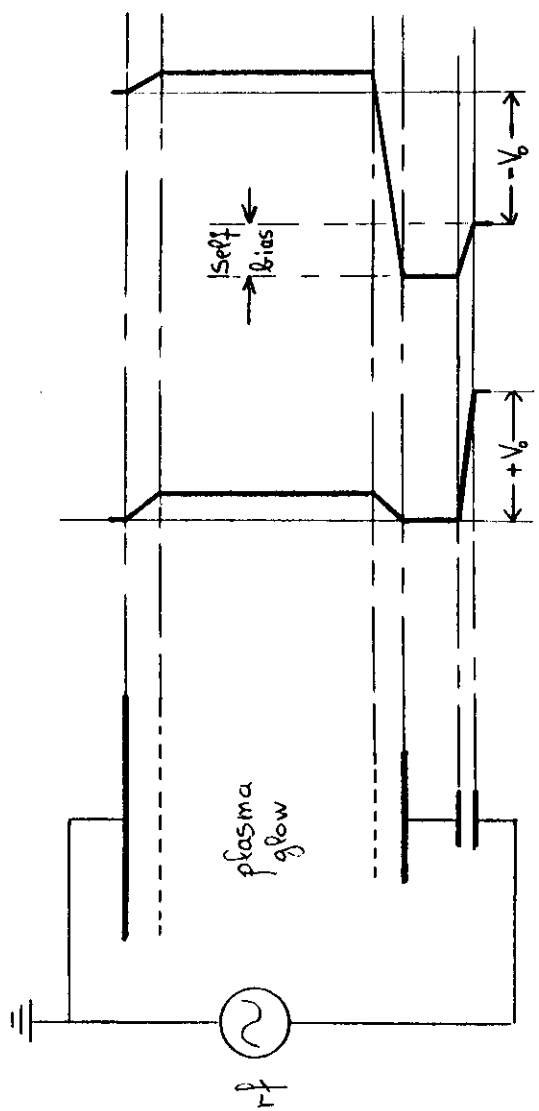


FIG III 10 Asymmetric capacitively coupled reactor. The blocking capacitor increases the self bias on the negative part of the cycle. The wafers are placed on the smaller powered electrode.

Table III.2 : RIE Etching of Various Materials

Etched Material	Feed Gas	Etch Rate (Å/min)	Selectivity M/resist	M/Si
Si	CF_4/O_2	3000 (at 100°C)		
Poly	Cl_2	500-800	5	
SiO_2	CF_4/H_2	500	5	20
PSG	CF_4/H_2	800	8	32
Si_3N_4	C_2F_6	1200	-	8
Al	BCl_3/Cl_2	500	5-8	3-5