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

IN REPLY PLEASE REFER TO

SMR/443 - 4

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. 1

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. 1

I.1 INTRODUCTION

These lectures are a fairly distant overview of one of the most advanced and exciting technologies that the world has to offer today, namely the Very Large Scale Integration of solid state electronic devices and circuits. We all need an overview – manager and technologist alike – but I would be wrong if I failed to remind you that success or failure of fabrication technique usually depends upon paying scrupulous attention to detail, most of which is absent from this brief course!

Present VLSI technology, on SILICON, involves circuits with 10^5 to 10^6 transistors carried out to design rules of about 1.2 microns. Medium term aims are targetted to about 5×10^6 transistors at 0.5μm. And the mid-1990s should see us down to design rules of 0.3μm and $> 10^7$ gates: this would be called ULSI, U for Ultra.

I.2 OUTLINE OF FABRICATION TECHNOLOGY

We are going to restrict ourselves, at first, to considering silicon and MOS transistors. The basic single most important circuit element is the INVERTER, and this will be used in order to illustrate the fabrication process. MOS transistors are localized layer structures on a silicon single crystal wafer. How to make such structures and interconnect them into circuits with reasonable device yield is what fabrication technology is largely about. The fabrication process is one of making various thin layers of material (LAYERING), making patterns on and through these layers (PATTERNING) and doping selected areas which are exposed by virtue of the patterning (DOPING); these basic steps are illustrated in Fig. I.1. There are a number of sequences of these steps, and finally the wafers have to be protected, cut up, mounted and tested.

An idea of the sequence of steps, in simplified form, can be gained by considering the nMOS process outlined below. But just before we go through that outline, a very brief discussion of the basic steps of layering, patterning and doping is necessary.

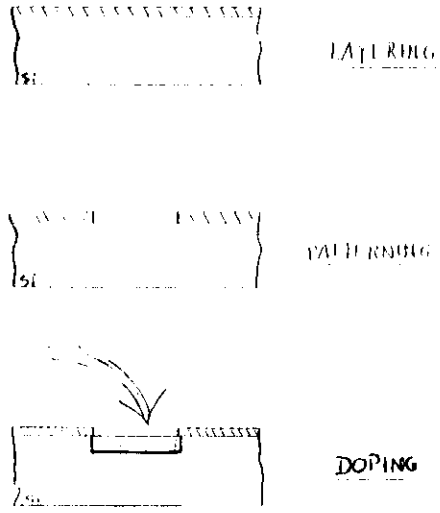


FIG I.1. THE BASIC FABRICATION STEPS

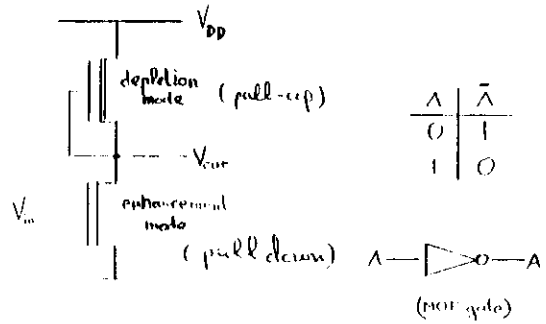


FIG I.2 The basic inverter circuit diagram: logic symbol and function shown

a) Layering

There are a number of different types of layers all of which, except the final coating layer, have to be patterned. There are four types of oxide layer. There is the thin (ca 250–500Å) gate oxide, which forms the insulator of the transistors and this film has to be treated with great care. There is the field oxide which separates interconnections between transistors from the underlying substrate with a relatively thick oxide (hence low capacitance). There is the chemical vapour deposition (CVD) oxide which serves to separate another level of interconnections. Then there is the overlayer of oxide which protects the structure. There are also other insulators like silicon nitride, Si_3N_4, which acts as an oxidation mask.

Other types of layers are heavily doped polycrystalline silicon (called "poly" or "poly-silicon") which acts as an interconnection (a wire) and gate material; and there are conductors like the metals aluminium and tungsten. Finally there is resist (or photo resist as it is also called) which is the thin layer in which patterns are created.

b) Patterning

This process is called lithography – which literally means an inscription on stone and as used here is a bit far fetched. For the sake of the justification of the word we take the silicon to be stone and the circuit to be the inscription. We will wish to dope certain areas of the silicon. The location of these areas is done by coating the layered silicon with a material called, in general, a resist. This material is exposed to radiation through a mask which gives the desired pattern. Say, for example, that the resist is an organic chemical which becomes water soluble on exposure to UV light. The pattern of solubilized polymer is washed away (this is "developing") and we now have a pattern of "windows" in the polymer. Now comes the step which explains why the material is called "resist". The polymer and exposed layer is subjected to a chemical etch (which may be wet or dry); the polymer is resistant to the etch but the oxide is etched away exposing the underlying silicon. At this stage etching is stopped, the polymer is stripped off and the silicon is doped through the windows in the oxide. Thus patterning is a two stage process, involving lithography and etching.

d) Doping

As just mentioned, doping is carried out through the opened windows. The process of doping can be by thermal indiffusion processes or by ion-implantation, followed by annealing.

We start with, say, a four or six inch diameter wafer of perfect single crystal silicon doped p-type (say, 3×10^{15} boron atoms cm^{-3}, 5 ohm cm). Only the basic steps will be given below, many operations of washing, etching, etc., which are not material to immediate understanding, will be omitted.

It will help if we have a particular device in mind as we go through the nMOS process. A depletion load inverter has been chosen. This consists of an enhancement mode transistor in series with a depletion mode transistor. The circuit is shown in Figure 1.2.

The mask layout for the depletion load inverter in nMOS technology is shown in Fig. 1.3. The design rules, which are of numerous types, might follow those laid out later on.

The first layer consists of a thin, thermally grown, 200Å film of SiO_2 (called the buffer or stress relief oxide layer) with a layer (typically 800Å thick) of silicon nitride, Si_3N_4, on top of it. Over this layer is added a layer of photo resist in which the first mask pattern is projected (this is called the ACTIVE AREA MASK and sometimes it is called the n^+ mask). Windows are etched in the oxide/nitride layer and boron atoms are implanted (doping) in these opened areas, see Fig. 1.4(a). We consider the complete mask sequence using only positive photo resist. The photo resist is stripped, the implanted areas flash annealed so as to render the boron active, making the areas p^+. The pattern underneath the Si_3N_4 will eventually contain within it all the active areas, i.e. all the areas containing transistors and MOS capacitors and buried contacts, and that is why we call it the n^+ mask. At this point the state of affairs is as shown on Fig. 1.4(a), the mask, and Fig. 1.4(b), which is a section view through AA'.

The second layering operation, which is the local oxidation process (LOCOS) involves placing the structure (Fig. 1.4(a) and (b)) in an oxidation system and growing the field oxide to about $\frac{1}{2}\mu m$ thickness, the nitride acting as a barrier to oxidation in the active area. Since this is a high temperature process there is some diffusion of the boron implanted acceptors and a spreading of the p^+ region, as indicated on Figure 1.5. The nitride, the buffer oxide and a small amount of the field oxide are etched away and a new thin oxide, ~ 250Å to 500Å, grown: this is the vital gate oxide (upper limit of electric field reliability on gate oxide is 2MV/cm).

Now follows a set of three ion-implantations, employing three mask steps. These implants are the enhancement mode transistor implant (just called the enhancement implant): photo resist, suitably patterned and developed, is used to delineate the required area, cf. Fig. 1.6. There is a self-aligning due to field oxide. A similar patterning is done for the depletion implant which follows as shown on Fig. 1.7. Finally, in this set of implants, the buried contact area, which is the first level of interconnection delineated; this is shown in Figure 1.8. A section through AA' on Figure 1.9 shows the enhancement transistor, and the buried contact, while Figure 1.11 shows buried contact and depletion mode transistor.

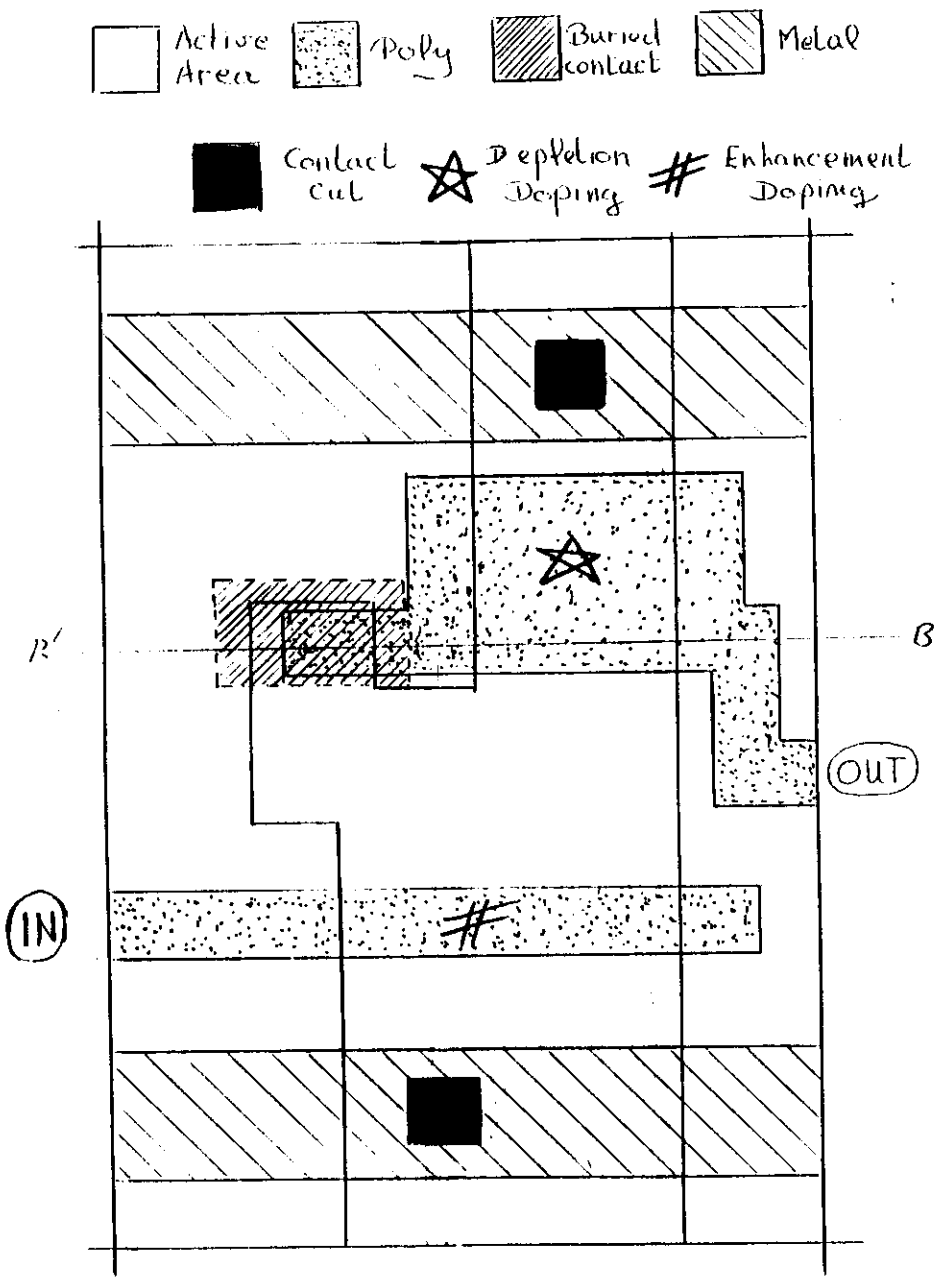


FIG I.3. Mask Layout for Depletion Load Inverter in nMOS.

FIRST MASK (n^+)

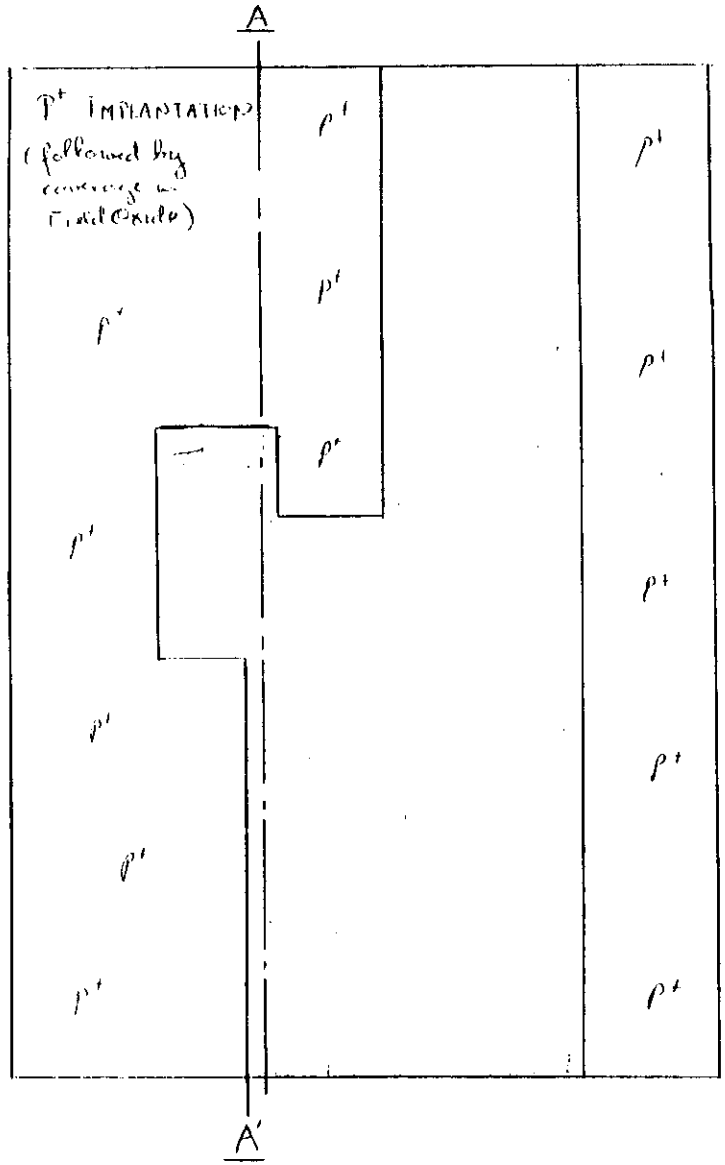


FIG I.4(a) Solid area is left covered in SiO_2/Si_3N_4. Exposed area is p^+ doped (boron) and then oxidized.

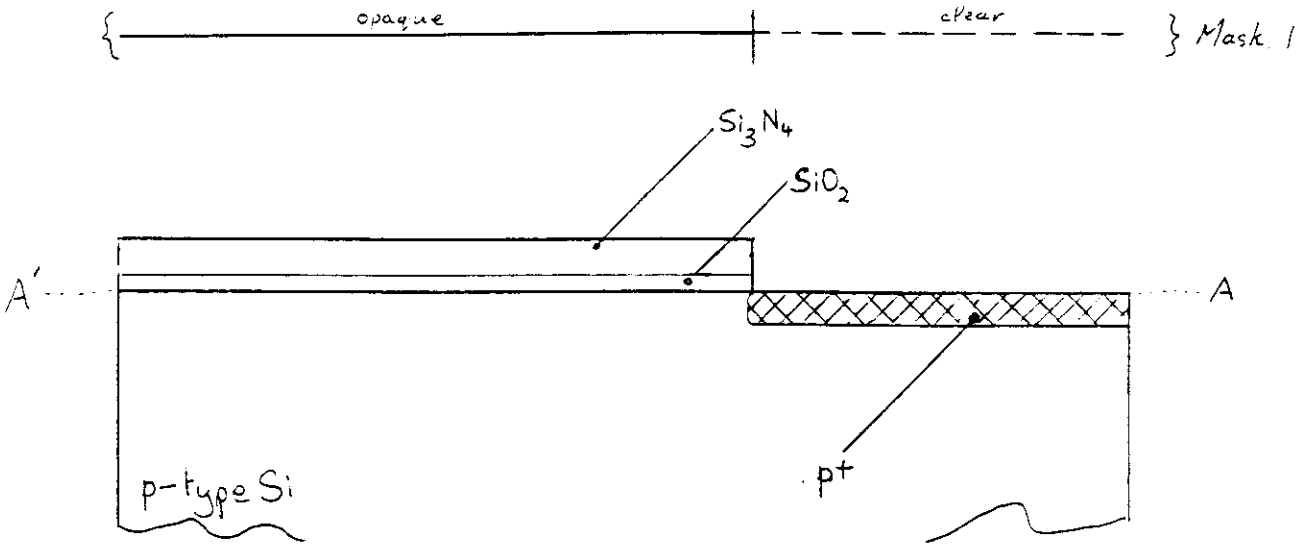


FIG I.4(b). Section through AA' showing p^+ implantation region. P^+ is ~ 1/4 μm deep. (prior to oxidation). SiO_2 200Å thick and Si_3N_4 800Å

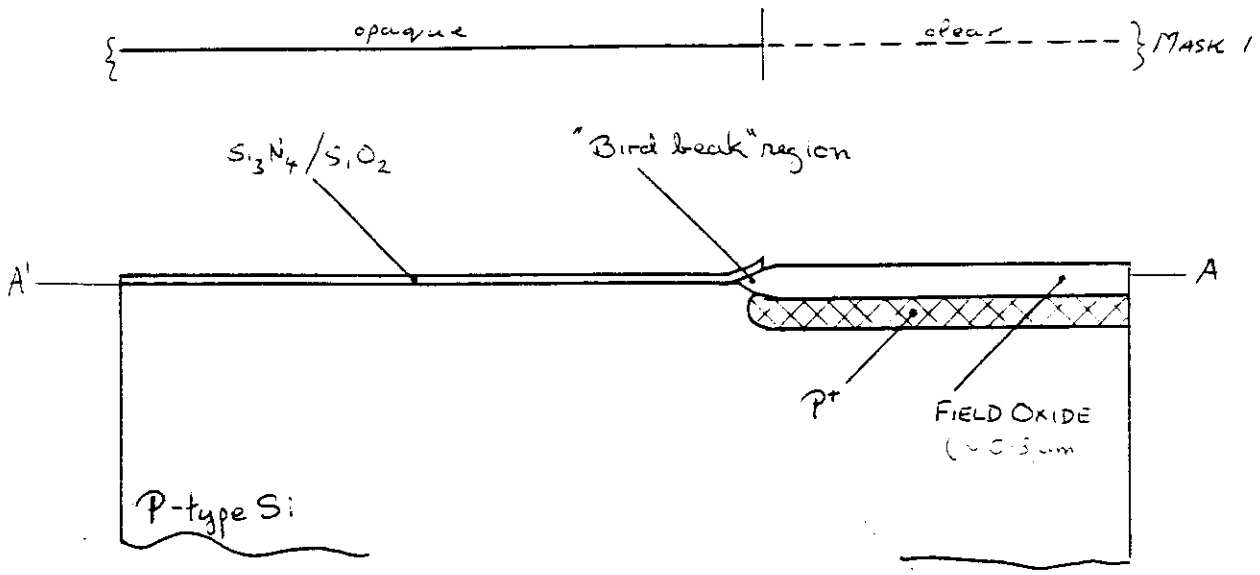


FIG I.5 Section through AA' showing P^+ implant after growth of field oxide. (Note advance of oxide to the left and formation of bird beak structure)

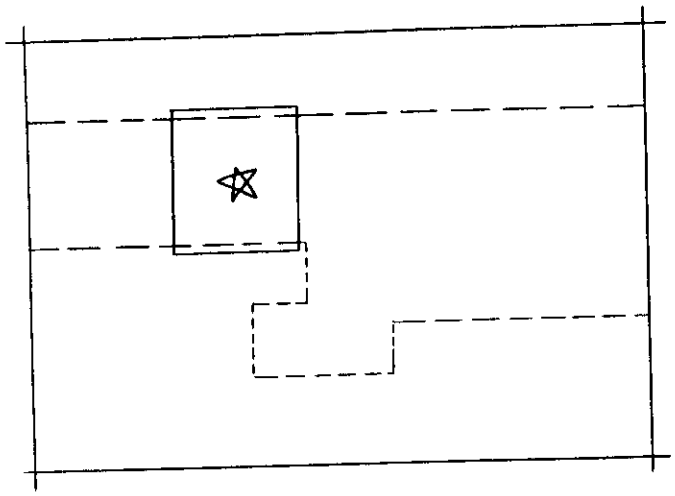


FIG. I.7. Third Mask: Heavy donor ion implantation to create permanently ON n-channel, which is the depletion (pullup) transistor

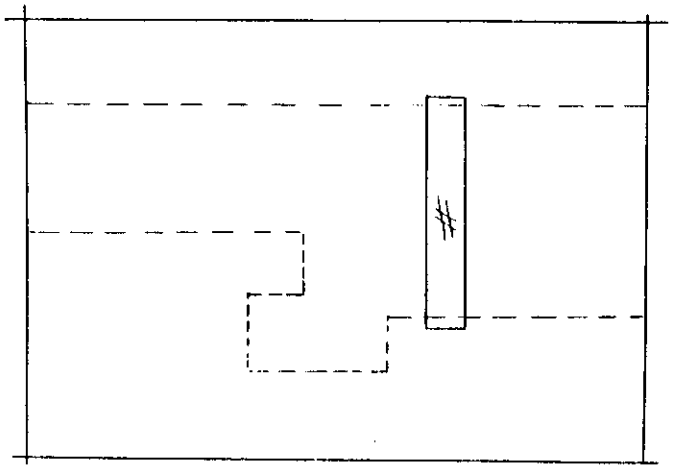


FIG I.6 Second Mask: Light acceptor ion implantation to adjust V_t to desired level. This will become the enhancement (pulldown) transistors. (Id marks shown as dotted lines)

Fourth Mask: n^+

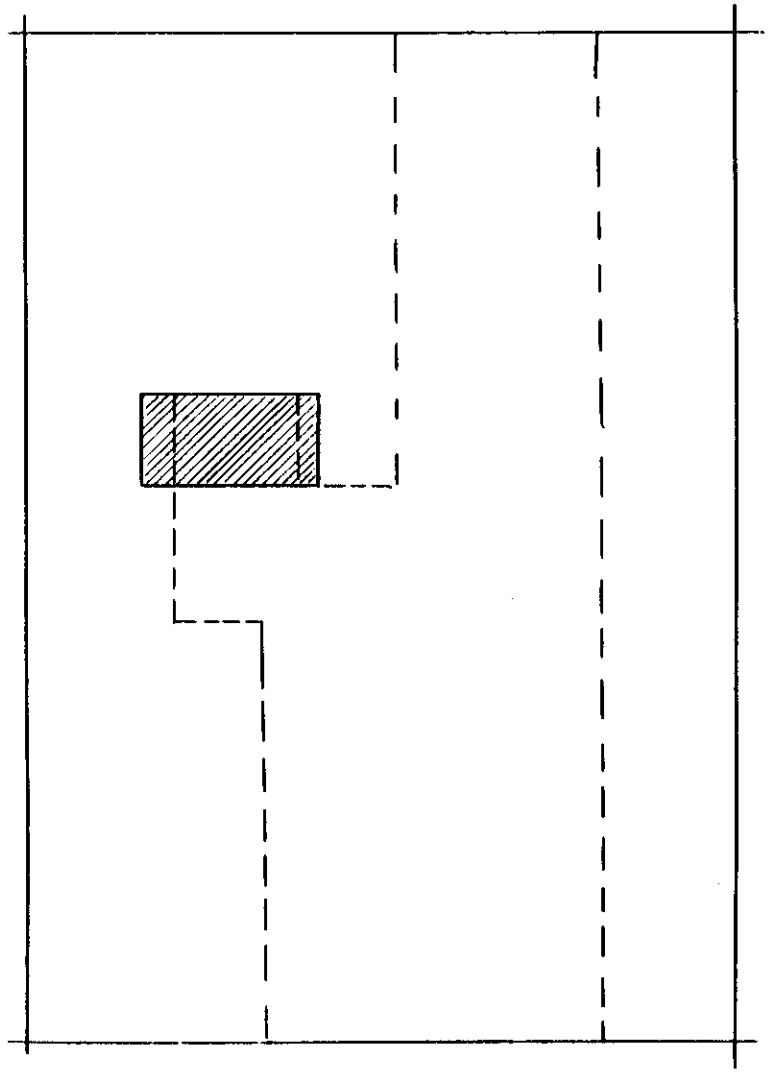


FIG. I.8 Fourth mask: Buried contact, n^+, contact area for Poly.

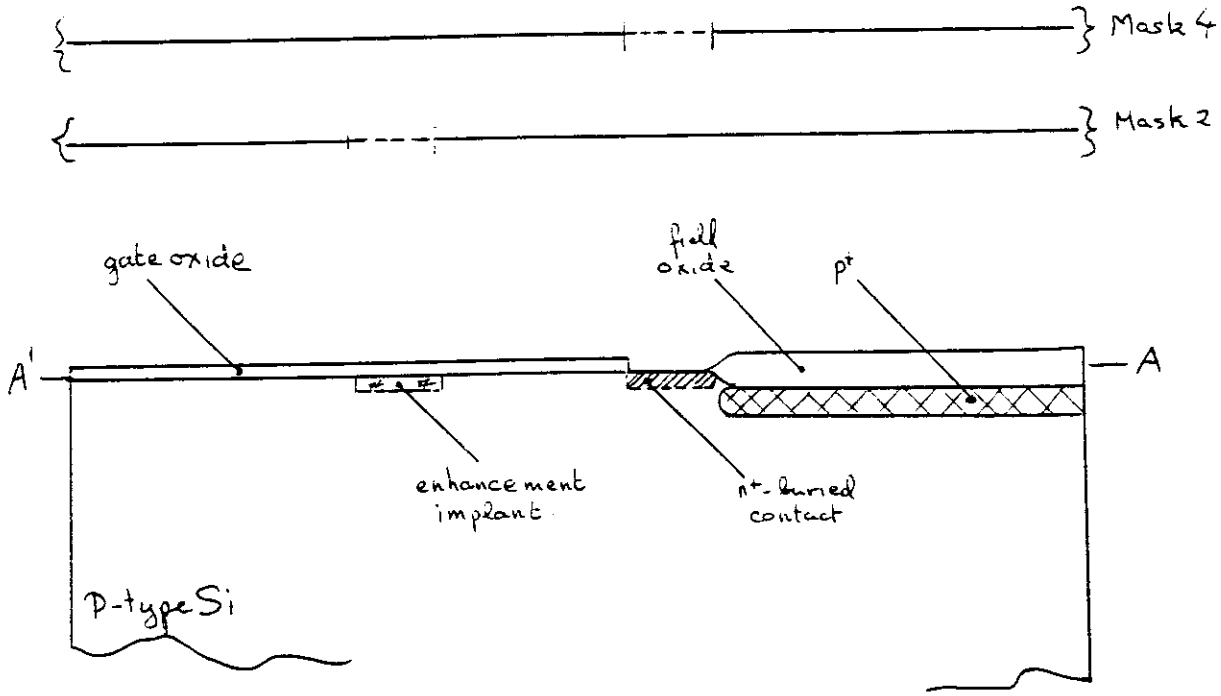


FIG I.9. Section through AA' showing enhancement implant (mask 2) and buried contact implant (mask 4) and then gate oxide removed over the buried contact area (ready for contact to Poly).

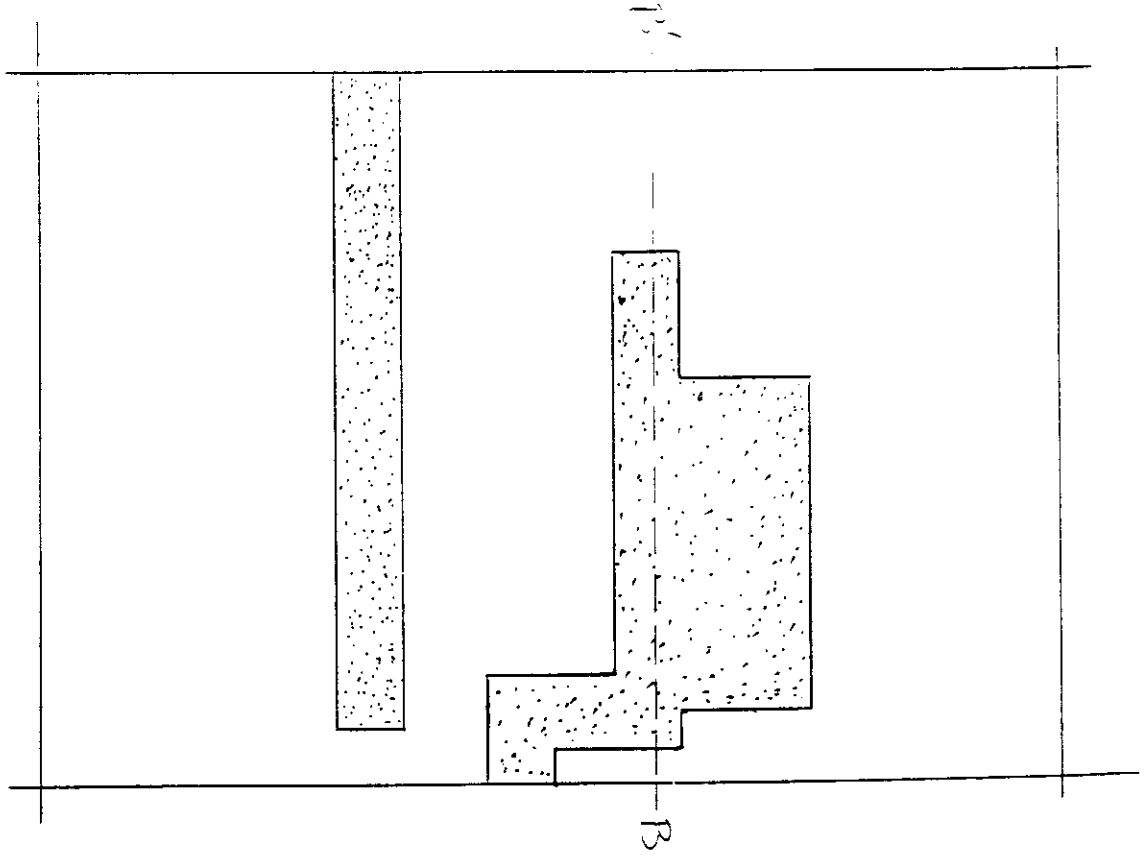


FIG I.10 Fifth Mask: Poly-silicon

A section through 'BB' is shown on FIG I.11.

Fifth Mask: Poly-Si

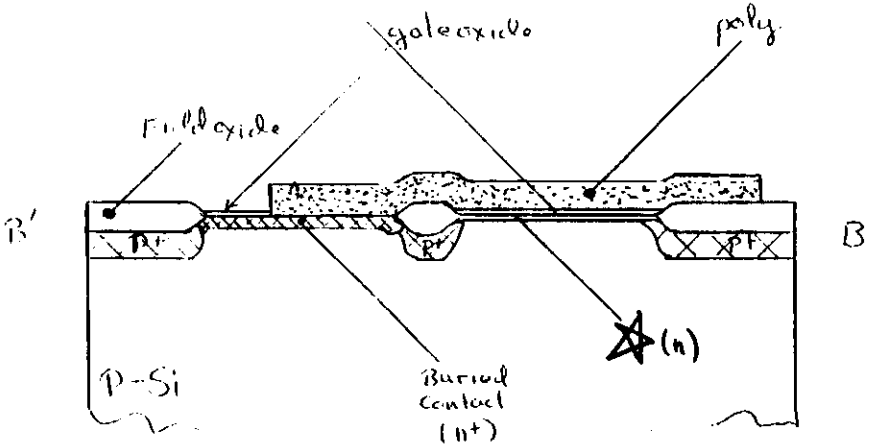


FIG I.11 Section through BB' (of Fig I.3) showing buried contact and full width of depletion mode transistor.

The last one of the three masks, the buried contact mask, is used again to delineate a window (in photo resist) on top of the buried contact. The thin gate oxide at the bottom of the photo resist window is etched away. We next proceed to the polysilicon steps.

Heavily doped n-type poly is deposited (by chemical means) to a layer thickness of about 4000Å, giving a sheet resistance in the range of 10-30 ohms per square. The poly is then coated with photo resist and patterned using mask 5, cf. Figure 1.10: this will form the second interconnection level. Where the active area is exposed n^+ regions are going to be created defining source and drain areas, the poly over active area being gate. The poly and gate oxide exposed is etched away and an n^+ ion implantation carried out, the poly having been the self-aligning mask. These n^+ implanted areas are then lightly re-oxidized producing thin, strong oxide, annealing out implant damage and improving the contact between poly and the buried contact. A section through BB' at this point in fabrication is shown in Fig. I.11.

The rest of the wafer processing is concerned with putting down another layer of oxide insulator (in fact a phospho-silicate glass) which is patterned (mask 6) and through which are etched contact holes down to the n^+ buried contact regions, and to the poly (such a contact hole is called a via). Aluminium is then deposited and patterned (mask 7). Finally the entire circuit is encapsulated in an over-glass, this is patterned (mask 8) and then holes are etched through to expose pad areas for wire bonding. Eight masks will have been used in this process, these are listed in Table 1.1, while Table 1.2 is a resume of the process.

Table I.1 : Masks used in nMOS process

1.	Active area where source and drain contacts and gates will appear, also buried contacts.
2.	Enhancement mode transistors.
3.	Depletion mode transistors.
4.	Buried contacts.
5.	Polysilicon
6.	Metal contact holes
7.	Aluminium pattern.
8.	Bonding holes in passivation layer.

Table 1.2 : Outline of nMOS Process

Start:	Si wafer p-type 3 x 10^{15} B/cm^3. DN zone, internal and back side gettering.	
	First layers:	200Å thermal oxide + 800Å Si_3N_4
Key :	PR	(layer of photo resist, positive)
	P	(UV exposure of mask pattern, followed by developing - lithography)
	E	(etch of pattern delineated by resist)
	LOCOS	(Oxidation of silicon)
	D	(Doping of silicon by Ion Implantation)
	L	(layering with poly or metal)
1st Mask :	P, E, D	This delineates the active area and surrounds it with a p^+ sea of implanted boron atoms. Next follows
Active area	LOCOS	over the p^+ area. Bird-beak and atom diffusion.
	Gate Oxide	is now created in the active area, by stripping the SiO_2/Si_3N_4 oxidation mask and reoxidizing: 500Å
2nd Mask :	PR, P, D.	This adjusts the p doping in gate region of the enhancement transistor
Pull down	(V, threshold). Finally, strip off photo resist.	
3rd Mask :	PR, P, D.	This is an arsenic implant to delineate the depletion mode transistor
Pull up	and its resistance. Strip photo resist.	
4th Mask :	PR, P, D.	This is another arsenic implant to create an n^+ buried contact region.
Buried Contact	The same mask is used for PR, P and then E to remove gate oxide in order to receive POLY in a reducing atmosphere for first level interconnection.	
5th Mask :	L with poly, PR, P, E. Strip photoresist	
Source Drain Connections	Now, using the patterned Poly as the self-aligned mask for D, n^+ (boron) implant into the exposed active area. Lightly re-oxidized and annealing, then L with CVD oxide	
6th Mask :	PR, P, E. Contact cuts	
Contact cuts	Strip photo resist	
7th Mask :	L with aluminium. PR, P, E strip photoresist	
Contacts & Interconnects	Followed by heat treatment for alloy contacts. Then overglass	
8th Mask :	PR, P, E, expose bonding pad area for external connection	
Bond Pads	Strip photoresist Followed by bonding and final encapsulation	

The nMOS process has been used as an illustration of VLSI technology and we will consider the modifications that CMOS technology requires shortly. But before we do that, it would be constructive to consider how the fabrication technology and the topological design rules are connected.

I.5 COMMENTS ON DESIGN RULES

Rules which govern the limits of the geometry of the various features of integrated circuits are called DESIGN RULES. These rules, if they are to be really useful, must be independent of any particular pattern adopted. The rules come about through a combination of the constraints of the fabrication processes and the electrical limitations of devices, and are generally a list of minimum feature sizes, or widths (W), separations between features (S) and overlaps (E), which are tolerable.

I assume that you are completely familiar with these rules from your VLSI design courses.

The basic metric, λ, is half the minimum source/drain separation attainable. This value is almost exclusively determined by the pattern transfer process, and in today's most advanced technology stands at about 0.8μm.

It is a combination of lithography and mask-to-mask registration which is at the core of what determines λ. But there are many inputs which determine the three characteristic features.

A simplified version of the rules is illustrated in Figures I.12, I.13 and I.15 and are as follows (see Table I.1 for mask levels).

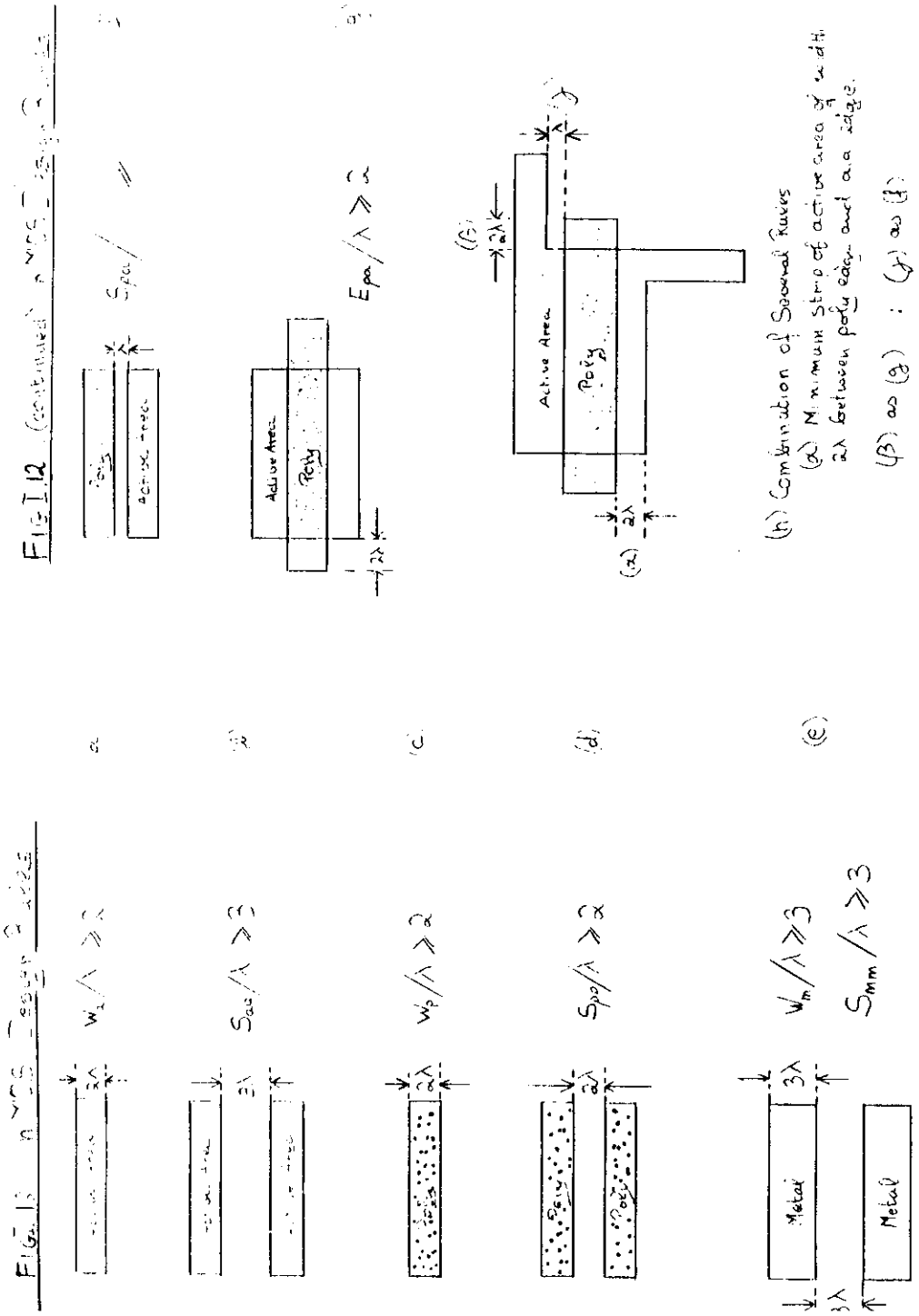
a) Minimum width (W_a) of <u>active area</u> is 2λ, and is determined essentially by patterning and subsequent atom diffusion or annealing. This is at the active mask level.

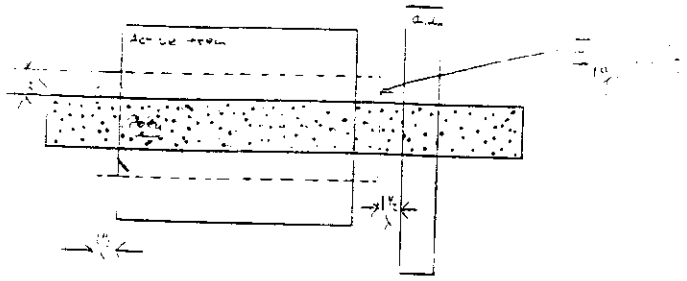
b) Minimum width of separation between <u>active areas</u> (S_{aa}) is 3λ. Here apart from lithography and processing we also have the pn junction depletion layer width to consider.

c) and d) and g) These refer to <u>polysilicon</u> and are at the polysilicon mask level (mask 5 in Table I.1). Here 2λ prevails for W, S and E: it is less than active area S_{aa}, there being no depletion layer to consider. While W_p and S_{pp} is single level process limited, E_{pa} involves lithographic registration between levels.
Thus it is necessary for poly (which will be the gate of a transistor) to reach beyond the active area to ensure that the source drain regions in the active area are not short circuited.

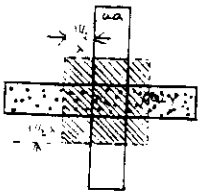
e) The <u>metal</u> mask is the seventh mask used and by this time the material is travelling over pretty rough ground, as shown in Fig. I.14. Thus the minimum width (W_m) and separation (S_{mm}) is increased to 3λ.

f) The minimum separation between poly and adjacent active areas (S_{pav}) is λ. Any





(a) Depletion implant. The implant for the depletion mode transistor must extend all around the gate area by $1\frac{1}{2}\lambda$, and must be separated from other active (and non-implanted) areas by at least $1\frac{1}{2}\lambda$. Thus $E_{ig}/\lambda \geq 1\frac{1}{2}$; $S_{ig}/\lambda \geq 1\frac{1}{2}$.

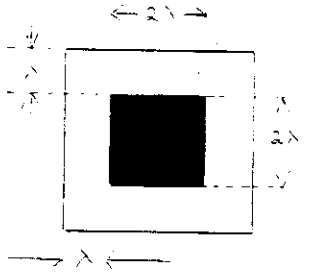


$E_{ic}/\lambda \geq 1\frac{1}{2}$

(b) Buried contact implant.

FIG. I.13 nMOS design rules for ion implantations.

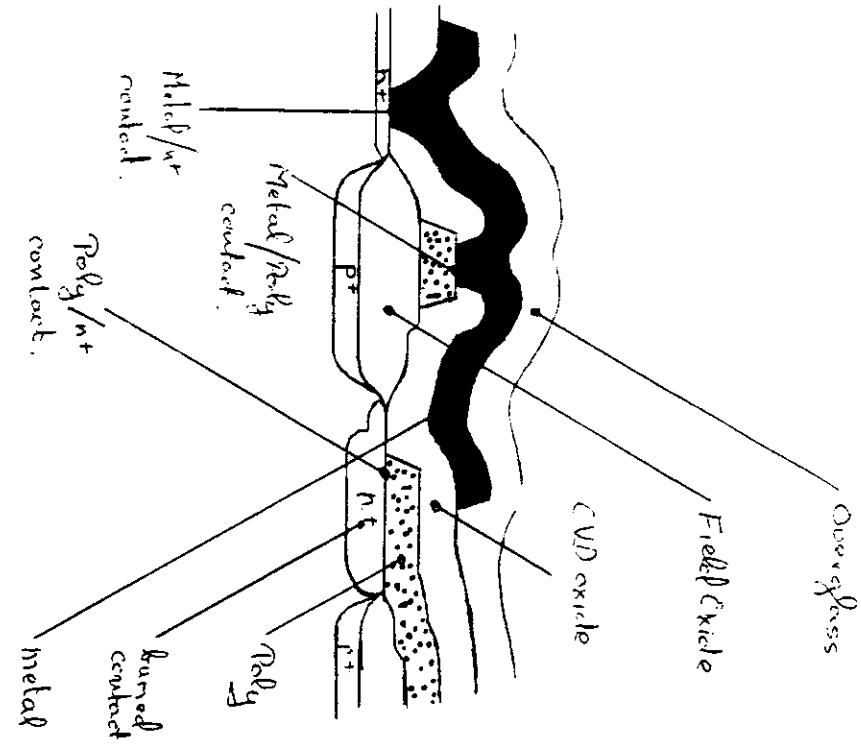
$W_c/\lambda \geq 2$



E_{ac}/λ; E_{pc}/λ; $E_{mc}/\lambda \geq 1$

as well as others

FIG I.15 nMOS design rules for various contacts shown on Fig I.12.



(P-type substrate)

FIG I.14 Various types of contact in nMOS technology.

overlap would result in unwanted capacitance, but nothing more serious.

i) The enhancement (mask level 2) and depletion (mask level 3) mode implants both have extensions around the gate area (E_{ig}) (ie the area of poly/active-area overlap) of $1\frac{1}{2}\lambda$. Otherwise we should have a maverick strip of source/drain region.

j) The buried contact, which is mask level four, requires a minimum extension around the via of λ, and a minimum spacing between the buried contact area edge to unrelated poly or active area of 2λ.

A final remark about design would seem to be in order, and that is the reminder that many of the electrical parameters which are important in design are of a "second order" nature. Thus mobility is usually taken as a constant, whereas in fact it depends upon interface state density, device temperature and substrate doping. Channel length depends upon the two-dimensional nature of the device (corners matter), and so forth and so on. These matters will be raised in context rather than in a special section.

Now to CMOS very briefly.

I.6 THE BASIC n-WELL CMOS PROCESS

In the CMOS (complementary MOS) process the n-channel enhancement transistor and the p-channel enhancement transistor are differentiated by putting the appropriate transistor into a well. Thus with a p-substrate we would use n-well technology putting the p-channel transistors into the n-wells.

There are great advantages to CMOS technology and some disadvantages. I leave that to the class. However, it may well be that of the submicron level nMOS may still win the day!

The n-well CMOS process is very similar to the nMOS process. An important difference occurs in the starting material and the first mask, which is used to define the well. This is illustrated in Fig. I.16 which shows a layer of p-silicon (epi-Si, essential to avoid latch up) grown on a p^+ doped substrate and an n-well which has been made by thermal in-diffusion of, say, arsenic. The buried contact is not used. Table I.3 gives an outline of the process.

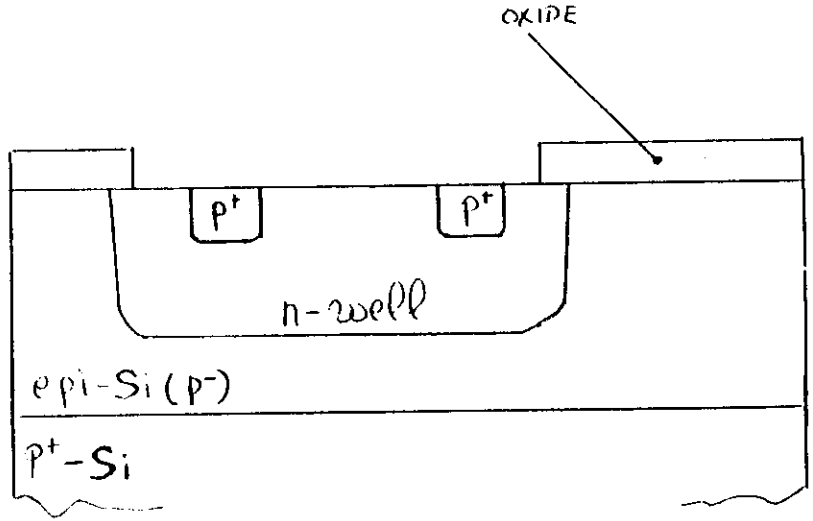


FIG I.16 An n-well diffusion in a CMOS process with epitaxial silicon (The epi-Si layer under the well must be thick enough to avoid punch through)

Table 1.3 : Summary of basic CMOS process for n-well devices

- o Grow oxide. Spin on photoresist and expose with n-well mask.
- o Develop photoresist and etch oxide. Strip photoresist, then diffuse n-well except where masked by oxide. Strip oxide.
- o Grow buffer oxide and deposit Si_3N_4. Spin on photoresist.
- o Expose active area mask to define eventual channel and diffusion regions. Develop photoresist and etch Si_3N_4.
- o Grow field oxide using the Si_3N_4 as a mask. Strip remaining Si_3N_4.
- o Grow field oxide using the Si_3N_4 as a mask. Strip remaining Si_3N_4.
- o Grow gate oxide.
- o Use p+ and n+ masks to adjust transistor thresholds.
- o Deposit the poly and pattern with the poly mask. Use poly as a mask to etch the gate oxide.
- o Use the n+ mask to define the source/drain regions to receive an arsenic implant. Also defines n+ plugs to n-well.
- o Use the p+ mask to define the source/drain regions to receive a boron implant. Also defines p+ plugs to substrate.
- o Deposit CVD oxide and pattern metal contact cuts with the contact mask. Etch contacts.
- o Deposit metal. Pattern with metal mask. Etch metal.
- o Deposit over-glass. Pattern with over-glass mask. Etch holes for bonding wires.
- o Probe wafers and ink bad die.
- o Saw and dice wafers.
- o Package, bond, seal and re-test.