

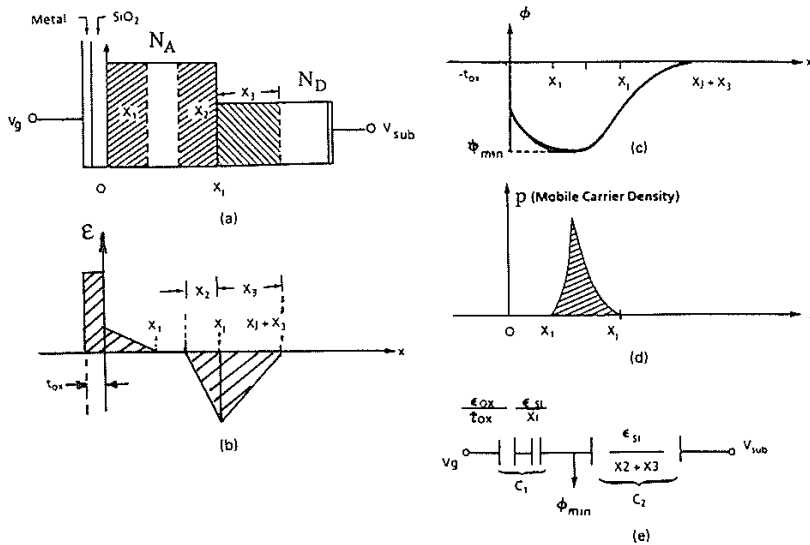


Fig. 5-59 Buried-channel PMOSFET when biased to inversion: (a) Doping profile and depletion regions; (b) electric field; (c) electron potential; (d) mobile carriers; and (e) equivalent capacitances in sub-pinchoff condition.¹²⁰ (© IEEE 1978).

5.8.2.2 Subthreshold Swing (S_t) in Buried-Channel PMOS. The PMOSFETs used in CMOS must be enhancement-mode devices; i.e., they need to be OFF when $V_{GS} = 0V$. If such PMOSFETs are buried-channel devices, the entire implanted p -layer in the BC PMOS devices must be depleted of holes when $V_{GS} = 0V$. On the other hand, upon applying a gate voltage equal to the threshold voltage, the p -layer of such devices is no longer depleted. That is, holes will have returned to the layer, forming a p -type conducting channel between the source and drain. But this channel will exist beneath instead of at the surface. Thus, for gate voltages greater than V_T , two depletion regions will be present in this p -layer; one that is formed between the surface and the buried channel (to a depth x_1) and the other due to the channel pn junction, as is shown in Fig. 5-59a. The BC MOSFETs thus have two capacitances in addition to the oxide capacitance $\kappa_{ox}\epsilon_0/t_{ox}$; namely the depletion capacitance from the surface ($\kappa_{si}\epsilon_0/x_1$), and the depletion capacitance from the pn junction ($C_2 = \kappa_{si}\epsilon_0/[x_2 + x_3] \sim \kappa_{si}\epsilon_0/x_3$). The potential minimum is located in the position between these two capacitors, as shown in Fig. 5-59e. As also noted earlier, the location of the potential minimum causes most of the free carriers to appear in the bulk Si rather than at the Si-SiO₂ interface. It is expected that since the free carriers are now farther away from the surface they will be harder to modulate by the gate voltage. (In the same way that

carriers in a surface-channel MOSFET with a thicker gate oxide are harder to modulate - because of reduced gate-to-channel capacitive coupling - than in one with a thinner gate oxide.) Hence, to achieve the same degree of turn-off, the gate voltage will need to be reduced more than in the case of a surface-channel device. Thus, we expect that the turn-off characteristic will be less sharp, which means that S_t will be larger.

Quantitatively the buried channel turn-off characteristic can be modeled by an expression that gives the subthreshold swing S_t in a structure with two such depletion regions, i.e.,

$$S_t = 2.3 (kT/q) (1 + C_2/C_1) \quad (5 - 96)$$

$$= 2.3 (kT/q) [1 + (\kappa_{ox}\epsilon_0 x_1 + \kappa_{si}\epsilon_0 t_{ox})/(\kappa_{ox}\epsilon_0 x_2 + \kappa_{ox}\epsilon_0 x_3)] \quad (5 - 97)$$

where C_1 is the oxide capacitance (C_{ox}) in series with the depletion capacitance from the surface ($\kappa_{si}\epsilon_0/x_1$), and C_2 is defined above. Let us compare Eq. 5-96 with Eq. 4-91 (which gives S_t for the case of surface-channel MOSFETs, and is repeated here for convenience).

$$S_t = \ln 10 (d \ln I_D / d V_{GS})^{-1} = 2.3 (kT/q) (1 + C_d/C_{ox}) \quad (5 - 98)$$

By noting that C_1 is smaller than C_{ox} , and C_2 is typically comparable in value to C_d , we see that S_t is expected to be larger in BC devices. Experimental results confirm that S_t is worse (i.e., larger) in buried channel devices, especially in short-channel PMOSFETs. That is, S_t values in buried-channel PMOSFETs range from 0.5 to 1.5 V/decade, compared to the 70-120 mV/decade S_t values observed in active NMOSFET devices. Figure 5-60 compares the subthreshold behavior of the two $0.8 \mu\text{m}$ BC PMOS

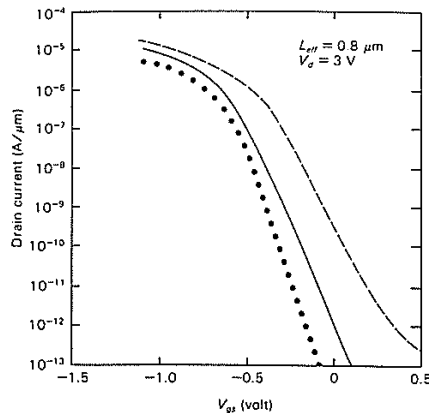


Fig. 5-60 Subthreshold I-V curves in PMOS device in CMOS. Dashed lines: buried channel PMOS in a lightly doped substrate. Solid lines: buried-channel PMOS in a retrograde n -well. Dotted lines: surface channel NMOS drawn for comparison.⁷² (© IEEE 1982).

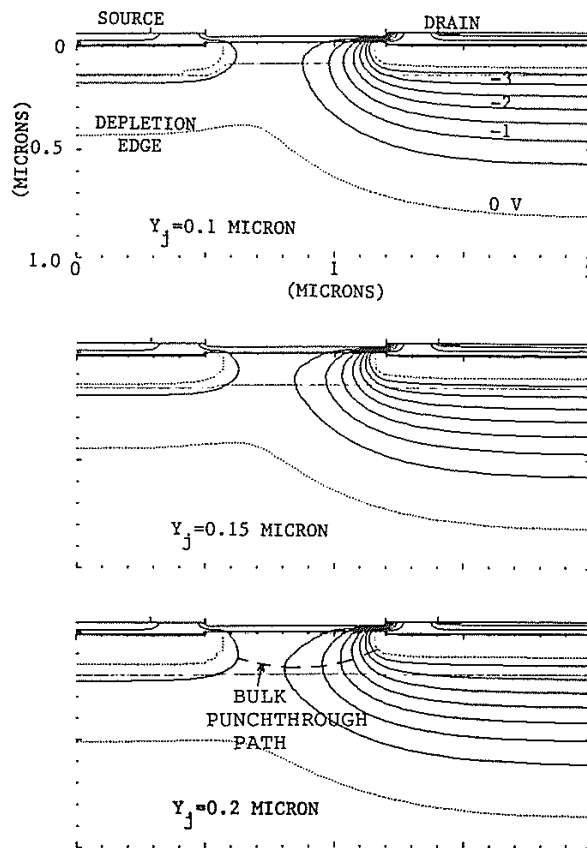


Fig. 5-61 Simulated 2-D potential profile for p -channel MOSFET with $Y_j = 0.1, 0.15$, and $0.2 \mu\text{m}$. $V_{GS} = 0\text{V}$ and $V_{DS} = -3\text{V}$.¹¹⁵ (© 1984 IEEE).

devices with that of a comparable surface-channel NMOSFET. (Note that the V_{GS} scale for the NMOSFET should be $+1.5\text{V}$ to -1.5V from left to right.) It indicates that S_t is larger for the BC PMOS devices.

The effect on S_t of the depths of the channel and source/drain junctions Y_j and r_j , respectively, and of the gate-oxide thickness t_{ox} has also been investigated with simulation studies. Figure 5-61 illustrates the 2-D simulated potential profiles for a $0.5 \mu\text{m}$ p -channel MOSFET as Y_j is increased from 0.1 to $0.2 \mu\text{m}$. It is observed that as Y_j is increased the potential contours spread more from drain to source, and the bulk punchthrough path passes through the depletion region near the pn junction. Figure 5-62 shows the calculated values of S_t for three Y_j values (i.e., $0.1, 0.15$, and $0.2 \mu\text{m}$), as r_j is varied in a BC PMOS device with an $L_{eff} = 0.5 \mu\text{m}$, a $t_{ox} = 25 \text{ nm}$,

bulk punchthrough path passes through the depletion region near the pn junction. Figure 5-62 shows the calculated values of S_t for three Y_j values (i.e., 0.1, 0.15, and $0.2 \mu\text{m}$), as r_j is varied in a BC PMOS device with an $L_{\text{eff}} = 0.5 \mu\text{m}$, a $t_{\text{ox}} = 25 \text{ nm}$, and $V_{\text{DS}} = -3\text{V}$. Each curve corresponds to a fixed value of Y_j . The threshold voltage [defined as the gate voltage at which the drain current is $100 \text{ nA} \cdot (Z/L)$] is specified at each point. The data shows that S_t is strongly dependent on Y_j and r_j . That is, as Y_j is increased for a fixed value of r_j , S_t increases. For a fixed value of Y_j , S_t also increases as r_j is increased. In Fig. 5-63 the simulated value of S_t for two different gate oxide thicknesses is shown as Y_j is varied. It is seen that S_t improves for thinner gate oxides, but the dependence of S_t on Y_j is only slightly reduced. The overall conclusion from these studies is that small values of Y_j , r_j , and t_{ox} are also needed to realize small values of S_t in submicron BC PMOS devices.

5.8.2.3 Punchthrough Susceptibility of Buried-Channel PMOS.

The larger S_t values exhibited by BC PMOS devices implies that they are susceptible to larger leakage currents in the OFF state. In fact, such relatively large leakage currents are commonly observed in short PMOSFETs, especially at high drain bias. For example, it has been shown that in submicron PMOS transistors ($L_{\text{eff}} = 0.8 \mu\text{m}$), I_{DST} will increase by two orders of magnitude if V_{DS} is increased from -1V to -10V .⁷⁴ In this case, if I_{DST} was 1 nA at -1V , it would increase to $0.1 \mu\text{A}$ if V_{DS} was raised to -10V . It would be useful to be able to predict the value of I_{DST} prior to fabricating the device.

As described in section 5.3.1 an increase in S_t has been correlated to the onset of subsurface punchthrough in surface-channel MOSFETs. The increase in S_t in BC PMOS devices as a result of various device parameters (i.e., L , Y_j , r_j) discussed in the previous section can also be considered from the perspective of the phenomenon of punchthrough (subsurface-DIBL), which was introduced in section 5.3. However, we will see that the subthreshold leakage behavior in BC PMOSFETs differs from that observed in surface-channel FETs. To help explain this difference, we will once again

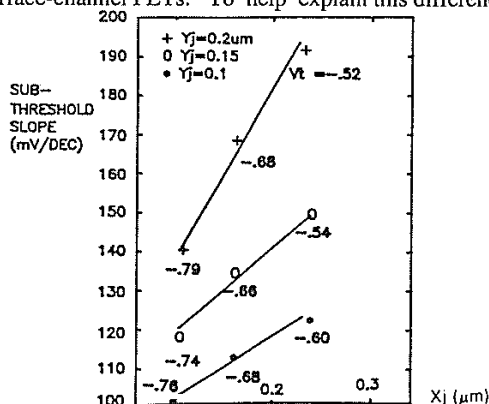


Fig. 5-62 Simulated subthreshold slope versus r_j and Y_j for p -channel MOSFETs in which $L_{\text{eff}} = 0.5 \mu\text{m}$, $t_{\text{ox}} = 250 \text{ \AA}$, and $V_{\text{DS}} = -3\text{V}$. V_T is the threshold voltage.¹¹⁵ (© 1984 IEEE).

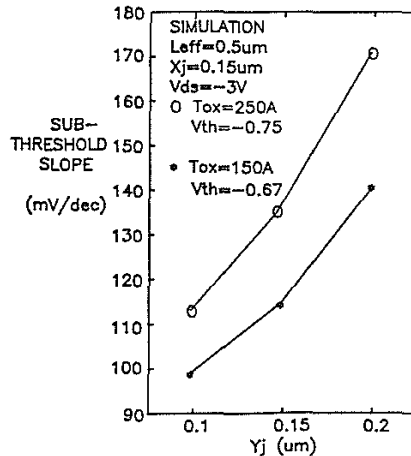


Fig. 5-63 Simulated subthreshold slope versus Y_j and t_{ox} .¹¹⁵ (© 1984 IEEE).

employ the punchthrough modeling technique described in section 5.3.2, but here it will be applied to quantitatively model I_{Dst} in BC PMOSFETs. The results of this analysis will permit us to compare the subthreshold leakage behavior of surface-channel and buried channel MOSFETs.

We recall that punchthrough current is determined by carrier injection over the potential energy barrier at the saddle point of the electrostatic potential in the bulk (see Fig. 5-64).¹¹⁷ Therefore, analysis of the channel potential, especially the change of the potential at the saddle point P of the BC PMOS device with V_{DS} is necessary. This analysis can be performed with the aid of 2-D device simulators,** and here we present the results of one report that performed such a study using the device simulator GEMINI.⁷⁴ The simulated potential along the direction from source to drain of a BC PMOSFET in an n -well with $L_{eff} = 0.55 \mu\text{m}$ as V_{DS} is increased is shown in Fig. 5-65. It is observed that the magnitude of ϕ^* (the potential at the saddle point P), begins to drop as soon as V_{DS} is increased above 0V. This is due to the fact that there is no pn junction along the channel region in a p -channel BC MOSFET. (Conversely, in the case of short, surface-channel MOSFETs, at low drain-bias values the V_{DS} drop occurs almost entirely across the drain-channel junction, and the barrier at the source does not decrease as long as V_{DS} remains small. Only when V_{DS} becomes large enough to cause subsurface-DIBL does the barrier begin to drop as V_{DS} exceeds this value.) The decrease of the potential barrier height ($\phi^* - V_S$) in a BC PMOSFET as V_{DS} is increased is plotted in Fig. 5-66. Here we see that $(\phi^* - V_S)$ is linearly proportional to the increase in drain voltage, even at low values of V_{DS} .

** Note that an entirely analytic model of punchthrough in PMOSFETs has also been published by Skotnicki et al.¹¹⁷

Once $(\phi^* - V_S)$ is found for a given value of V_{DS} , it can be used to calculate I_{DSt} with the aid of an analytic expression (Eq. 5-18b) as discussed in section 5.3. This expression for I_{DSt} is repeated here for the case of a PMOS device:

$$I_{DSt} = -q D_p Z \frac{t^*}{L^*} \frac{n_i^2}{N_D} \exp\left[\frac{q}{kT}(V_S - \phi^*)\right] \quad (5-99)$$

To exploit the results plotted in Fig. 5-66, we express Eq. 5-99 in logarithmic form as

$$\log(I_{DSt}) = \{\log[-q D_p Z (t^*/L^*) (n_i^2/N_D)]\} - q(\phi^* - V_S)/k \quad (5-100)$$

or

$$\log(I_{DSt}) = A - q\phi_b/kT \quad (5-101)$$

where $A = \log[-q D_p Z (t^*/L^*) (n_i^2/N_D)]$, and $\phi_b = (\phi^* - V_S)$. Now, Fig. 5-66 shows that the decrease in ϕ_b is linearly proportional to V_{DS} , i.e.,

$$\phi_b = (\phi_{b0} - b|V_{DS}|) \quad (5-102)$$

where ϕ_{b0} is the barrier height at $V_{DS} = 0V$, and b represents the slope of the curve of Fig. 5-66 (i.e., $d\phi_b/dV_{DS}$). We also note that the values of both ϕ_{b0} and b can be determined from Fig. 5-66. Then we can write Eq. 5-101 as

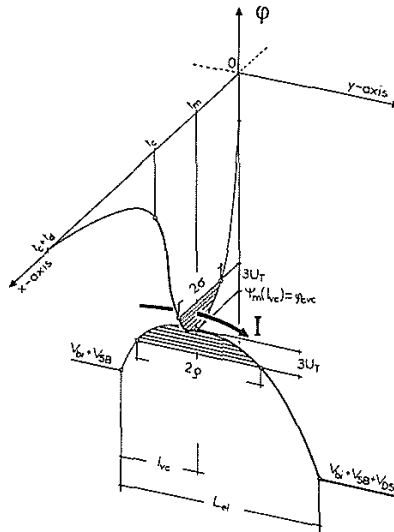


Fig. 5-64 Perspective view of the electrical potential distribution in two perpendicular cross sections of the channel of a buried-channel PMOSFET. Shaded regions are those corresponding to t^* and L^* in Eq. 5-99.¹¹⁷ (© 1989 IEEE).

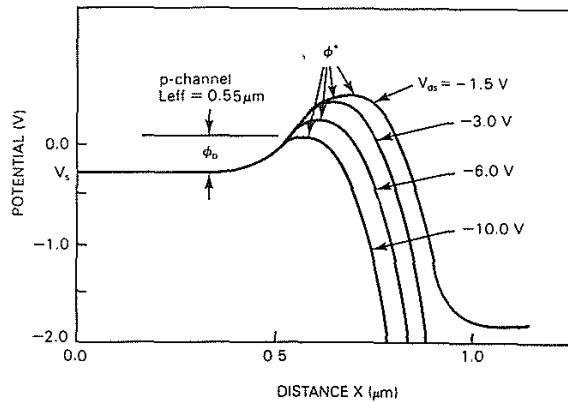


Fig. 5-65 Simulated potential distributions along the buried channel of a PMOSFET for various drain voltages.⁷⁴ (© 1988 IEEE).

$$\log(I_{Dst}) = A - q(\phi_{b0} - b|V_{DS}|)/kT \quad (5-103)$$

or

$$= A' + V_{DS}/R \quad (5-104)$$

where $A' = A - q\phi_{b0}/kT$ and $R = kT/bq$. The parameter R is the incremental voltage that the drain can sustain before I_{Dst} increases by another order of magnitude ($R = \partial V_{DS}/\partial \log I_{Dst}$). The larger the value of R , the lower the DIBL rate, hence the better the MOSFET is able to maintain low subthreshold leakage as the drain bias rises.

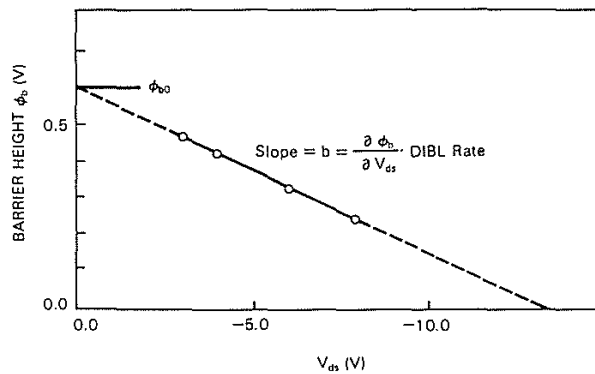


Fig. 5-66 Potential barrier height as a function of drain voltage.⁷⁴ (© 1988 IEEE).

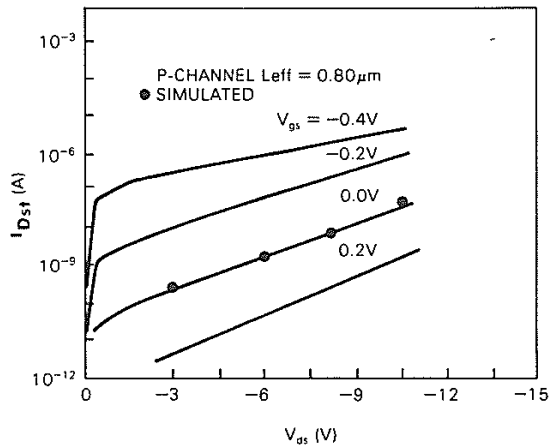


Fig. 5-67 Subthreshold I_{Dst} - V_{DS} characteristics for a submicron PMOSFET.⁷⁴ (© 1988 IEEE).

Equation 5-104 implies that the relationship between $\log I_{Dst}$ and V_{DS} should be linear in BC PMOSFETs operating in subthreshold. Figure 5-67 shows the measured $\log I_{Dst}$ - V_{DS} curves for a $0.8 \mu\text{m}$ BC PMOSFET, together with some simulated I_{Dst} values at $V_{GS} = 0\text{V}$. The agreement between the experimental and simulated data is good. We also see that R in this device is about -4.4V/decade for $V_{GS} = 0\text{V}$. This means that if I_{Dst} is 10 pA at $V_{DS} = -1\text{V}$, it will increase to 10 nA at $V_{DS} = -10\text{V}$.

Now we can compare the I_{Dst} behavior of SC and BC MOSFETs with respect to externally applied biases. We note that in the case of surface channel MOSFETs two regions of different slope are observed in the $\log I_{Dst}$ versus V_{DS} curves (see Fig. 5-11). At low values of V_{DS} the value of R is large (e.g., in Fig. 5-11b $R = 12.6\text{V/decade}$ when $V_{DS} < 6.0\text{V}$), but when V_{DS} exceeds about 8.0V , the value of R drops to about 1.0V/decade . This change implies that in surface-channel MOSFETs there is a transfer from surface-DIBL to subsurface-DIBL (punchthrough) - e.g., in the device of Fig. 5-11b at about 6.0V . Also, up to some value of V_{DS} (e.g., V_{PT}), the drain bias has little effect on I_{Dst} . In BC PMOSFETs, however, the value of R is relatively constant (e.g., it has a magnitude of $R = 4.4\text{V/decade}$ for all values of V_{DS} in the case of the device characterized in Fig. 5-67). This implies that the nature of subthreshold leakage current in BC PMOSFETs does *not* undergo a change from surface- to subsurface-DIBL, but instead the DIBL phenomenon in a BC device is a combination of both effects. Furthermore, the intermediate value of R also supports the view that *both* the gate and the drain voltages affect the potential of the p -type layer, and thus both can impact I_{Dst} at all values. Finally, this indicates that the measured slope of $\log I_{Dst}$ versus V_{DS} (defined as $1/R$), reflects the effect of V_{DS} on ϕ_b through DIBL, and permits a comparison of the subthreshold leakage behavior of various PMOS device structures.

5.8.3 Techniques Used to Suppress Short-Channel Effects in Buried-Channel PMOS

5.8.3.1 Techniques Used to Keep the Implanted p -Layer as Thin as Possible. The most obvious solution to suppressing short-channel effects in PMOS devices is to increase the PMOS device channel length. This is done in many CMOS technologies, and it is the reason that the minimum channel lengths of PMOS devices are often larger than those of the NMOS devices on the same chip. But we also noted that to achieve small values of S_t , less V_T -roll-off, and larger punchthrough voltages in BC PMOS devices, the p -buried layer depth Y_j must be as shallow as possible. That is, the thinner the p -layer, the more the device resembles a surface channel FET. Several methods have been reported to realize shallow p -layer implant profiles, including the following:

1. Using the lowest practical implant energy, which in ordinary medium-current implanters is typically 20-25 keV.
2. Implanting BF_2^+ (which produces shallower boron layers than implantation with elemental boron, see Vol. 1, chap. 9).
3. Using a high-energy n -implant (e.g., P^+ at 400 keV) to form a retrograde n -well in the regions where the PMOS devices are formed. This places more n -type dopant atoms beneath the channel pn junctions, and the more heavily doped regions absorb the drain voltage in a shorter distance, while simultaneously squeezing the channel pn junction toward the surface, as shown in Fig. 5-68.

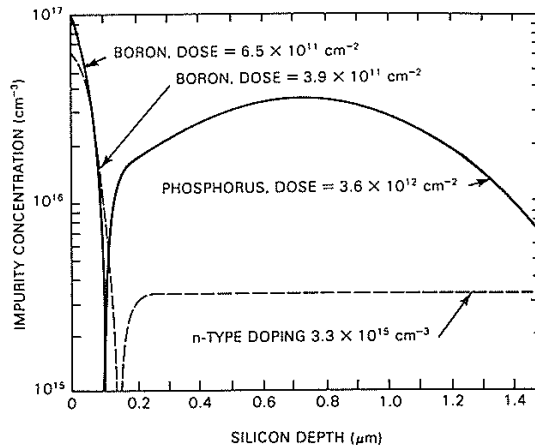


Fig. 5-68 Simulated doping profiles in PMOS channel regions: dashed lines lightly doped substrate; solid lines, retrograde n -well.⁷² (© 1982 IEEE).

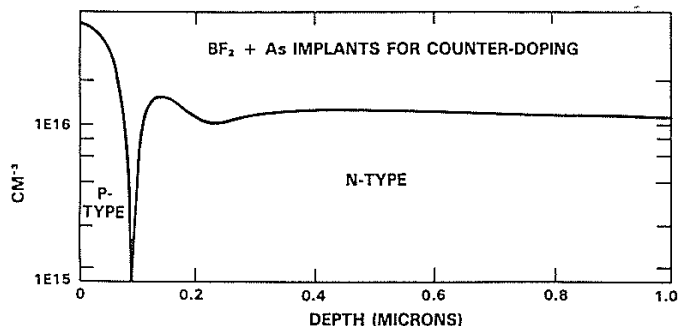


Fig. 5-69 (a) Simulated channel profile for *p*-channel with BF_2 and As counterdoping.¹¹⁵ (© 1984 IEEE).

4. An extra *n*-type implant (typically arsenic) can be performed into a conventional *n*-well or *n*-substrate, with its peak located below the *p*-layer immediately following the BF_2^+ implant (Fig. 5-69a). The *n*-doping under the *p*-layer is thus raised, making the junction depth lie further up the tail of the boron distribution (and thereby making it shallower). The arsenic diffuses relatively slowly, and thus will largely remain in the vicinity of where it was implanted. This forms a steep profile at the counter-doping junction, and maintains a shallow junction depth. The simulation shown in Fig. 5-69a indicates that Y_j with such an arsenic implant is $\sim 0.09 \mu\text{m}$. Such an extra implant can also be used with a retrograde *n*-well.

Two drawbacks of this technique are that since the boron channel implant is compensated by an arsenic implant very good process control must be maintained for it to be successful, and it also requires an extra masking step to prevent the arsenic implant from entering the *n*-channel transistor regions.

To prevent a shallow implanted-boron layer from growing thicker, it is necessary to use a reduced thermal budget in order to restrict boron diffusion following the implant. Specific steps for restricting boron redistribution include the following:⁷⁶

1. Implanting the boron through the gate oxide. This avoids the oxidation-enhanced diffusion of boron that would occur during the growth of the gate oxide if the implant were performed first. In this case, it is necessary to prevent the gate oxide from becoming contaminated during implant, either by material sputtered by the beam line or by vaporized resist material used as a mask against the implant. To prevent such contamination, a thin layer of polysilicon may be

deposited on the gate oxide prior to the implant (in fact, immediately after the oxide is grown).⁷⁷

2. After the implant has been performed, the remainder of the polysilicon film is deposited. In order to avoid the 900°C thermal cycle that would have to be used if the poly were doped following deposition, this polysilicon is doped *in situ* at ~600°C with phosphorus.

3. A BPSG glass layer is used as the dielectric between the gate and the first level of metal. A significantly lower temperature can be used to flow BPSG than PSG (e.g., 900°C versus 1000°C). A lower temperature cycle can thus be used to smooth the surface topography (flow step) and gently taper the contact holes after etch (reflow step).

5.8.3.2 Shallow p^+ Source/Drain Junctions to Suppress Punchthrough in PMOSFETs. As noted in our discussion of MOSFET scaling in section 5.6.1.2, reducing the source/drain junction depths also decreases the susceptibility to short-channel effects. By reducing the vertical depth of the source/drain junctions the lateral spread of the dopants in these regions beneath the gate also becomes smaller. Hence, a longer effective channel length is possible for a fixed drawn gate length. In addition, the lateral spread of source and drain depletion regions below the surface layer makes a MOSFET more prone to punchthrough. Consequently, shallow source and drains effectively suppress subsurface punchthrough paths.

In NMOSFETs, it is easier to fabricate shallow n^+ source/drain regions (i.e., $\leq 0.2 \mu\text{m}$ in depth) because dopants with heavy ionic masses (usually arsenic) are implanted to form these regions. They yield implant profiles with small projected ranges and little channeling. Furthermore arsenic has a low diffusivity, so that little redistribution occurs during postimplant annealing. In PMOSFETs, however, the only practical p -type dopant for forming the p^+ regions is boron, and junctions formed with boron are much deeper than those formed with arsenic for several reasons. First, the projected range of the low-mass boron is relatively larger, even at low implantation energies. Ion channeling also causes the implanted boron profile to have a long tail, even if the target wafers are tilted to minimize this effect. Thus, the minimum junction depth which results when boron is implanted into a lightly doped n -type crystalline silicon wafer is about $0.35 \mu\text{m}$.

The high diffusivity of boron compounds the problem. That is, in CMOS processes both the n - and p -implants that form the NMOS and PMOS source and drain regions, respectively, must be annealed. The anneal used for the arsenic implant will cause significant ($>0.1 \mu\text{m}$) redistribution of the boron implant. In CMOS technologies with channel lengths of $2 \mu\text{m}$ or longer, this issue does not arise because the boron can be implanted after the arsenic implantation and anneal steps (i.e., allowing the boron to be annealed only for the minimum time and temperature needed for full activation). Thus, boron redistribution after implantation can be minimized. However, in submicron CMOS processes it is generally not possible to perform the boron implant after the n^+ arsenic implant. This is because submicron NMOS devices usually use LDD structures,

in which the arsenic implant is performed after spacers have been formed. The boron p^+ implant, however, must be carried out before the spacers are formed for two reasons. First, this ensures that the boron implant sufficiently counterdopes the n^- implant (used to form the LDD tip region) everywhere in the areas of the PMOS source and drain (i.e., the n^- implant is typically done unmasked to maintain process simplicity). Second, this also guarantees that the p^+ doping will extend under the gate. Thus, in CMOS processes that use LDDs in the NMOS devices the p^+ regions must experience the same anneal cycle as the n^+ regions. Alternative approaches have been proposed to avoid subjecting the implanted boron to the n^+ anneal, as will be described below.

Techniques to decrease the p^+ junction depths fall into two groups: (1) those which set out to reduce the junction depth of the implanted boron profile; and (2) those which attempt to reduce the diffusion of the boron during the post-implantation anneals.

The approaches mentioned earlier when discussing techniques to reduce the depth of the as-implanted boron region in the channel of PMOS devices also apply to reducing the source/drain junction depths. These include use of BF_2^+ instead of atomic boron as the implant species, and implanting through a screen oxide. Pre-amorphizing the wafer surface region into which the junction dopants will be implanted to a depth of several tenths of a micron (with an earlier implant of silicon or germanium) has also been investigated (see Vol. 2, chap. 3). The amorphous layer effectively prevents channeling, and subsequent annealing at a low temperature is sufficient to achieve recrystallization by solid-phase epitaxy. Use of very-low-energy ion implanters (<5 keV) for implanting the boron has also been studied, as have techniques for implanting the boron into silicide layers formed above the source/drain regions and then diffusing the boron into the substrate from these silicide layers.

Several techniques to limit the redistribution of boron after implant have also been suggested. The first involves use of rapid thermal processing instead of a conventional furnace treatment following the implant. Relatively high temperatures are employed to realize adequate dopant activation, while keeping the times short enough so that boron redistribution is minimized. The second makes use of lower temperature flow and reflow steps by employing BPSG as the interlevel dielectric material instead of PSG. The third modifies the process sequence so that the boron implant is done after the arsenic implant. If the spacers required for NMOS LDD formation are in place, the boron implant must be sufficiently redistributed to ensure that the p -regions extend under the p -channel gates.¹⁰⁵ On the other hand, some process sequences have been reported in which the spacers are removed prior to implanting the boron. Other materials than SiO_2 must be used for such removable spacers (e.g., silicon nitride,¹⁰² TiN,¹⁰³ or polysilicon¹⁰⁴) since removing SiO_2 spacers would significantly thin the field oxide. Concerns about damaging the gate oxide at the gate edge have been raised if dry etching is used for such spacer removal. In the reports describing silicon nitride and TiN removable spacers, wet etching is used to remove them.

5.8.3.3 LDD Structures in PMOS to Limit Punchthrough. Another way to achieve a shallow junction in a PMOS device is to use an LDD structure. The junction depth of the lightly-doped LDD tip region is shallower than the heavily doped

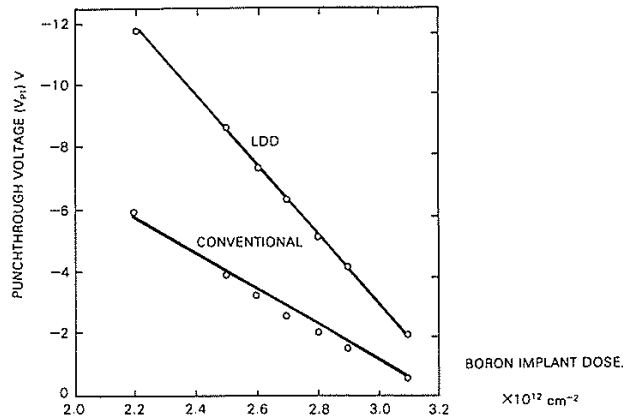


Fig. 5-70 Punchthrough voltage as a function of boron implant dose for a $0.4 \mu\text{m}$ PMOSFET with and without LDD structure.⁷⁵ (© 1986 IEEE).

p^+ source/drain junction depth implanted at the same energy. Figure 5-70 indicates that the punchthrough voltage V_{PT} is increased by 2-3V when an LDD is incorporated into a PMOS device. (The LDD structure, of course, may also be used to increase the hot-carrier resistance of the PMOS devices, especially as the channel length is decreased below about $0.8 \mu\text{m}$.) The LDD, however, will generally reduce I_D due to a rise in the series resistance of the drain region. Thus, a tradeoff analysis between an increase in V_{PT} and a decrease in the drain current must be made.

Adding an LDD structure to a PMOS device in CMOS involves relatively straightforward changes to the process sequence. The gate sidewall spacers are already available from the creation of NMOS LDDs, but there is some increase in process complexity, (especially if the former process used a maskless n^- implant to form the tip regions of the NMOS LDD). That is, to form both PMOS and NMOS LDDs four masking steps are generally needed (n^- , p^- , n^+ , p^+), rather than just two. However, a two-mask process, which employs removable polysilicon spacers, has also been reported.¹⁰⁴

5.8.3.4 New PMOS Device Structures Designed to Suppress Punchthrough. New PMOS device structures have also been proposed to minimize punchthrough current. For example, a small segment of an n -type region can be created by ion implantation between the p -type S/D and the channel as a punchthrough stopper (refer to Fig. 5-71). This structure has been termed a double-implanted or *halo structure*. One report of such a halo device with $L = 0.5 \mu\text{m}$ used a phosphorus implant dose of $1 \times 10^{12} \text{ cm}^{-2}$ at 130 keV.⁷⁸ A sidewall spacer was used to form the lateral offset of the n -type region in this device. However, since the p^+ source/drain is offset from the gate by the sidewall spacer, this method of forming the local n -regions degrades the drive current and impairs hot-electron immunity.⁷⁹ Furthermore, as this technique also places a highly doped n -region adjacent to a highly doped p -region, the maximum halo doping

is determined by the onset of low-level avalanche at the chosen supply voltage and channel length.

A large-tilt-angle implant can be used to fabricate halo-type structures without a sidewall spacer, thereby gaining the benefits while avoiding the problems of the structure of Fig. 5-71.⁷⁹ The dopant is placed precisely between the p^+ source and the buried p -channel junction. The lateral penetration and depth of the n -type region are controlled by the implant energy and tilt-angle. The peak concentration is set by the dose. A typical implant sequence to form such a *large-angle tilt implanted punchthrough stopper* (LATIPS) structure as shown in Fig. 5-72 is: (1) implant phosphorus at 90 keV with a dose of $2 \times 10^{13} \text{ cm}^{-2}$ at 25° tilt, in two segments, with the target rotated 180° between segments; then (2) implant BF_2^+ (source/drain) at 40 keV (no tilt) to a dose of $3 \times 10^{15} \text{ cm}^{-2}$.⁸⁰

Devices fabricated using the LATIPS process were compared to conventional PMOSFETs. LATIPS devices showed dramatically improved punchthrough characteristics for physical gate lengths as small as 0.5 microns. As an example, the OFF-state leakage current was eight orders of magnitude smaller (at -3V drain bias) than that of a conventional control device fabricated without the LATIPS structure. This dramatic improvement in punchthrough resistance was achieved without extreme scaling of the source/drain junction depth or the n -well background doping. That is, the p^+ source/drain junction depth was about $0.2 \mu\text{m}$ and the n -well concentration was $1 \times 10^{16} \text{ cm}^{-3}$. Furthermore, the benefits were obtained without significant degradation of device performance. That is, no gate-oxide degradation was observed as a result of using BF_2^+ source/drain implants at 8 keV, even at doses of $5 \times 10^{15} \text{ cm}^{-2}$. Better high-current drive capability than conventional devices was also exhibited, as well as a smaller body (back-gate bias) effect.

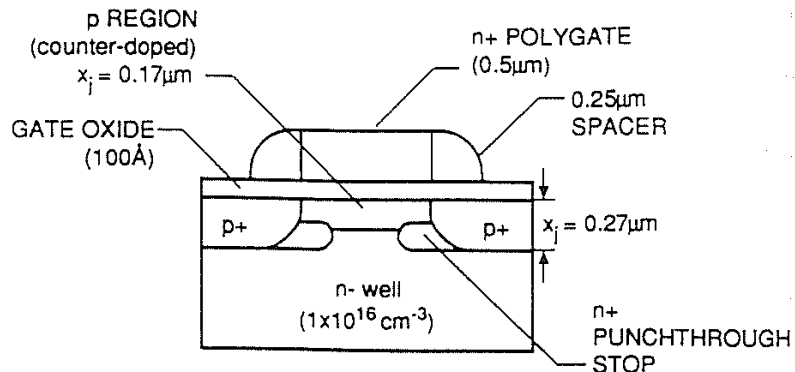


Fig. 5-71 PMOS Punchthrough-stop structure.⁷⁸ (© 1985 IEEE).

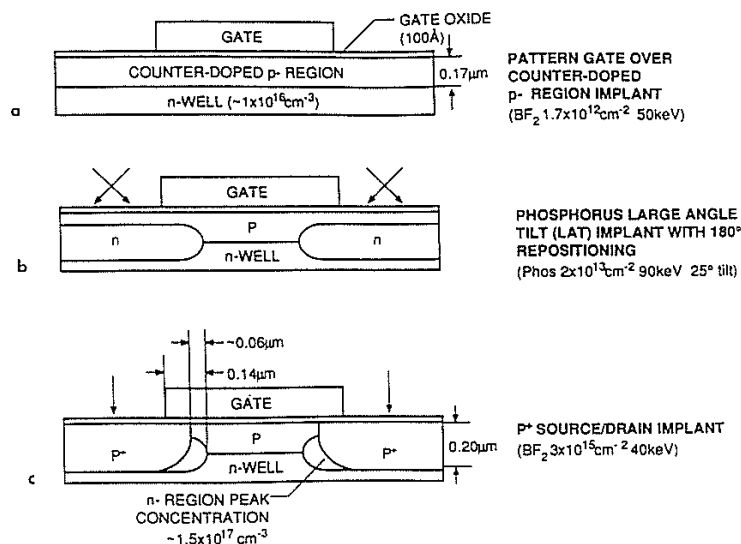


Fig. 5-72 LATIPS process sequence for PMOSFET: (a) Pattern gate over an n -well region that has been counterdoped with a p -type implant (BF_2); (b) Large angle tilt implant with phosphorus; (c) p^+ Source/Drain implant (BF_2).⁷⁹ (© 1988 IEEE).

5.8.4 PMOS Devices with p^+ -Polysilicon Gates

The use of a boron implant to adjust V_{TP} becomes less feasible as devices use even thinner gate oxides, since larger doses of boron are needed. Y_j thus becomes deeper, and the punchthrough problem worsens. In addition, buried channel devices are subject to more process variability stemming from the difficulty of controlling the compensating implant. Solutions involving the use of gate electrodes other than n^+ polysilicon must therefore be explored. One alternative is to use p^+ polysilicon for PMOS devices. This approach appears to become necessary for CMOS and BiCMOS once L becomes smaller than 0.5 μm .

When p^+ polysilicon is used for the gate material, V_{TP} is shifted from the values that occur when n^+ polysilicon is used. Figure 3-20 also plots V_{TP} as a function of substrate doping when p^+ polysilicon is employed (using the scale on the right side of the figure). V_{TP} can be seen to be less negative than -0.7 V over the substrate doping range of 10^{15} - 10^{17} cm^{-3} . Thus, V_{TP} can easily be made more negative through the implantation of phosphorus or arsenic into the channel. If p^+ polysilicon is used throughout the circuit, however, the NMOS device then has to be overcompensated in order for V_{TH} to be reduced to sufficiently small values. (Note that some reports have been published on the use of p^+ polysilicon alone.^{81,82}) This implies it would be ideal

to use both n^+ and p^+ poly gates on the same chip (with n^+ poly for NMOS devices, and p^+ poly for PMOS devices).⁸³

Such a dual-doped poly approach, however, adds process complexity, and also introduces other problems. One drawback of is that the sheet resistance of p^+ poly is $\sim 3\times$ higher than that of an n^+ poly film of the same thickness, because the mobility of holes is $\sim 3\times$ smaller than that of electrons in Si. A second problem arises from the need to connect the two types of poly (e.g., at the input of an inverter). Such problems occur when a silicide overlayer (or strap) is used to make the connection between n^+ and p^+ poly (as a method to avoid a separate, space-consuming metal contact). Because the silicide strap provides a very rapid diffusion path for boron and arsenic, one type of poly can be counterdoped by the other when the device is subjected to high-temperature excursions. This counterdoping can occur to the degree that a region of poly can change doping types (i.e., from n^+ to p^+). In such cases, the threshold voltage of devices with counterdoped poly will be shifted from their designed value.

If the processing temperature is limited to 800°C after the two types of poly have been connected by the silicide, such lateral diffusion does not produce significant shifts in V_T .⁸⁴ On the other hand, temperatures of 900°C will produce sufficient counterdoping to significantly shift V_T . Since one of the last high temperature steps in CMOS is activation of the source/drain implants, this would mean either using a lower-temperature activation step, or performing the step prior to formation of the silicide layer.

Another approach is to form a polysilicon electrode with an overlying conductor layer that suppresses such counterdoping. One such structure is a W-TiN-poly electrode structure.⁸⁵ The thin (30-nm) TiN film acts as a diffusion barrier to the dopants in the poly and also prevents reaction between the poly and the W. Very little of the dopant present in the poly is found to diffuse into the W, even following an anneal at 900°C for 1 hour.

Another problem encountered with p^+ poly gates when a thin gate oxide is used is poor V_T process control in the PMOS devices, due to penetration of the boron into the oxide (or further, into the silicon). It is reported that boron will penetrate gate oxides that are ≤ 12.5 nm thick during a 900°C 30-minute post-implant anneal in N_2 .⁸⁶ This implies that if p^+ poly gates are used, a lower processing temperature may be needed. If too low a temperature is used, however, the boron implanted into the polysilicon will not be sufficiently redistributed. The polysilicon dopant concentration at the polysilicon-gate oxide interface could thus end up being less than the mid- $10^{19}/\text{cm}^3$. This would make the work function of the polysilicon different from the desired degenerately doped value, creating V_T control problems in the MOS devices.

It has also been found that the presence of fluorine in the gate oxide worsens the boron penetration problem (Fig. 5-73).^{87,88} Such fluorine can be introduced into the gate oxide if the PMOS source drain regions are implanted using BF_2^+ . Elemental boron is therefore considered inherently superior to BF_2^+ as the implant species for surface-channel PMOS devices in CMOS technologies that use p -doped polysilicon. A

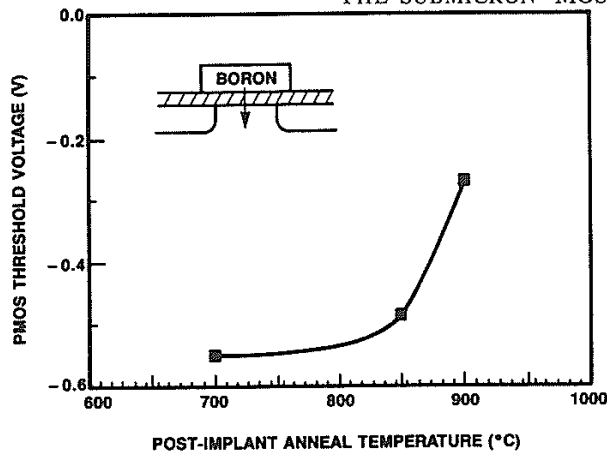


Fig. 5-73 Boron penetration through the gate oxide causes V_T in BF_2 implanted PMOSFETs to shift positive at anneal temperatures above $\sim 800^\circ\text{C}$.⁸⁷ (© IEEE 1989).

study of enhanced boron diffusion through thin SiO_2 layers in a wet oxygen atmosphere is also reported in reference 89.

It also appears that nitrided oxides and reoxidized nitrided oxides (see chap. 9) are effective in blocking boron diffusion.⁹⁰ For this, and other reasons, they are being extensively studied for use as candidates to replace SiO_2 as the gate insulators in dual-gate CMOS processes. An alternative approach to introducing nitrogen into the gate oxide to block boron penetration is through ion implantation. Recently, ion implantation of nitrogen into the gate p - and n -doped poly electrodes of a $0.25\text{ }\mu\text{m}$ CMOS process has been reported. The heavy nitrogen dose ($1 \times 10^{15}\text{ cm}^{-2}$) resulted in the gate oxide incorporating high concentrations of nitrogen. This had the effect of suppressing boron penetration of the gate oxide.¹¹⁰

Hot carrier effects in p^+ -poly $0.5\text{-}\mu\text{m}$ PMOSFETs have also been examined. Stress experiments showed significantly more electron trapping in short stress times than is observed in NMOSFETs. The device degradation due such trapping can be severe enough that it may pose a problem in such devices.¹¹⁸

REFERENCES

1. J.E. Chung et al., "Performance and Reliability Design Issues of Deep Submicron MOSFETs," *IEEE Trans. Electron Devices* **ED-38**, March 1991, p. 545.
2. J. Greenfield and R. Dutton, "Nonplanar VLSI Device Analysis Using the Solution of Poisson's Equation," *IEEE Trans. Electron Dev.*, **ED-27**, p. 1520 (1980).
3. L.D. Yau, "A Simple Theory to Predict the Threshold Voltage in Short-Channel IGFETs," *Solid-State Electronics*, **17**, p. 1059, (1974).

4. W. Fichtner and H.W. Poetzel, "MOS Modelling by Analytical Approximations - Sub-threshold Current and Subthreshold Voltage," *Int. J. Electronics*, **46**, p. 33, Jan. 1979.
5. C.R. Viswanathan, B.C. Burkery, G. Lubberts, and T.J. Treadwell, *IEEE Trans. Electron Dev.*, **ED-32**, p. 932, (1985).
6. R.R. Troutman, "VLSI Limitations from Drain-Induced Barrier Lowering," *IEEE Trans. Electron Dev.*, **ED-26**, p. 461, (1979).
7. F.-C. Hsu et al., *IEEE Trans. Electron Dev.*, **ED-30**, p. 571 (1983).
8. S.C. Jain and P. Balk, "A Unified Analytical Model for DIBL ... in Short-Channel MOSFETs," *Solid-State Electronics*, **30** (5), p. 503, May 1987.
9. T. Skotnicki et al., *IEEE Electron Dev. Letts.*, **EDL-9**, p. 109, March 1988.
10. G.W. Taylor, *Solid-State Electronics*, **22**, p. 701, (1979).
11. B.J. Sheu, D.L. Scharfetter, P.K. Ko, and M.-C. Jeng, "BSIM: Berkeley Short-Channel IGFET Model for MOS Transistors," *IEEE J. of Solid State Circuits*, August 1987, p. 558.
12. G. Merkel, *Process and Device Modeling for IC Design*, (F. Van de Wield et al., Eds.), Noordhoff, Leyden, (1977), p. 705.
13. L. A. Akers et al., "A Model of a Narrow-Width MOSFET including Tapered Oxide and Doping Encroachment," *IEEE Trans. Electron Dev.* **ED-28**, p. 1490, 1981.
14. C.R. Ji and C.T. Sah, "Two-Dimensional Analysis of the Narrow-Gate Effect in MOSFET," *IEEE Trans. Electron Dev.*, **ED-30**, P. 635, June 1983.
15. T. Furukawa and J.A. Mandelman, "Process and Device Simulation of Trench Isolation Corner Parasitic Device," *Ext Abs. Mtg. Electrochem. Soc.*, Fall 1988, p. 329.
16. L. A. Akers, "The Inverse Narrow-Width Effect," *IEEE Electron Dev. Letts.*, **EDL-7** (7), p. 419, July 1986.
17. E.H. Li et al., "The Narrow-Channel Effect in MOSFETs with Semi-Recessed Oxide Structures," *IEEE Trans. Electron Dev.*, **ED-37** (3) p. 692, March 1990.
18. S. S. Chung and T.-C. Li, "An Analytical Threshold-Voltage Model of Trench-Isolated MOS Devices with Nonuniformly Doped Substrates," *IEEE Trans. on Electron Dev.*, March 1992, p. 614.
19. K. Ohe et al., "Narrow-Width Effects of Shallow Trench-Isolated CMOS with n^+ -Polysilicon Gate," *IEEE Trans. Electron Dev.*, **ED-36** (6), p. 1110, June 1989.
20. G. Merkel, "A Simple Model of the Threshold Voltage of Short and Narrow Channel MOSFETs," *Solid-State Electronics.*, **23**, p. 1207, December 1983.
21. S. Selberherr et al., "MINIMOS - A Two-Dimensional MOS Transistor Analyzer," *IEEE J. Solid State Circuits*, **SC-15**, p. 605 (1980).
22. Z.-H. Liu et al., "Threshold Voltage Model for Deep-Submicrometer MOSFETs," *IEEE Trans. Electron Dev.*, **ED-40**, January 1993, p. 86.
23. J.R. Brews, W. Fichtner, E.H. Nicollian, and S.M. Sze, "Generalized Guide for MOSFET Miniaturization," *IEEE Electron Dev. Letts.*, **EDL-1**, p. 2, 1980.
24. C. Hu, "Future CMOS Scaling and Reliability," *Proceedings of the IEEE*, **81**, p.682, May 1993.
25. F.M. Klaassen, "Design and Performance of Micron-Sized Devices," *Solid-State Electron.*, **21**(3), p. 565 (1978).
26. C.F. Codella and S. Ogura, "Halo Doping Effects in Submicron DI-LDD Device Design," *Tech. Dig. IEDM*, p. 230, 1985.

27. A. Hori et al., "A Self-Aligned Pocket Implantation (SPI) Technology for 0.2 μm -Dual Gate CMOS," *Tech. Dig. IEDM*, p. 641, 1991.
28. H. Nihara et al., "Anomalous Drain Current in NMOS and its Suppression by Deep Ion Implantation," *Tech. Dig. IEDM*, p. 487 (1978).
29. L.C. Parrillo, "CMOS Active and Field Device Fabrication," *Semicond. International*, April 1988, p. 64.
30. T. Skotnicki, G. Merkel, and T. Pedron, "A New Punchthrough Current Model Based on the Voltage-Doping Transformation," *IEEE Trans. on Electron Dev.*, **ED-35**, July 1988, p. 1076.
31. C. Duvvury, "A Guide to Short-Channel Effects in MOSFETs," *IEEE Circuits and Systems Magazine*, November 1986, p. 6.
32. J. A. Cooper and D.F. Nelson, "Measurement of the high-field drift velocity of electrons in the inversion layer in silicon," *IEEE Electron Dev. Lett.*, **EDL-2**, July 1983, p. 169.
33. R.H. Crawford, *MOSFETs in Circuit Design*, McGraw-Hill, New York, 1967.
34. D. Frohman-Bentchkowsky, "On the effect of mobility in inversion and accumulation layers on thermally oxidized silicon surfaces," *Proc. IEEE*, vol. **56**, p. 217, 1968.)
35. A.G. Sabnis and J.T. Clemens, "Characterization of Electron Velocity in the Inverted $\langle 100 \rangle$ Si Surface," *Tech. Dig. - IEDM*, p. 19, 1979.
36. M.S. Liang et al., "Inversion layer capacitance and mobility of very thin gate oxide MOSFETs," *IEEE Trans. Electron Dev.*, **ED-33**, March 1986, p. 409.
37. K.K. Hung, *Electrical characterization of the Si-SiO₂ interface for thin oxides*, Doctoral Thesis, Hong Kong University, (1987).
38. C.G. Sodini, T. Ekstedt, and J. Moll, "Charge Accumulation and Mobility in Thin Dielectric MOS Transistors," *Solid-State Electronics*, **25**, p. 833, (1982).
39. K. Lee et al., "Physical Understanding of Low Field Carrier Mobility in Silicon Inversion Layer," *IEEE Trans. Electron Devices*, **ED-38**, August 1991, p. 1905.
40. S.R. Hofstein and G. Warfield, "Carrier mobility and current saturation in the MOS transistor," *IEEE Trans. Electron Dev.*, **ED-12**, p. 129, (1965).
41. F.N. Trofimenkoff, "Field-dependent mobility analysis of the field-effect transistor," *Proc. IEEE*, **53**, p. 1765, (1965).
42. D.M. Caughey and R.E. Thomas, "Carrier mobilities in silicon empirically related to doping and field," *Proc. IEEE*, **55**, p. 2192 (1967).
43. C.G. Sodini, P.K. Ko, and J.L. Moll, "The effect of high fields on MOS device and circuit performance," *IEEE Trans. Electron Dev.*, **ED-31**, October 1984, p. 1386.
44. E.A. Talkhan, I.R. Mansour, and A.I. Barboor, "Investigation of the effect of drift-field-dependent mobility on MOSFET characteristics," Parts I and II. *IEEE Trans. Electron Dev.* **ED-19**, August 1972, p. 899.
45. P.K. Ko, "Approaches to Scaling," Ch. 1, in *Advanced MOS Device Physics*, N.G. Einspruch and G. Gildenblatt, Eds. Academic Press, Orlando, FL., 1989, p. 1
46. K.Y. Toh, P.K. Ko, and R.G. Meyer, "An engineering model for short-channel MOS devices," *IEEE J. Solid-State Circuits*, August 1988, p. 950.
47. M.C. Jeng et al., "Performance and reliability of deep-submicron MOSFETs," *Tech. Dig. IEDM*, 1987, p. 710.

48. V.G. Reddi and C.T. Sah, "Source to Drain Resistance Beyond Pinchoff in Metal-Oxide-Semiconductor Transistor (MOST)," *IEEE Trans. Electron Dev.*, **ED-12**, p. 108, (1969).
49. D. Frohman-Bentchkowsky and A.S. Grove, *IEEE Trans. Electron Devices* **ED-16**, p. 108, (1969).
50. G. Baum and H. Beneking, "Drift Velocity Saturation in MOS Transistors," *IEEE Trans. Electron Devices*, **ED-17**, June 1972, p. 481.
51. Y.A. Elmansy and A.R. Boothroyd, "A simple two-dimensional model for IGFET operation in the saturation region," *IEEE Trans. Electron Dev.*, **ED-24**(3), p. 254-62, (1977).
52. P.K. Ko, R.S. Muller, and C. Hu, "A Unified Model for Hot-Electron Currents in MOSFETs," *Tech. Dig. IEDM*, p. 600, 1981.
53. P.K. Ko, *Hot-electron effects in MOSFETs*, " Doctoral Thesis, University of California, Berkeley, (1982).
54. R.S. Muller and T.I. Kamins, *Device Electronics for Integrated Circuits*, John Wiley & Sons, New York, 1986, p. 491.
55. T.Y. Chan, P.K. Ko, and C. Hu, "Dependence of Channel Electrical Field on Device Scaling," *IEEE Electron Dev. Lett.*, **EDL-6**, October, 1985, p. 531.
56. B. J. Sheu et al., "BSIM: Berkeley Short-Channel IGFET Model for MOS Transistors," *IEEE J. Solid-State Circuits*, Aug. 1987, p. 558.
57. B. J. Sheu, D. L. Scharfetter, and H. C. Poon, "Compact short-channel IGFET Model," Univ. of Calif., Berkeley, CA Memo. UCB/ERL M84/20, March 1984.
58. M.-C. Jeng, *Design and Modeling of Deep-Submicron MOSFETs*, ERL Memo No. ERL M90/90, Electronics Research Laboratory, University of California, Berkeley, October 1990.
59. A.H.-C. Fung, "A Subthreshold Conduction Model for BSIM," Univ. of Calif., Berkeley, Memo, UCB/ERL M65/22, March 1985.
60. B. Eitan and D. Frohman-Bentchkowsky, "Surface Conduction in short-channel MOS Devices as a limitation to VLSI scaling," *IEEE Trans. Electron Dev.* vol. **ED-29**, p. 254, Feb. 1982.
61. R. H. Dennard, et al., "Design of II MOSFETs with Very Small Physical Dimensions," *IEEE J. Solid-State Circuits*, **SC-9** (5), 256, (1974).
62. G. Bacarani, M. R. Wordeman, and R. H. Dennard, "Generalized Scaling Theory and its Application to a 0.25 μm MOSFET Design," *IEEE Trans. Electron Dev.*, **ED-31**(4), 452 (1984).
63. J.R. Brews et al., "Generalized Guide to MOSFET Miniaturization," *IEEE Electron Dev. Letts.*, **EDL-1**, January 1980, p. 2.
64. F.M. Klaassen, "Design and Performance of Micron-Size Devices," *Solid-State Electronics*, **21**, March 1978, p. 565.
65. J.R. Brews, K.K. Ng, and R.K. Watts, in "The Submicron Silicon MOSFET," Ch. 1, p. 19, *Submicron Integrated Circuits*, Ed. R.K. Watts, Wiley Interscience, New York, 1989.
66. *ibid.* pp. 23- 37.
67. R.A. Haken et al., "BiCMOS Process Technology," Chap. 3, p. 85, in *BiCMOS Technology and Applications*, A. Alvarez Ed. Kluwer, Norwell MA, 1989.

68. A. Alvarez et al., "Application of Statistical Design and Response Surface Methods to Computer-Aided VLSI Device Design," *IEEE Trans. Computer-Aided Design, CAD-7*, February 1988, p. 272.
69. *ibid.* ref. 65, p. 24.
70. S.M. Sze, Ed. *VLSI Technology*, 2nd edition, chap. 11, McGraw-Hill, New York, 1988.
71. T. Ohzone et al., *IEEE Trans. Electron Dev.*, **ED-32**, 1789, (1980).
72. G.J. Hu et al., *Tech. Dig. IEDM*, 1982, p. 710.
73. K.M. Cham et al., *Computer-Aided Design and VLSI Device Development*, Boston, Kluwer Academic Publishers, 1986, p. 182.
74. J. Zhu, R.A. Martin, and J.Y. Chen., "Punchthrough Current for Submicrometer MOSFETs in CMOS VLSI," *IEEE Trans. Electron Dev.*, February 1988, p. 145.
75. A. Schmitz and J. Y. Chen, "Design, Modelling, and Fabrication of Submicron CMOS Transistors," *IEEE Trans. Electron Dev.*, **ED-33**, p. 148, (1986).
76. L.C. Parrillo, "CMOS Active and Field Device Fabrication," *Semicond. Internat.*, April 1988, p. 64.
77. R. A. Chapman et al., "A 0.8 μm CMOS Technology," *Tech. Dig. IEDM*, 1987, p. 362.
78. S. Odanaka et al., "A New Half-Micron P-Channel MOSFET with Efficient Punchthrough Stops," *1985 Symp. VLSI Tech., Digest of Tech. Papers*, p. 62.
79. T. Hori et al., "A New P-Channel MOSFET with Large Tilt Angle Implanted Punch-through Stopper (LATIPS)," *IEEE Electron Dev. Letts.*, **EDL-9**, Dec. 1988, p. 641.
80. M. Stinson et al., "Effects of Ion Implantation on Deep-Submicrometer Drain-engineered MOSFET Technologies," *IEEE Trans. Electron Dev.*, **ED-38**, March 1991, p. 487.
81. L. C. Parrillo et al., *Tech Dig. IEDM*, 1984, p. 418.
82. K. M. Cham et al., *IEEE Electron Dev. Lett.*, January 1986, p. 49.
83. C.Y. Wong et al., *Tech. Dig. IEDM*, 1988, p. 238.
84. S. J. Hillenius et al., "A Symmetric Submicron CMOS Technology," *Tech. Dig. IEDM*, 1986, p. 252.
85. P.-H. Pan, J.G. Ryan, and M.A. Lavoie, *Ext. Abs. of the Electrochem. Soc. Meeting*, Spring 1989, Abs. 138, p. 193.
86. M.L. Chen et al., "Constraints in p-Channel Device Engineering for Submicron CMOS Technologies," *Tech. Dig. IEDM*, 1988, p. 390.
87. F.K. Baker, et al., *Tech. Dig. IEDM*, 1989, p. 443.
88. J.M. Sung, *Tech. Dig. IEDM*, 1989, p. 447.
89. Y. Sato, K. Ehara, and K. Saito, *J. Electrochem. Soc.*, June 1989, p. 1777.
90. H. Fang et al., "Low-Temperature Furnace-Grown Reoxidized Nitrided Gate Dielectrics as a Barrier to Boron Penetration," *IEEE Electron Dev. Letts.*, **EDL-13** (4) p. 217 April 1992.
91. M. Nishida and H. Onodera, "An anomalous increase of threshold voltages with shortening of the channel lengths fro deeply boron-implanted n-channel MOSFETs," *IEEE Trans. Electron Dev.*, **ED-28**, p. 1101, 1981.
92. M. Orlowski, C. Mazure, and F. Lau, "Submicron short-channel effects due to gate re-oxidation induced lateral diffusion," *IEDM Tech. Dig.* 1987, p. 632.
93. C.Y. Lu and J.M. Sung, "Reverse Short-Channel Effects on Threshold Voltage in Submicron Salicide Devices," *IEEE Electron Dev. Lett.*, **EDL-10**, p. 446, October, 1989.

94. N.D. Arora and M.S. Sharma, "Modeling the Anomalous Threshold Voltage Behavior of Submicrometer MOSFETs," *IEEE Electron Dev. Lett.*, **EDL-13**, p. 92, February 1992.
95. H. Hanafi et al., "A Model for Anomalous Short-Channel Behavior in MOSFETs," *IEEE Electron Dev. Lett.*, **EDL-14**, p. 575, December 1993.
96. D. Sadana et al., "Enhanced Short-Channel Effects in NMOSFETs due to Boron Redistribution Induced by Arsenic Source and Drain Implant," *IEDM Tech. Dig.* 1992, p. 849.
97. H. Jacobs et al., "MOSFET Reverse Short Channel Effect Due to Silicon Interstitial Capture in Gate Oxide," *IEDM Tech. Dig.* 1993, p. 307.
98. C.S. Rafferty et al., "Explanation of Reverse Short Channel Effect by Defect Gradients," *IEDM Tech. Dig.* 1993, p. 311.
99. M.D. Giles, *Appl. Phys. Lett.*, **62**(2), p. 1940, (1993).
100. J.Y. Chen, "An *n*-well CMOS with self-aligned channel stops," *Tech. Dig. IEDM* 1983, p. 526.
101. T.M. Liu and W.G. Oldham, *IEEE Electron Dev. Letts.*, **EDL-5**, 299 (1984).
102. J.R. Pfister et al., "A Poly-Framed LDD Sub-half-Micrometer CMOS Technology," *IEEE Electron Dev. Letts.* **EDL-11**, November 1990, p. 529.
103. J.R. Pfister et al., "An Integrated 0.5 μm CMOS Disposable TiN LDD/Salicide Spacer Technology," *Tech. Dig. IEDM*, 1989, p. 781.
104. L.C. Parrillo, "CMOS Active and Field Device Fabrication," *Semiconductor International*, April 1988, p. 64.
105. S. Meguro et al., *Tech. Dig. IEDM*, 1984, p. 59.
106. J.S.T. Huang, J.W. Schrankler, and J.S. Kueng, *IEEE Trans. Electron Dev.*, **ED-31**, 1889, (1984).
107. S. Ogura et al., *IEEE Trans. Electron Dev.*, **ED-27**, 1359 (1980).
108. Y. Omura and K. Ohwada, *Solid-State Electronics*, **24**, 301 (1981).
109. N. Kotani and S. Kawazu, *Solid-State Electronics*, **22**, p. 63, (1979).
110. T. Kuroi et al., "Novel NICE (Nitrogen Implantation into CMOS Gate Electrode and Source-Drain) Structure for High Reliability and High Performance 0.25 μm Dual Gate CMOS," *Tech. Dig. IEDM*, 1993, p. 325.
111. R.R. Troutman, *IEEE J. Solid-State Circuits*, **SC-9**, p. 55 (April 1974).
112. J.Y. Chen et al., "A Fully Recessed Field Isolation Technology Using Photo-CVD Oxide," *Tech. Dig. IEDM*, 1982, p. 233.
113. T. Shibata et al., "An optimally designed process for submicrometer MOSFETs," *IEEE Trans. Electron Dev.* **ED-29**, p. 531, (1982).
114. G.J. Hu and R.H. Bruce, "Design Tradeoffs Between Surface and Buried-Channel FETs," *IEEE Trans. Electron Dev.*, **ED-32**, March 1985, p. 584.
115. K.M. Cham and S.Y. Chiang, "Device design for the Submicrometer *p*-channel FET with n^+ polysilicon gate," *IEEE Trans. Electron Dev.*, **ED-31**, p. 964 (1984).
116. K.M. Cham et al., "Characterization and modeling of the trnch surface inversion problem for the trench isolated CMOS technology," in *IEDM Tech. Dig.*, p. 23, 1983.
117. T. Skotnicki, G. Merkel, and T. Pedron, "Analytical Study of Punchthrough in Buried-Channel PMOSFETs," *IEEE Trans. Electron Dev.*, April 1989, p. 690.
118. Y. Hiruta et al., "Impact of Hot Electron Trapping of Half-Micron PMOSFETs with p^+ Poly -Si Gate," *Tech. Dig. IEDM*, 1986, p. 718.

119. J. Cooper and D. Nelson, "High-Field Drift Velocity of Electrons at the Si-SiO₂ Interface as Determined by the Time-of-Flight Technique," *J. Appl. Phys.* **54**, 1445 (1983).
120. T.E. Hendrikson, "A Simplified Model for Subpinchoff Condition in Depletion Mode IGFETs," *IEEE Trans. Electron Dev.*, **ED-25**, p. 435 (1978).

PROBLEMS

1. Calculate the threshold voltage (a) in 1 Ω -cm p -type silicon and (b) in 1 Ω -cm p -type silicon. The MOS devices for each case have : (i) n^+ -polysilicon gate; (ii) $t_{ox} = 100$ nm; and (iii) the oxide is free of charge except for a surface density of $Q_f = 5 \times 10^{10} \text{ cm}^{-2}$.
2. For exercise 1, what are the threshold voltages if p^+ polysilicon is used for both the n - and p -channel devices.
3. Describe punchthrough current and subthreshold current and explain the difference between them.
4. An n -channel MOSFET has a t_{ox} of 20 nm, an $L_{eff} = 1 \mu\text{m}$, $r_j = 0.2 \mu\text{m}$, and $V_T = 0.6\text{V}$. If the device is biased at $V_{GS} = 3\text{V}$ and $V_{DS} = 5\text{V}$, calculate the saturation voltage V_{DSSat} and the maximum electric field.
5. Use one-dimensional junction theory to estimate the punchthrough voltage of a MOSFET with a channel length of $1 \mu\text{m}$. Assume a substrate doping of $3 \times 10^{16} \text{ cm}^{-3}$, a substrate bias of 0V, and a uniformly doped substrate (no V_T -adjust implant).
6. Using Eq. 5-97, calculate the buried-channel S_t value for $x_1 = 500\text{\AA}$, $t_{ox} = 250\text{\AA}$, $x_2 \ll x_3 = 3000\text{\AA}$ and compare it with the S_t value calculated for a surface-channel MOSFET with the same t_{ox} and a channel depletion width of $d = x_3$.
7. How does a MOSFET's threshold change when constant field scaling rules are applied? Does it scale proportionately?
8. Consider a long-channel MOSFET with $L = 3 \mu\text{m}$, $Z = 21 \mu\text{m}$, $N_A = 5 \times 10^{15} \text{ cm}^{-3}$, $C_{ox} = 1.5 \times 10^{-7} \text{ F/cm}^2$, and $V_T = 1.5\text{V}$. Calculate V_{DSSat} for the case when $V_{GS} = 4\text{V}$. If a constant scaling factor is used to reduce the channel length to $1 \mu\text{m}$, find the following scaled-down parameters: Z , C_{ox} , and I_{psat} .
9. A MOSFET with gate oxide thickness of 20 nm is to be fabricated on a p -type substrate of doping $N_A = 10^{16} \text{ cm}^{-3}$. The source/drain regions have a depth of $0.2 \mu\text{m}$ and $N_D = 10^{19} \text{ cm}^{-3}$. The device is to be operated with $V_{BS} = 0\text{V}$, but V_{DS} could be as large as 3V. From Eq. 5-84 estimate the minimum channel length this MOSFET can have while still exhibiting electrically long-channel behavior. (The gate length is the channel length plus twice the junction depth.) A smaller value of minimum channel length could be obtained if a larger N_A was used. Suggest some undesirable side effects that such a choice could produce.
10. Qualitatively describe the differences between a surface-channel and a buried-channel MOSFET.
11. We have stated in this chapter that the saturation of electron velocity can limit the drain current in a MOSFET with a channel length of $1 \mu\text{m}$ or less operating from a 5-V supply. Examine this effect by using SPICE to generate the I_D - V_{DS} characteristic of an NMOSFET with $L = 1 \mu\text{m}$ formed on a substrate of doping $N_A = 10^{16} \text{ cm}^{-3}$ with a 50 nm thick gate oxide. Set $V_{GS} = 5\text{V}$ and sweep V_{DS} from 0 to 5V using the Level 2 model. First run a simulation in which V_{MAX} is not specified, and then repeat the simulation with $V_{MAX} = 8 \times 10^6 \text{ cm sec}^{-1}$, which corresponds roughly to the scatter-limited velocity of electrons in a MOSFET channel.

320 SILICON PROCESSING FOR THE VLSI ERA - VOLUME III

- 12.** Qualitatively summarize why a surface channel NMOSFET is prone to suffering punchthrough only after the punchthrough voltage is exceeded, but that a buried-channel PMOSFET exhibits punchthrough current behavior even when drain voltages are very small.
- 13.** Why does the subthreshold slope of the I_D - V_{GS} curve appear to increase at the onset of subsurface punchthrough?

CHAPTER 8

WELL FORMATION IN CMOS

Both n - and p -channel transistors must be fabricated on the same wafer in CMOS technologies. Obviously, only one type of device can be fabricated on a given starting substrate. To accommodate the device type that cannot be built on this substrate, regions of a doping type opposite to that present in the starting material must be formed. These regions are called *wells* (or sometimes *tubs*) and are usually (but not always) the first features to be defined on a starting wafer. The well-formation process starts with a uniform, lightly doped p - or n -type substrate. Excess dopants of the opposite type are then selectively implanted or diffused to attain the proper well depth and doping profile in the well regions. The doping type in the wells becomes the identifying characteristic of the CMOS technology (e.g., p -well or n -well CMOS). Conventional well-implant doses range from 2×10^{12} – $8 \times 10^{12}/\text{cm}^2$ at energies from 100–200 keV. The term "conventional" in this sense refers to wells formed by low-energy (<200 keV) ion implantation and a thermal drive-in. Alternative well-formation techniques that do not require a drive-in will also be discussed separately. Epitaxial n -on- n^+ or p -on- p^+ starting substrates have also come into widespread use, since they help reduce the susceptibility of CMOS circuits to latchup.

As well as being of the correct doping type, the well and substrate must be electrically isolated from each other under every operating condition. As noted in chap. 6, to ensure that pn junctions are not forward-biased during circuit operation, if a p -type substrate is used as the starting material of an IC, the substrate must be connected to the most negative circuit voltage. If an n -type substrate is used, it must be connected to the most positive circuit voltage. When wells are present, however, the well regions must also be connected to the appropriate circuit voltages to prevent forward-biasing of the junctions within the well (and between the well and substrate). Because the wells are totally junction-isolated from the rest of the wafer, it is especially important that provisions for well connections be made at the top surface of the wafer. That is, in non-well technologies, it may still be possible to contact the substrate from the backside of the wafer even if no provision is made for a substrate contact from the top surface. Obviously, such backside connections cannot be established to the well regions.

The dopant concentration near the surface of the well is one of the factors that impact V_T , carrier mobility, source/drain capacitance, and field isolation. The well concentration deeper beneath the surface affects other characteristics of MOSFET behavior such as latch-up and vertical punchthrough (i.e., between drain and well-substrate junctions).*

* To ensure that vertical punchthrough will not occur, the depletion regions of the drain/well and well/substrate junctions must be prevented from merging. That is, the depth of the well/substrate junction must be deep enough (typically $>1.5 \mu\text{m}$) so that it is greater than the sum of the widths of the two depletion regions in the well, even when a reverse bias of V_{DD} is

8.1 *p*-WELL, *n*-WELL, and TWIN-WELL TECHNOLOGIES

There are many trade-offs involved in the optimization of a CMOS process. The choices revolve around the highly interrelated parameters of circuit performance, layout density, fabrication cost, and susceptibility to latchup. As described in previous chapters, obtaining the best circuit performance from an MOS device involves maximizing the drive current and minimizing junction capacitances and body effect - all characteristics which favor lower doping concentrations in the device body. Optimizing density, however, requires raising these same doping concentrations (to avoid punchthrough and to achieve high values of threshold voltage in the parasitic field-transistors). Higher density is thus achieved by allowing closer packing of adjacent *n*- and *p*-channel transistors. Conversely, latchup tolerance can also be improved by keeping *n*- and *p*-channel transistors farther apart (see chap. 6, Vol. 2), but this in turn lowers circuit density.

These complex interacting tradeoffs converge on several processing configurations that are determined at the outset of the processing sequence through the selection of the type of well doping.^{1,2} It is therefore useful to consider the ramifications of the well-type choice with respect to its impact on circuit applications. Because at least one of the device types must be located in the well, the selection of well architecture influences device and circuit behavior. In particular, the performance of devices in the well for the most part suffer as a consequence of the higher well doping. That is, they exhibit higher junction capacitance, increased sensitivity to body effect, and decreased transconductance (due to reduced carrier mobility). Furthermore, substrate current (e.g., due to hot-carrier effects) is harder to collect from the well regions. The issue thus becomes one of deciding which device type (if any) should be subjected to such performance degradation.

Figure 8-1 shows the five basic well structures employed in CMOS technologies. Figure 8-1a illustrates the *single p-well structure*, in which *p*-wells are fabricated in an *n*-type starting substrate. NMOS devices are made in the well, and PMOS transistors in the substrate. If the dopant types of the *p*-well technology are reversed, a *single n-well structure* is obtained (Fig. 8-1b). The structures in Figs. 8-1c to 8-1e involve approaches in which *both p*- and *n*-wells are formed. CMOS technologies employing such structures are called *twin-well*, *dual-well*, or *twin-tub CMOS*. The first two of these (Figs. 8-1c and 8-1d) are designed to be fabricated on a particular substrate type, like the single *p*- and *n*-well structures. The third of these twin-well technologies (Fig. 8-1e), however, is independent of starting material type.

In the next sections we will outline the pros and cons of both the single *p*-well and *n*-well configurations, as well as those of the more complex well configurations that have been developed (e.g., twin-well and retrograde-well CMOS).

* (cont.) applied to the drain (see example 8-1 in this chapter.). In conventional-well CMOS structures, this depth is achieved by a long, high-temperature drive-in (up to 1200°C) following the well implants. For example, in one twin-well process the well-implants are followed by a drive-in at 1100°C for 500 minutes (8 1/3 hours).¹³

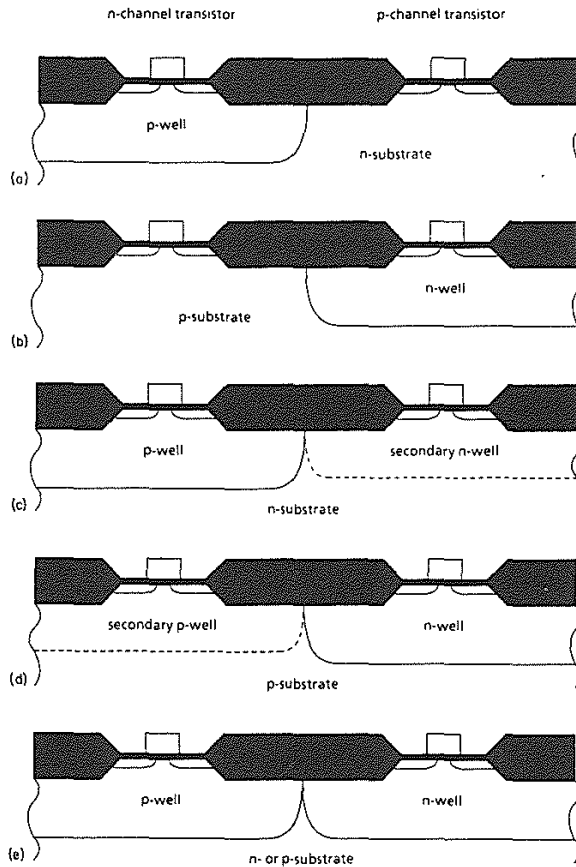


Fig. 8-1 Well structures available for CMOS circuit fabrication: (a) *p*-well CMOS; (b) *n*-well CMOS; (c) *p*-well with secondary *n*-well (twin-well constrained to *n*-substrate); (d) *n*-well with secondary *p*-well (twin-well constrained to *p*-substrate); (e) twin-well CMOS on *n* or *p*-substrate. A.G. Lewis and J.Y. Chen, "Current Trends in MOS Process Integration," in N.G. Einspruch and G. Gildenblatt, Eds. *Advanced MOS Device Physics*, Academic Press, San Diego CA, 1989, Reprinted with Permission.

8.1.1 *p*-Well CMOS

The *p*-well process, illustrated in Fig. 8-1a, involves the creation of *p*-regions in an *n*-type substrate to permit the fabrication of NMOS devices. The *p*-wells are formed by implanting a *p*-type dopant into an *n*-substrate at a high enough concentration to overcompensate for the substrate doping and to give adequate control over the *p*-type

doping in the well. The starting *n*-type substrate, however, must also have sufficient doping to ensure that the characteristics of devices fabricated in the substrate regions are adequate (a minimum doping concentration of 3×10^{14} – $1 \times 10^{15}/\text{cm}^3$ is required). If the *p*-well is doped too highly, however, the performance of the *n*-channel devices will be degraded through lower carrier mobility, increased source/drain to *p*-well capacitance, and increased sensitivity to body-biasing effects. The *p*-well doping is therefore usually selected to be about five to ten times higher than the doping in the *n*-substrate.

p-well CMOS was the first type of CMOS that could be practically manufactured, and such circuits were introduced in the late 1960s. The first companies to commercially offer CMOS components produced many designs in *p*-well technology. As this experience was spread throughout the industry, *p*-well CMOS became widely established.

There were several reasons for choosing *p*-well CMOS at that time. First, the earliest MOS ICs were based on a metal-gate PMOS process which utilizes an *n*-substrate. The work-function of aluminum is such that enhancement-mode PMOS devices can be made in a lightly doped *n*-type substrate without the need for additional channel doping. Thus, when CMOS circuits were invented (to take advantage of the reduced power consumption possible with CMOS), it was simplest to manufacture them merely by adding a *p*-well to the basic PMOS process. Second, higher channel doping levels were needed to permit sufficiently large V_T values to be established in the enhancement-mode NMOS transistors. Such higher doping was also needed in the field regions of the NMOS sections of the circuit to ensure adequate isolation (see chap. 6). These requirements are compatible with the need to have well surface concentrations five to ten times larger than the substrate doping level. Finally, it was relatively simple to manufacture *p*-wells with the existing process fabrication technology. That is, in the late 1960s ion implantation was not yet available and it was necessary to introduce the well dopants by chemical diffusion. At the doping levels required in the wells it was much easier to achieve adequate doping profile control using boron rather than phosphorus diffusion processes.

There are also several circuit and system advantages of *p*-well over *n*-well CMOS. First, *p*-well technology may be the better choice for pure-static logic, in which a good balance between the performance of both MOS device types is beneficial. (The performance of the PMOS and NMOS devices are "more balanced" because the carrier mobility reduction associated with higher well doping is compensated by forming the NMOS devices in the *p*-well; i.e., the electrons have higher mobility.) Second, it is attractive for applications that require an isolated *p*-region (such as those using an *nnp* bipolar transistor as an on-chip driver or an *n*-channel FETs for an analog input). Third, *p*-well CMOS is less susceptible to field-inversion problems than *n*-well CMOS (which uses a lightly doped *p*-substrate), and can thus be slightly easier to fabricate. (We will describe later how *p*-well CMOS can use the inherent substrate and well doping as the channel stops, whereas *n*-well CMOS must use a separate substrate channel-stop process.) Fourth, if the so-called retrograde-well process (rather than a diffusion of a shallow implant) is used to form the wells, *p*-well technology is more feasible. It is easier to form a *p*-retrograde well than an *n*-retrograde well, since boron ions penetrate deeper than arsenic or phosphorus ions at the same implant energy. Finally, *p*-well

CMOS may be better for fabricating SRAMs. Since the alpha-particle-induced soft-error rate (SER) becomes significant even in SRAMs if feature sizes are scaled to submicron dimensions, the cells should be made inside a well. The sensing of the state of an SRAM cell depends on the current provided by the cell. As a result, high-gain NMOS devices are more desirable for the pass gates and drivers in the cell, and must be built in a p -well. Furthermore, it facilitates the use of polysilicon load resistors in 4-transistor memory cells and buried contacts.

Some of the disadvantages of p -well CMOS are the following: (1) Since NMOS transistors are fabricated in the p -well, their characteristics will be degraded compared to those made in a lightly doped substrate. In CMOS circuits that rely heavily on optimum NMOSFET behavior, this can lead to reduced circuit performance; (2) Substrate currents, the result of impact-ionization generated carriers in short-channel devices (see chaps. 7 and 9), are much larger in NMOS than PMOS devices. It is more difficult to collect these carriers from a well than from the substrate. Since the NMOSFETs in p -well CMOS are in the well, this represents a disadvantage of submicron CMOS p -well technology.

8.1.2 n -Well CMOS

By the late 1970s, the technological barriers that prevented enhancement-mode NMOS transistors from being manufactured were overcome. Primarily, the introduction of ion implantation made V_T control in NMOSFETs a simple, straightforward procedure. As a result, NMOS technologies ousted PMOS in high-speed, high-packing-density ICs. By the mid-1980s, however, the power dissipation in VLSI NMOS ICs became intolerably high (see chap. 6, Vol. 2), and CMOS development was pursued as a mainstream replacement technology for NMOS. For those companies that had been producing NMOS as their main IC product, n -well CMOS appeared to be an attractive approach for making the switch to CMOS.

In the n -well process, shown in Fig. 8-1b, the p -channel devices are formed in the more heavily doped n -well, while the NMOS devices are made in the substrate. Since it was now feasible to build NMOS devices in lightly doped substrates, this made it possible to make n -channel devices in an n -well CMOS process that were identical to those made by NMOS technology. Thus, virtually all of the experience that had been amassed in fabricating high-performance NMOS could be transferred to an n -well CMOS process. Furthermore, CMOS circuits could be made that exploited existing NMOS device design. The availability of ion implantation for introducing well dopants into the substrate (instead of having to rely on diffusion), now made dopant control of n -wells much more feasible. This led to the successful development of n -well CMOS technology.

As a result, virtually all EPROMs, microprocessors, and dynamic RAM designs in the generations of technology built with 1.25-2.0 μm dimensions were implemented with n -well CMOS. Mixed CMOS-NMOS designs were also made, in which the high noise immunity and negligible stand-by-power dissipation of CMOS was exploited in the input/output sections, and faster NMOS devices were used internally.

However, it should be noted that n -well technology also has some disadvantages. First, as mentioned earlier, it is more sensitive to field-inversion problems than p -well CMOS. Second, it may be more difficult to build pure-static, high-performance logic circuits with n -well CMOS.

8.1.3 CMOS on Epitaxial Substrates

As described in vol. 2, chap. 6, heavily doped substrates with a more lightly doped surface epitaxial layer have been utilized to suppress latchup in CMOS.⁴ When such starting material is used with single-well CMOS, the epitaxial layer is doped to a concentration equal to that of the substrate in a nonepitaxial wafer used for that process. If a twin-well CMOS process is used, the epitaxial layer is doped to a level significantly lower than that required for building either the p - or n -channel MOSFETs (see section 8.1.4).

The epitaxial layer is made thicker than the well depth, since the dopants in the heavily doped substrate under the epi layer diffuse toward the surface as the well dopants are diffused toward the bulk (Fig. 8-2). Thus, some of the epitaxial layer becomes more heavily doped during the CMOS process flow. The process is designed so that the bottom of the well is eventually adjacent to the heavily doped substrate region.

Either n -epi on n^+ substrates or p -epi on p^+ substrates can be used, with each method having advantages and drawbacks. Because the problems with n -epi on n^+ are more serious, p -epi on p^+ is more widely used. The major limitation of the latter

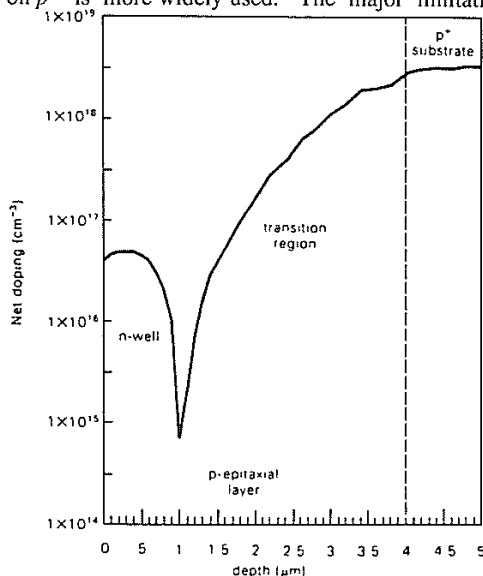


Fig. 8-2 A p -on- p^+ -epi doping profile and diffused n -well profile measured after the entire CMOS process has been completed. The as grown epi thickness is $4\text{ }\mu\text{m}$ and the final lightly doped epi thickness is about $2\text{ }\mu\text{m}$.³ (© 1987 IEEE).

approach is that the outdiffusion of boron from the p^+ is much more severe than is the outdiffusion of As or Sb in n -epi on n^+ . (The reason for this is that boron diffuses much more rapidly than antimony, which is the most widely-used n^+ dopant material.) Thus, a thicker p^- epitaxial layer must be used.

In addition, the transition region between the p^+ substrate and the p^- -epi layer is thicker, producing a larger series resistance (R_{sub}), which in turn reduces latchup immunity. On the other hand, the p -on- p^+ material is less sensitive to process-induced defects, and the p -type substrate provides higher conductivity under NMOS devices. Such additional conductivity is desirable, since it reduces the voltage drops caused by the substrate currents (generated as a result of the hot-carrier effects in short-channel devices). It is especially important in the regions containing NMOS devices, since the hot-electron substrate current is much higher in such devices than in PMOS devices.

The n -on- n^+ epitaxial substrates also offer some advantages. First, SRAMs are often built on n -type substrates, because the p -well-to- n -substrate junction provides protection from radiation-induced discharging of the memory's n -type storage nodes in the p -well.⁵ Second, retrograde p -wells are easier to implement because their implantation energy requirements are much lower. Third, the n -to- n^+ transition region is smaller than that in p -on- p^+ substrates, and the smaller value of R_{sub} provides improved latchup protection.

The limitations of n -epi on n^+ involve the process by which the heavily doped substrate is grown. Antimony (Sb) is the n -type dopant used, both because its diffusion coefficient is so low and because it exhibits much less lateral autodoping than arsenic (the other slow-diffusing n -type dopant). The problem with Sb is that its segregation coefficient, k_0 , is very small (i.e., $k_{0,\text{Sb}} = 0.023$; see, Vol. 1, chap. 1), and a large quantity of Sb must therefore be put into the Si melt to ensure that a sufficiently heavily doped ingot will be produced. Furthermore, in even the most highly refined Sb there are high concentrations of unwanted heavy metals, which become incorporated into the growing silicon crystal. Finally, the oxygen content of the Sb-doped crystal is relatively small, due in part to the special growing conditions used when the Sb-doped ingot is pulled.^{6,7} As a result, intrinsic gettering techniques that would immobilize the metals in the substrate are not as effective as they are in p -on- p^+ epi. For these reasons, n -on- n^+ substrates are less frequently chosen when epi-CMOS is implemented.⁴

A problem that exists with both types of epitaxial substrates for CMOS is that the wells cannot be made too deep, since the lateral diffusion would then take up too much area. On the other hand, if the wells are too shallow, vertical punchthrough (between the substrate and the source/drain regions of devices in the well) will ensue.⁸ In this case, appreciable leakage current can flow along this vertical path. As noted, punchthrough can occur if the depletion regions of the p^+n -well junction and the n -well/ p^+ -substrate junction merge. The problem is even more severe in the case of the heavily doped p^+ substrate, due to the high degree of boron outdiffusion. A major limitation is thus imposed on the minimum epitaxial-layer thickness.

Another problem of epi-CMOS is back-surface autodoping. For example p -type 20 $\Omega\text{-cm}$ (7×10^{14} boron/ cm^3) epi on a 0.005 $\Omega\text{-cm}$ (2×10^{19} boron/ cm^3) substrate is representative of epi for CMOS devices. If the back surface of the substrate is not

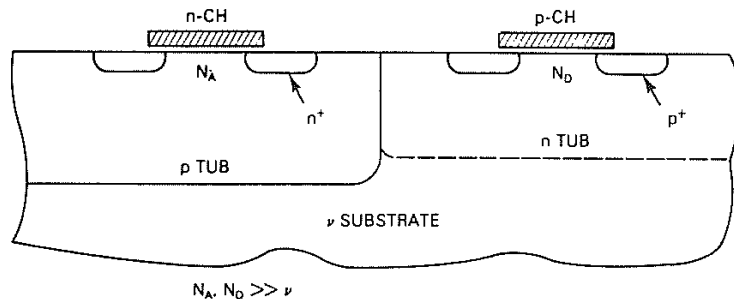


Fig. 8-3 Cross section of a twin-well 1.2- μm CMOS structure.⁵ Reprinted with permission of Semiconductor International.

sealed (e.g., by using a sealing layer such as undoped silicon dioxide or silicon nitride), boron evaporation from the substrate can contribute to autodoping on the front surface during the entire epi deposition cycle. This widens the epi/substrate interface and may even prevent the epi from reaching the 20 $\Omega\text{-cm}$ specified resistivity. Note that if silicon nitride is used as a sealing layer it should only be used in thin layers (e.g., less than 100 nm thick) since its high intrinsic stress causes it to bow the silicon.⁹

8.1.4 Twin-Well CMOS

With the twin-tub approach, two separate wells are formed for n - and p -channel transistors on a lightly doped substrate (Figs. 8-1c and 8-3). The substrate may be either a lightly doped wafer of n or p material, or a thin, lightly doped epitaxial layer on a heavily doped substrate. In either case, the level of the well surface doping is significantly lower than that required for building either the p - or n -channel MOSFETs. Each of the well dopants is implanted separately into the lightly doped surface region and is then driven in to the desired depth.

In twin-well CMOS, the doping profiles of each of the device types can be set independently, since the constraint of single-well CMOS does not exist (i.e., that the well doping must always be higher than the doping of the substrate in which one type of device is made). This was originally cited as an advantage of twin-well CMOS over single-well CMOS, with the argument made that both device types could thus be optimized.¹² This benefit for 1-2 μm CMOS has been questioned by Chen,² who points out that in modern single-well CMOS processes, an additional implant is used to prevent punchthrough without the need to raise the entire substrate-doping concentration. By incorporating this additional implant, it is possible to build higher-performance devices than can be achieved with the twin-well approach, in terms of junction capacitance and sensitivity to body effect.

Twin-well CMOS does offer some significant benefits for devices with submicron-channel lengths (although these advantages are gained at the cost of greater process

complexity). The first, and most important advantage arises when devices with submicron channel lengths are fabricated. Since it is recognized that the two device types perform similarly as channel lengths approach $0.5\ \mu\text{m}$, it is useful to provide symmetrical n - and p -channel devices. Furthermore, at submicron dimensions the body doping of both transistor types must be raised significantly to prevent punchthrough and to maintain adequate threshold-voltage levels. Thus, the advantage of having one type of MOS transistor in a lightly doped region (to optimize its performance at the expense of the other) disappears. It is instead more beneficial to produce two types of active device wells, each with its own optimized doping profile (i.e., formed by separate implants into a lightly doped substrate).*

The second advantage of the twin-well process is that it is compatible with the technologies of either isolation by selective epitaxial growth (SEG) or trench isolation.¹ Both approaches restrict the lateral diffusion of the dopants in each of the wells. In addition, sidewall inversion along the trench is less likely when both device sidewalls are butted against a trench that has been formed in a more highly doped well.¹⁰ Finally, when deep trenches are used with a thin epitaxial layer on a heavily doped substrate, latchup can be eliminated. The combined use of the twin-well process and advanced isolation techniques allows n^+ to p^+ spacing to be dramatically reduced in comparison to single-well technologies.

A third benefit is that the combination of epitaxial substrates and the twin-well process allows either substrate type to be chosen with no effects on transistor performance and essentially no change in process flow. Such flexibility is useful since some applications are best met with n^+ substrates and others with p^+ substrates. This advantage may be important if a single process is needed for implementing designs with different applications.

Finally, self-aligned channel stops can be easily implemented in the twin-well approach, allowing the spacing between n - and p -channel devices to be reduced. Although this spacing reduction is not as great as it would be if trenches or SEG were used, the process is much less complex with the twin-well method.

A twin-retrograde-well $0.7\text{-}\mu\text{m}$ -CMOS process for fabricating 1-Mbit SRAMs has been reported.¹¹ The high energy implants also allow a restricted thermal budget to be used, thus reducing the up-diffusion from the p^+ substrate. This permits the use of a thinner epi layer (which also helps prevent latch-up), and allows implantation of the channel-stop dopants after the field oxide has been grown and planarized (consequently, minimizing lateral diffusion of the boron channel-stop dopants). We discuss retrograde-well CMOS in more detail in section 8.3.

* It is not useful to start with a substrate doped to the optimum level needed for just one of the submicron devices. If this were done, a single well of much higher doping would have to be established for the other type of submicron device, which would unnecessarily degrade the device performance. It is possible, with the twin-well approach, to have both types placed in a well of optimum doping profile.

8.1.5 Summary of CMOS Well-Technology Issues

The selection of a particular CMOS process depends largely upon circuit applications, and to a lesser degree on technology evolution. For VLSI circuits with 1-2 μm design rules *n*-well CMOS has been the more widely used single-well technology, both because *n*-channel devices can be optimized in the *p*-substrate and because most circuits use more NMOS than PMOS devices. In circuits that use transistor channel lengths smaller than 1 μm , twin-well and retrograde-well technology are more attractive. As noted, the price paid for their benefits is increased process complexity.

The optimum process-integration design strategy is to first select the well technology (and to decide whether or not to use epitaxial substrates), and then to select the isolation method. Once these decisions have been made, the well depth and doping profile can be determined. The well depth impacts the lateral-diffusion distance of the well and the vertical-punchthrough voltage. The doping profile affects transconductance, threshold voltage, source/drain punchthrough, junction capacitance, carrier mobility, source/drain-to-substrate breakdown, sensitivity to body effect, and hot-electron effects.

As previously discussed, early CMOS used lightly doped *n*-type wafers and diffused *p*-wells (in which the NMOSFETs were fabricated). The *p*-well approach was initially adopted because NMOSFETs could only be reliably produced in more heavily doped regions until ion implantation was developed in the mid 1970's (see chaps. 5 and 6, Vol. 2). Eventually, however, *n*-well as well as *p*-well single-well CMOS processes were implemented. Later twin-well processes were introduced, and these now dominate the newest CMOS fabrication lines (although older *n*- and *p*-well single-well technologies are still used to fabricate a variety of CMOS products).

The well-formation process was also one that quickly exploited the technology of ion implantation. The doping precision required in the well gave implantation a significant edge over diffusion as a method for introducing the dopants into the substrate. Advances in implantation technology are allowing development of processes involving multiple implants and higher energies to avoid the long, high-temperature drive-in and to form retrograde well profiles (as will be discussed in sections 8.3 and 8.4).

8.2 CONVENTIONAL FORMATION OF WELLS AND CHANNEL STOPS

There are several issues associated with conventional-well and channel-stop formation that are covered in this section, including: (1) alignment of subsequent patterns to the well locations; (2) process sequences for forming twin-well and channel-stop regions, and; (3) novel well-formation schemes that reduce mask count and process complexity.

8.2.1 Aligning Subsequent Patterns to the Wells

The wells are often the first features to be formed in the CMOS process flow. However, the implant process that introduces the well dopants does not produce visible features on the wafer surface. Thus, it is necessary to provide some indirect indication of the well locations so that subsequent layers can be aligned to them. Here we illustrate some

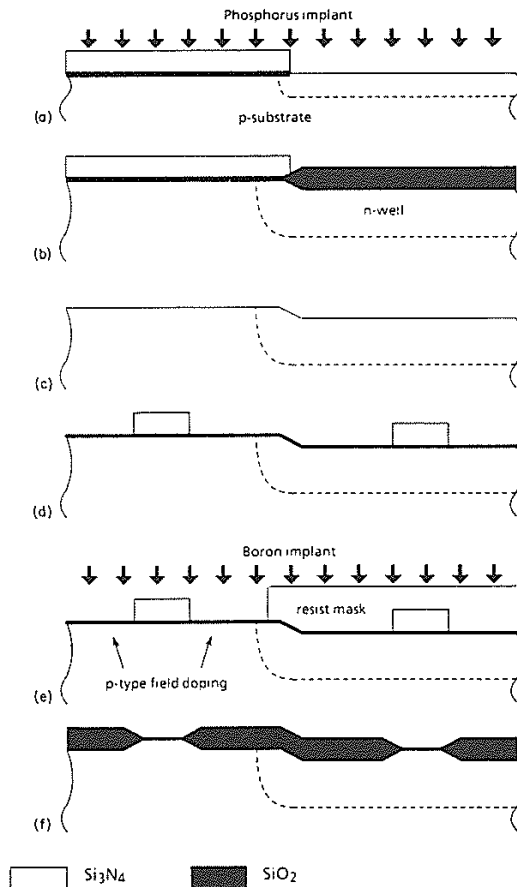


Fig. 8-4 Conventional n -well architecture with n -well as first mask. (a) Pad oxide/nitride well mask and well implant. (b) Well oxidation and drive-in. (c) Nitride/oxide strip. (d) LOCOS pad oxide/nitride mask and active area definition. (e) NMOS field implant. (f) LOCOS field oxidation. A.G. Lewis and J.Y. Chen, "Current Trends in MOS Process Integration," in N.G. Einspruch and G. Gildenblatt, Eds. *Advanced MOS Device Physics*, Academic Press, San Diego CA, 1989, Reprinted with Permission.

methods that have been developed to provide such alignment marks. For simplicity we utilize an n -well CMOS structure in depicting the process flows, but later we show how these methods can be incorporated into twin-well CMOS processes.

In the first approach the wells are the initial features created on the wafer. Here a nitride mask is employed to both define the well locations and to serve as an oxidation mask (see Fig. 8-4a and 8-4b). That is, the patterned nitride/resist layer blocks the phosphorus n -well-implant from penetrating to the p -substrate (or p -well) regions. Part of the drive-in step after the implant is carried out in an oxidizing ambient with the nitride mask layer left in place. This causes a thermal oxide layer to be grown selectively over the n -well regions. At the conclusion of the drive-in process, this oxide is stripped. Since the oxide growth consumes some silicon in the n -regions and since no such silicon loss occurs in the regions covered by the nitride mask, a step is produced on wafer surface when the oxide is stripped (Fig. 8-4c). This step is also self-aligned to the well edge, therefore providing a visible indication of the well boundaries that allows the active area features to be aligned to them. (Subsequent layers, however, are usually aligned to the active area.) The disadvantage of this approach is the step that is created on the wafer surface. In submicron processing such nonplanar topography is undesirable.

The second approach is a simpler process and it also avoids the creation of the step produced in the previous method (see Fig. 8-5). Here, however, the first features to be defined are the active areas (Mask #1). That is, a nitride mask stack layer is patterned first and all subsequent layers are aligned to the active area pattern. Then the well mask (Mask #2) is aligned to the active area pattern, and the well implant is performed. Note that the implanted dopants must penetrate the nitride mask stack. This does not present a serious difficulty, but it does mean that the n -well tends to be shallower and slightly less-heavily doped beneath active regions than under the field areas. Both the first and second approaches require three masking steps (including the boron channel-stop implant that is performed after the well implant - but prior to the LOCOS field oxidation process). However, the second approach offers the benefit of needing fewer oxidations and depositions.

The first method can be applied to a twin-well process by growing an oxide after the n -well implant (Mask #1), but not doing much of the drive-in step during this oxidation. Then the nitride mask is stripped and a second p -well mask is aligned to the n -well oxide (Mask #2). After doing the p -well implant the well drive-in process is performed. An even easier (i.e., one mask) procedure would be to use the n -well oxide to mask the p -well implant (instead of using the second mask). The second method can also be applied to a twin-well process but two masks must be used to create the twin wells. However, if a shallow p -well is acceptable, the same resist mask that is used to form the channel-stop region in Fig. 8-4e (Mask #3) can be used to form the shallow p -well. In this case the p -well implant must be performed at a higher energy than the field implant in order to penetrate the thin nitride mask layer.

The third method utilizes an initial alignment mark etched into the silicon surface prior to forming any device features. This mark is also used as an automatic alignment target by the wafer stepper. Figure 8-6 shows an example of such an alignment pattern.¹⁶ The use of such initial alignment marks is likely to increase with the proliferation of automatic stepper alignment.

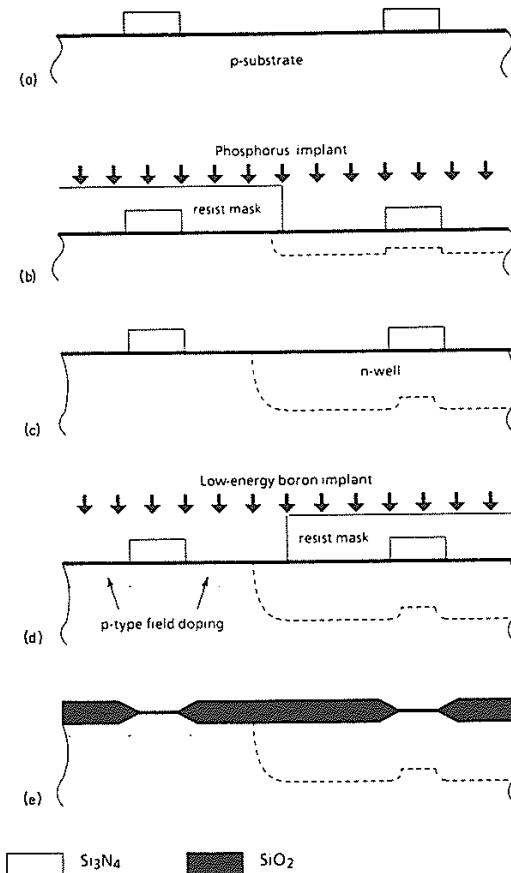


Fig. 8-5 Conventional *n*-well architecture with active area as first mask. (a) Active area definition using LOCOS nitride mask. (b) *n*-well mask and implant. (c) *n*-well drive-in. (d) boron field implant, and (e) LOCOS field oxidation. A.G. Lewis and J.Y. Chen, "Current Trends in MOS Process Integration," in N.G. Einspruch and G. Gildenblatt, Eds. *Advanced MOS Device Physics*, Academic Press, San Diego CA, 1989, Reprinted with Permission.

8.2.2 Twin-Well and Channel-Stop Formation.

The most obvious method of forming twin wells is to use two masking steps, each of which blocks one of the well implants. A procedure for forming twin-wells with a single masking step, however, has also been suggested (see Fig. 8-7).¹² In this method,

a single mask is used to pattern a nitride/oxide film formed on the bare silicon surface (*Mask #1*). The openings in the film become the *n*-well regions, and phosphorus is implanted into them (e.g., at 80 keV). Next, a thermal oxide is selectively grown on the *n*-well regions, to a thickness that is sufficient to block the boron implant used to form the *p*-wells. (The nitride mask layer is used to prevent the oxide from growing in the *p*-well regions.) The pad-oxide/nitride mask layer is then stripped, exposing the silicon in the *p*-well regions. The *n*-well regions are however, still covered with the oxide grown earlier. Thus, during the *p*-well implant (e.g., at 50 keV), this oxide masks the *n*-well regions, permitting the boron to enter the silicon only in the *p*-well areas.

Next the wells are driven in (e.g., at 1100°C for 500 min). At the conclusion of this drive-in step the concentrations in the wells for an 0.8- μm -CMOS process are $\sim 1 \times 10^{16}/\text{cm}^3$ for the *p*-well, and $\sim 3 \times 10^{16}/\text{cm}^3$ for the *n*-well. The *n*-well is selected to have a higher doping concentration (i.e., to improve the punchthrough performance of the PMOS devices and to eliminate the need for a separate channel-stop implant for the *n*-well). A higher concentration in both wells would still produce devices with relatively low capacitances at the bottoms of the source/drain-to-well junctions.

The channel-stop implant procedure is also usually included in the part of the processing sequence in which the wells are formed. In one reported twin-well process,¹³ only a *p*-well channel-stop implant (boron) is used, because the doping in the *n*-well is high enough ($3 \times 10^{16} \text{ cm}^{-3}$) that a separate channel stop implant is not required. In this case, an unmasked boron implant is performed following both well implants, but prior to field-oxide growth.

In a second approach,¹⁴ an additional mask can be used to provide both *n*-well and *p*-well channel stops (Fig. 8-8a). Note that in this procedure, a mask is saved by

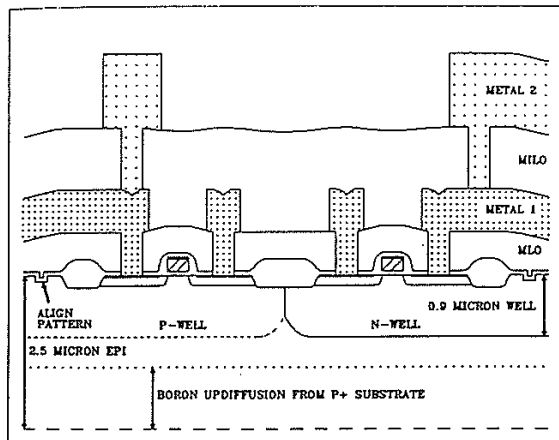


Fig. 8-6 Twin-well CMOS showing initial alignment mark.¹⁶ (© 1991 IEEE).

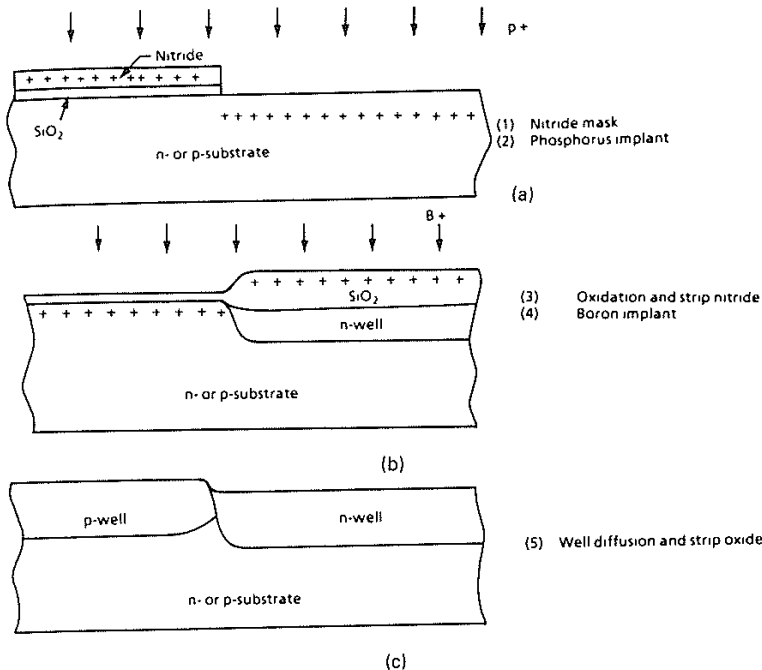


Fig. 8-7 Twin-well CMOS process using a single mask to form both wells. (a) P⁺ implant for the n-well. (b) B⁺ implant to form the p-well (blocked by the n-well oxide). (c) well drive-in for both wells.¹² (© 1982 IEEE).

performing a blanket boron implant. As a result, the boron channel stop dopant ends up in the n-well as well as in the p-well. Hence, the phosphorus channel stop dose must be increased to compensate for the boron's presence in the n-well field regions. The disadvantages of this method include the additional alignment step between the channel stops and the well masks, the interdiffusion that occurs during the oxidation step, and the asymmetry of the doping profiles. (The latter is due to the increased phosphorus concentration that must overcompensate the nonselective boron channel stop).

A variation of this method that requires no masks at all to create the channel stops has been developed to overcome the above drawbacks (Fig. 8-8b).^{14,15} In this sequence, both arsenic and phosphorus are co-implanted into the n-well regions. This places both the dopants that form the well and the channel stop into the n-well regions prior to implanting of the p-wells. An oxide is then selectively grown on the n-well regions and the boron dopant for the p-wells is implanted. After the wells have been driven in, the oxide over the n-wells is retained, and a second boron implant is carried out. This implant serves as both a channel stop in the p-well field regions and a punchthrough-prevention implant in the active regions of the p-well.

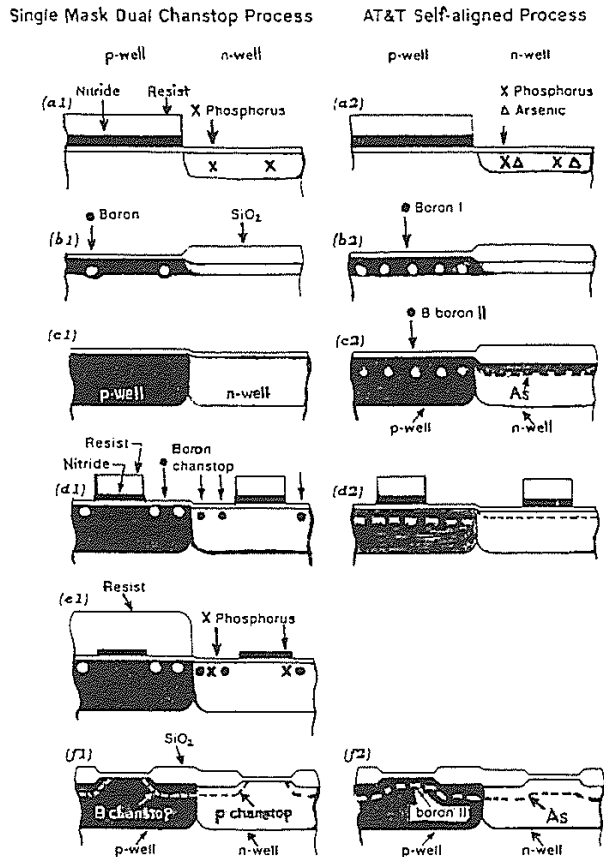


Fig. 8-8 (a) Twin-well CMOS with a single-mask dual channel stop process. (b) ATT twin-well CMOS process which requires no masks to form the dual channel stops (self-aligned process.)¹⁵ (© IEEE 1981).

It should also be noted that the minimum separation between two wells of the doping type at two different potentials is limited by electrical isolation. In single-well processes, to prevent punchthrough in the lightly-doped substrate between two such wells, adequate spacing between them must be allowed.

8.2.3 Split Well-Drive-In Process

In another report, the long, high-temperature drive-in step is performed in a novel manner, allowing the p-well transistors to be fabricated without use of a channel-stop

implant.¹⁷ After the *p*-well implant is carried out, the field oxide is first grown without driving the implanted well dopants. During this step, some of the implanted boron impurities segregate into the field oxide, leaving a surface concentration as low as would exist after any field oxide step. Next, the remainder of the well drive-in step is performed in a non-oxidizing ambient. The well dopant deep in the bulk now back-diffuses toward the surface, raising the Si-SiO₂ interface concentration by a factor of 2-3 (to $2\text{--}3 \times 10^{16} \text{ cm}^{-3}$ in this case, see Fig. 8-9). The well concentration at surface of the field regions is only about 1.5x lower than at the surface in the active regions. V_{TFISO} greater than 10 V (at I_D at $0.5 \text{ pA}/\mu\text{m}$ at $V_{\text{DS}} = 8 \text{ V}$) can be produced for isolation lengths of $1.2 \mu\text{m}$. It has been observed that for very narrow NMOSFETs (e.g., $0.35 \mu\text{m}$) fabricated with the split-well drive-in scheme, V_T is decreased by $\sim 100 \text{ mV}$ compared to wider channel devices (e.g., $5 \mu\text{m}$). This occurs because boron also diffuses from the channel regions to the field regions during the post-oxidation drive-in step. Such diffusion occurs because, as described earlier, the boron concentration at the silicon surface in the field regions has been reduced during the field oxide step by boron segregation effects.

8.3 RETROGRADE-WELL CMOS

As previously discussed, conventional wells are formed in single- and twin-well CMOS technology by implanting the well dopants and then diffusing them to the desired depth. However, lateral as well as vertical diffusion occurs, which has the effect of reducing packing density. For example, since the lateral diffusion is ~ 0.7 times the junction

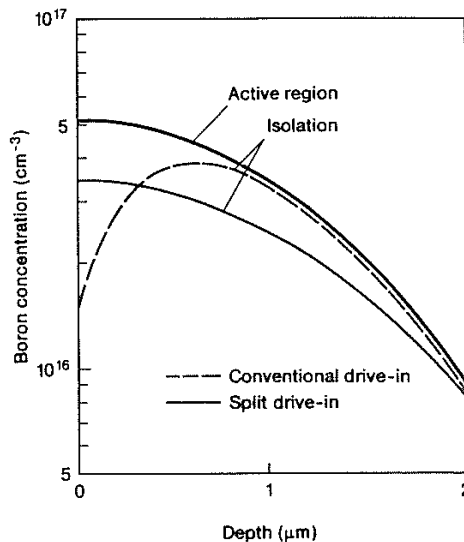


Fig. 8-9 SIMS profiles of boron well underneath field oxide and in the active area for split and non-split drive-in. Implanted boron dose: $8 \times 10^{12} \text{ cm}^{-2}$. The active area profile is the same for both drive-in cases.¹⁷ (© 1990 IEEE).

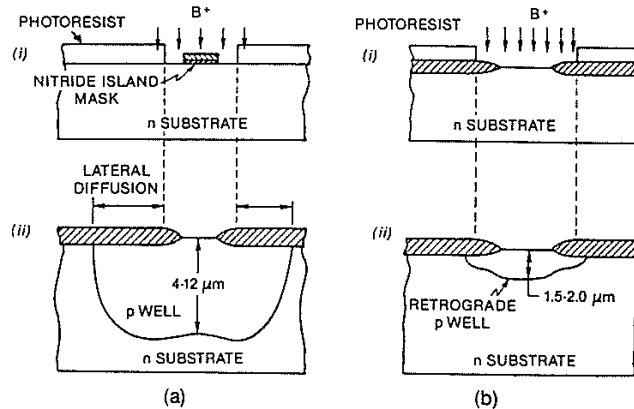


Fig. 8-10 Formation of (a) conventional and (b) retrograde *p*-well CMOS structures.¹⁹ (© 1981 IEEE).

depth, a typical diffused well of 3 to 4 μm results in a lateral spread of well dopant of ~ 2 to 3 μm . If a higher energy implant is used to place the dopants at the desired depth without significant additional diffusion (i.e., the high-energy implant is performed after the active area isolation structures [field oxide] is already in place, and only a short anneal at a relatively low temperature is used to activate the implant), much less lateral dopant spread occurs. This is because ion implantation is a more anisotropic dopant distribution process than diffusion. That is, the lateral implant straggle is $\sim 0.7 \mu\text{m}$ and barely exceeds one micron after all thermal processing.¹⁸ In addition, since the high-energy implanted wells are formed after the LOCOS isolation step, lateral spread of the well dopants is further reduced. Hence, the 7-9 μm n^+ -to- p^+ spacing required for a twin-well 1.2 μm CMOS process which uses conventional wells, can be reduced to $\sim 5 \mu\text{m}$ if retrograde wells are used instead.

Figure 8-10 depicts the cross-section of a high-energy implanted *p*-well in comparison with a conventional diffused well. In this example, the implanted *p*-well is formed by boron implantation at 400-600 keV, followed by a brief 30 min anneal at 1000°C (instead of the 6-20 hour drive-ins at 1100-1150°C used for conventional-well formation). Notice the lateral spread associated with the *p*-well is greatly reduced.

Such high-energy implants also cause the peak of the implant to be buried at a certain depth within the silicon substrate (depending on the implantation energy, but at roughly 1 μm for the energies most frequently used), and the impurity concentration decreases as the doping profile approaches the wafer surface (see Fig. 8-11). Since the well profile in this case is different from that of conventionally formed wells (in which the doping concentration after drive-in is highest at the surface, and decreases monotonically with depth), such deeply implanted wells are known as *retrograde wells*. It is estimated that, in order to avoid excessive spread of the active device electrical

parameters, the surface concentration after the retrograde implant should be at least five times smaller than its final value, which is set by shallower low-dose implants specifically designed for threshold and punchthrough control. Comparison of the doping profiles in conventional and implanted p -well profiles (obtained by simulation) is shown in Fig. 8-11.¹⁹ The retrograde-doping profile is retained by minimizing the temperature cycles of later process steps (e.g., by using RTP). Retrograde wells can be implemented on both bulk and epitaxial wafers.²⁰

Besides the potential benefit of increased packing density, retrograde wells offer the following advantages:¹

1. A retarded electric field is created in the parasitic vertical bipolar transistor, thereby reducing its current gain. This affords some protection against latchup.
2. Susceptibility to vertical punchthrough is reduced without having to significantly raise the surface doping concentration.
3. The conductivity in the bottom of the well is increased, which also provides additional latchup protection.
4. A higher threshold voltage can be achieved in the field regions of the p -wells if the boron implant is done following field oxidation (i.e., the boron does not segregate out into the oxide during the field oxide growth step).¹⁹
5. Lateral diffusion of the channel-stop boron is also eliminated, thereby reducing encroachment of the boron into the active regions. This lessens narrow width effects in the active devices.

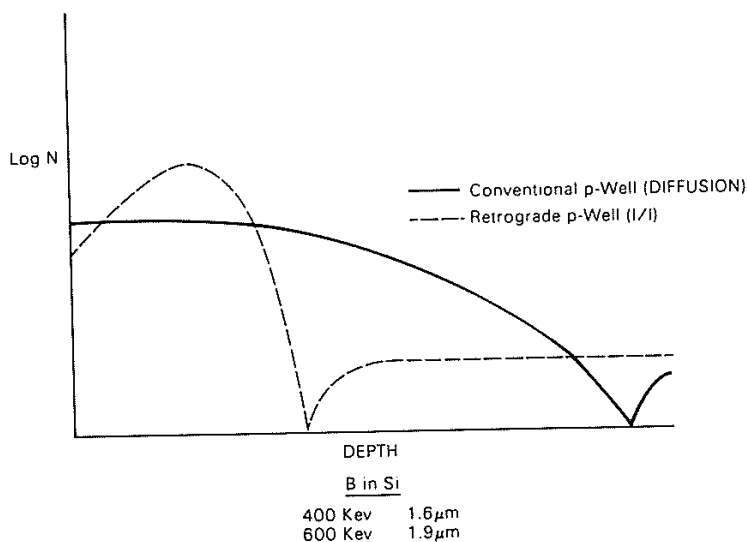


Fig. 8-11 Comparison of conventional and retrograde well p -well profiles.¹⁹ (© 1981 IEEE).

6. Greater flexibility exists in retrograde structures to optimize channel doping profiles for active device performance. That is, in conventional wells (especially shallow ones), the surface doping is primarily determined by the need to maintain sufficient vertical isolation. In retrograde structures, however, vertical isolation is secured by the high doping level towards the well bottom. Thus, it is possible to adjust the surface doping profile to a larger extent. This helps to achieve adequate short-channel device characteristics with retrograde wells.

7. The shallower well depth should also make the drain of the parasitic n^+ -to- n -well (or p^+ -to- p -well) parasitic FETs shallower. This should improve both the short-channel threshold roll-off and the punchthrough characteristics of these parasitic structures, as is discussed further in chap. 6.

However, several concerns associated with the retrograde process should also be mentioned. First, the process requires a thick photoresist as a masking material for the high-energy implant. Typical resist thicknesses are about 5 μm , and special exposure and development techniques are required for this process. Second, there is concern about the annealability of the damage caused by the high-energy implant. Third, the retrograde implant may cause both the junction capacitance and the body factor to be significantly increased. The p^+ -to- n -well junction capacitances measured in retrograde-well structures formed with a 700 keV implant have been shown to be about twice the values obtained in conventional wells. While this can adversely impact the propagation delay, circuit performance is also impacted by a number of other parasitic capacitances, and the net result on typical CMOS circuit speed and power dissipation is often relatively small. For example, in a simulation of a 32-bit CMOS arithmetic logic unit (ALU), it was found that the circuit delay increases by about 7% as the p -channel junction capacitance increases by 30 percent.¹

When the retrograde well is formed by means of very high-energy implants, however, the doping concentration under the bottom of the source and drain regions is decreased, which reduces the junction capacitance. This implies that if such high-energy ion implanters (~ 1 MeV) become available for production, some of the disadvantages of retrograde wells can be overcome.¹⁸ In addition, the higher doping that will be required in the wells for fabrication of submicron CMOS devices may mean a less severe junction capacitance penalty.

Although both p -type²¹ and n -type²² retrograde wells have been demonstrated (see Figs. 8-12a and 8-12b,²⁰ and Fig. 8-12c,²³ respectively), the p -type technology has been more widely implemented. This is because at least 600-keV beam energies are required for the formation of n -type (phosphorus) retrograde wells suitable for 0.5-0.8 μm CMOS. As of 1994 such machines were commercially available, but had only begun to be introduced into mainstream production environments. Retrograde p -type wells adequate for 0.5 μm devices require implants with energies of ~ 300 keV at a dose of $\sim 1 \times 10^{13}$ cm^{-2} . Such p -well implants can be achieved by implanting singly ionized boron at 300 keV or doubly ionized boron at 150 keV. (The required p -well energy is smaller because boron has a larger projected range than either As or P; see Vol. 1, chap. 9). Although the doubly ionized boron approach is achievable with conventional

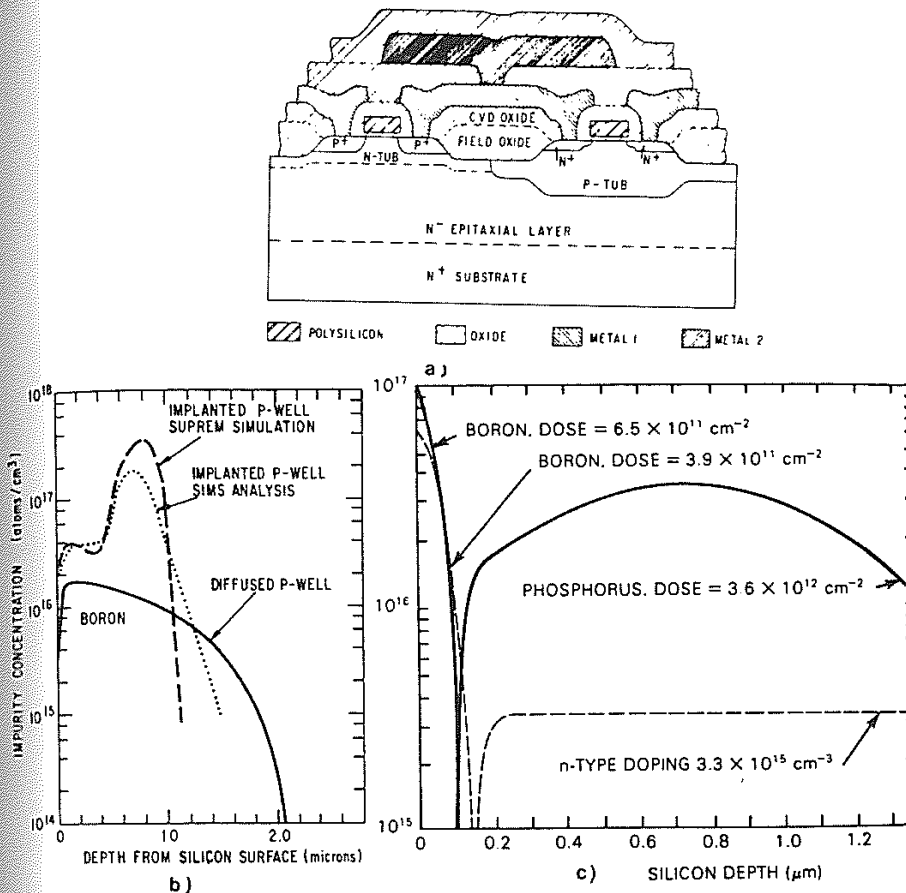


Fig. 8-12 (a) Twin-tub device cross section in which a retrograde *p*-well is used. General Electric AVLSI Process. (b) Retrograded *p*-well implanted impurity concentration profile. General Electric AVLSI process. Also shown is a conventionally thermally diffused well.²⁰ (© 1986 IEEE). (c) Simulated doping profiles in PMOS channel regions: dashed lines, lightly doped substrate; solid lines, retrograde *n*-well.²³ (© 1982 IEEE).

production implanters it is not a routine process to implement.* Subject to the precautions mentioned in the footnote, retrograde well CMOS processes have been

* One problem is that a large fraction of double-ionized ions, as high as 20%, can interact with gas molecules at the entrance of the acceleration column and become single-charged ions because of partial neutralization. These ions would only penetrate into silicon about half the desired range, distorting the profile and causing abnormalities of the threshold

the successfully implemented in a production environment with good results. However, these have primarily been limited to custom and advanced system applications, where higher costs of retrograde well processing are justified by the higher density and performance benefits offered by the retrograde-well device structures.

Two examples in which a 300 keV boron implant is used to form a retrograde p -well CMOS structure have recently been published. In the first, this p -well structure (together with a $1.2\text{ }\mu\text{m}$ epi layer) allowed an NMOS device-to-device spacing of less than $1\text{ }\mu\text{m}$ to be used when an optimized field implant was employed.²⁵ In the second, the p -well was fabricated with a 320 keV implant (using, in fact, doubly-charged boron accelerated at 160 keV), and a 5 min 1100°C RTP anneal. Together with the use of a $2\text{--}2.5\text{ }\mu\text{m}$ epi layer, a sub-half micron CMOS structure was fabricated (which permitted an n^+ to p^+ spacing of less than $2.5\text{ }\mu\text{m}$).¹⁶

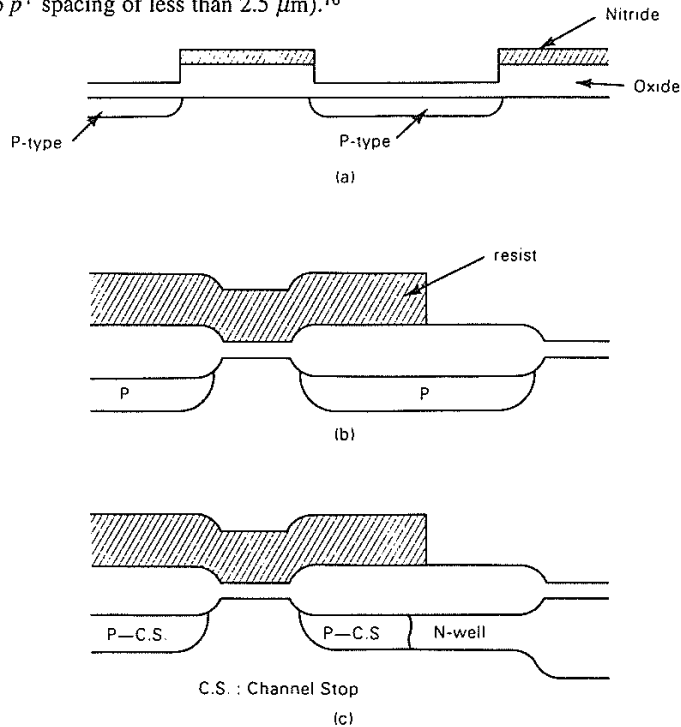


Fig. 8-13 A simplified retrograde n -well process. (a) Nitride mask for active area and B^+ implant for p -type channel stop.; (b) photoresist mask for n -well; and (c) high energy P^+ implant and any additional shallow implants in the n -well.¹⁸ (© 1986 IEEE).

* (cont.) voltage and other electrical parameters. Thus, if a high-current implanter is not available, a medium-current implanter of the post-analysis type should be used to eliminate single-charge ions before they reach the wafers.²⁴

However, an n -well retrograde CMOS process using a 700-keV phosphorus implant has also been demonstrated.¹⁸ The process sequence used to form this retrograde n -well CMOS structure (with the well implant performed after field oxidation) is shown in Fig. 8-13. The active areas are first defined using nitride masking. Next, a blanket low-energy boron channel-stop implant is performed. The field oxidation is then carried out and the retrograde n -well implant step is performed. A 30 minute, 1000°C anneal follows the implant. Note that the blanket boron channel-stop dopant is overcompensated in the n -well regions by the phosphorus implant. This retrograde-well process permits very small ($\sim 2\ \mu\text{m}$) n^+ -to- p^+ spacings.

Some advanced twin-well CMOS processes with both retrograde n - and p -wells, (using short [~ 5 min] RTP well-drive-in steps), have also been reported for submicron applications.²⁶ Scaling CMOS to 0.5 μm and below reduces the well depths required to prevent vertical punchthrough and latchup. In addition, the well depth and well-dopant concentration can be reduced if additional, lower energy implants are precisely located with respect to the transistor structure to suppress punchthrough of the MOSFET channels and to enhance the field isolation characteristics.

One novel method of achieving such short well-drive-in times is to enhance the effective diffusion rate. It has been demonstrated that the diffusion of boron and phosphorus is substantially enhanced by oxynitridation in an NH_3 ambient with a thin oxide layer over the diffusion region.^{27*} These factors combine to produce a trend of lower implant energy in retrograde wells for scaled devices at and below 0.5 μm .

A twin-retrograde-well 0.7- μm -CMOS process for fabricating 1-Mbit SRAMs on p -on- p^+ epi substrates has been reported.¹¹ The high energy implants also allow a restricted thermal budget to be used, thus reducing the up-diffusion from the p^+ substrate. This permits the use of a thinner epi layer (which also helps prevent latch-up), and also allows implantation of the channel-stop dopants after the field oxide has been grown and planarized (thus minimizing lateral diffusion of the boron channel-stop dopants).

The use of retrograde wells for both the n - and p -wells of a twin-tub CMOS process has also been reported.²⁹ The approach was called *quadruple-well CMOS* to emphasize the fact that for each type (n - or p -well), two wells were needed, one in the active area, and another, shallower one in the field region. The deep and shallow wells, however, are created simultaneously using the same higher energy implant steps. That is, since the implant is carried out after the field oxidation (and after the LOCOS nitride mask has been stripped), in regions of the wafer where no field oxide exists the implant penetrates deeply, but in regions where the implant must penetrate the field oxide, it enters the Si substrate less deeply (creating a shallower well there).

Another approach for fabricating retrograde wells without the need for very high ion implantation is to use epitaxy. Such a method for forming an n -well CMOS structure is shown in Fig. 8-14. Shallow, heavily doped wells are formed in the p -substrate by ion implantation of a slow diffusing dopant (typically arsenic) in order to prevent significant redistribution during subsequent processing. A p -type epitaxial layer is then grown to a thickness of a few microns, creating an n^+ buried layer. A low-energy phosphorus implant is then used to complete the n -wells. No drive-in is needed if it is

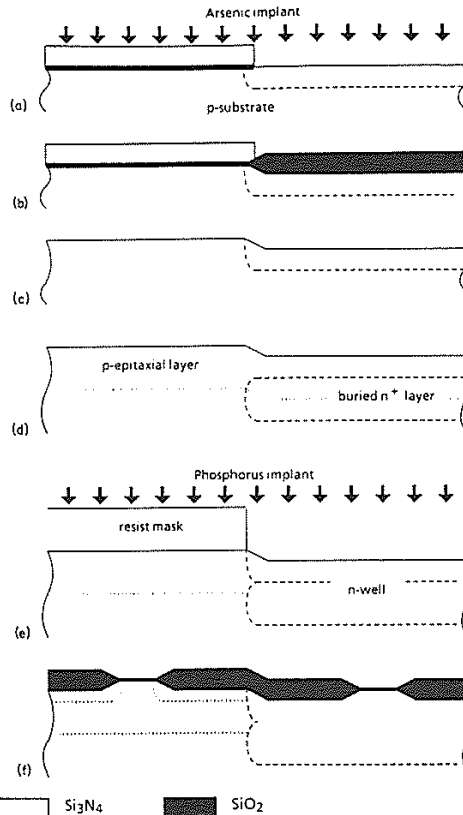


Fig. 8-14 Retrograde n -well architecture based on epitaxy. (a) Oxide/nitride well mask and buried layer implant. (b) Well oxidation. (c) Nitride/oxide strip. (d) Epitaxial growth. (e) Surface n -well implant. (f) Boron field implant and standard LOCOS field oxidation. From A.G. Lewis and J.Y. Chen, "Current Trends in MOS Process Integration," in N.G. Einspruch and G. Gildenblatt, Eds. *Advanced MOS Device Physics*, Academic Press, San Diego CA, 1989, Reprinted with Permission.

arranged that the surface phosphorus implant merges with the buried layer to complete the n -well. The formation of buried layers in this way is especially well-suited to BiCMOS technologies.

8.3.1 Retrograde Well Technologies to Improve Isolation

Since retrograde wells are formed by a high-energy implantation step following the LOCOS oxidation growth process, lateral diffusion of the well dopants is greatly reduced compared to a conventional well process. This results in significant saving of layout space. Isolation among opposite type devices is also expected to be improved.

Here we present data that supports this latter claim, as a supplement to the material on isolation presented in chap. 6.

The comparison of isolation performance between conventional-well and retrograde-well CMOS structures was performed using an n -well process and aluminum-gate MOSFETs.¹⁸ To establish a fair comparison, equal well depths were used in the active area of both the conventional-well and retrograde-well structures.

The process sequences used to form both the conventional-wells and retrograde-wells in an n -well CMOS process are shown in Figs. 8-15a and 8-15b, respectively.³⁰ The active areas are first defined in both structures using nitride masking. In the retrograde structures, a blanket low-energy boron channel-stop implant is performed next. In the case of the conventional well, the n -well resist mask is aligned to the patterned nitride, and the n -well implant (phosphorus) is carried out, followed by a blanket low-energy boron channel-stop implant. The field-oxide is now grown on both structures. In the retrograde-well structure, the n -well implant step is carried out after the field oxidation, followed by a short (~ 30 min) anneal at 1000°C . In both structures, the boron channel stop dopant is overcompensated in the n -well regions by the phosphorus implant.

Because the retrograde well is formed by high-energy ions penetrating the field oxide in the field regions the well depth there is only half of the well depth in the active regions. The shallower well depth in the field-regions of the retrograde-well structure (compared to the well field-region depth in the conventional-well structure) is advantageous. That is, it creates a shallower drain for the n^+ -to- n -well parasitic FET,

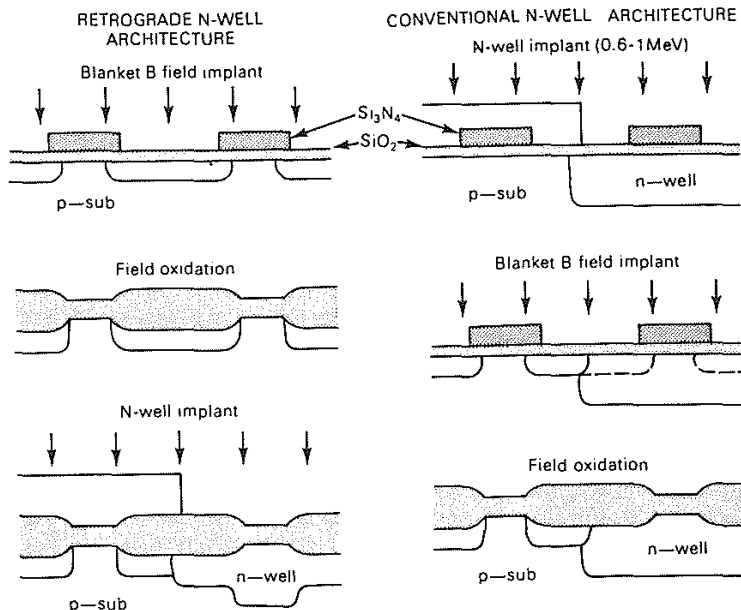


Fig. 8-15 Process architectures for retrograde and conventional n -well CMOS structures.³⁰

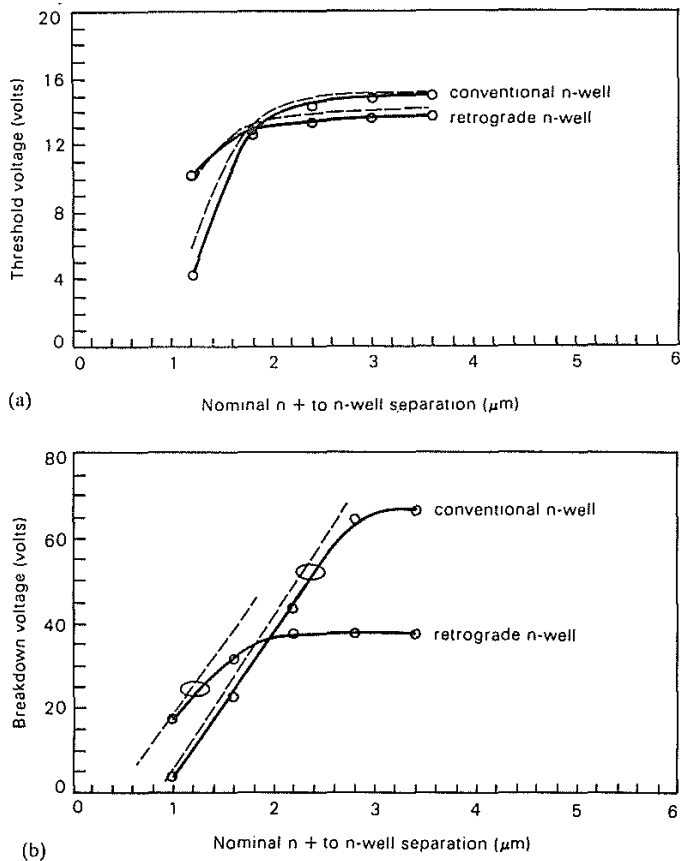


Fig. 8-16 Comparison of retrograde and conventional well structures for n^+ to n -well parasitic FETs: (a) field threshold voltage; and (b) n^+ -to- n -well breakdown voltage. Solid lines: experimental data, dashed lines: modeled results, breakdown voltage is modeled by punchthrough only.³⁷ (© 1987 IEEE).

thereby improving isolation. At the same time, the depth of the well in the active region is greater, so that the performance of the devices in the well are not significantly degraded because the n -well doping concentration near the surface remains sufficiently low. In addition, in the active regions the peak of implanted profile is near the bottom of the well, which is beneficial for vertical isolation. Meanwhile, the implant profile peaks near the Si surface in the field areas, which is again better for producing a channel stop and for preventing p^+ -to- p -substrate lateral punchthrough.

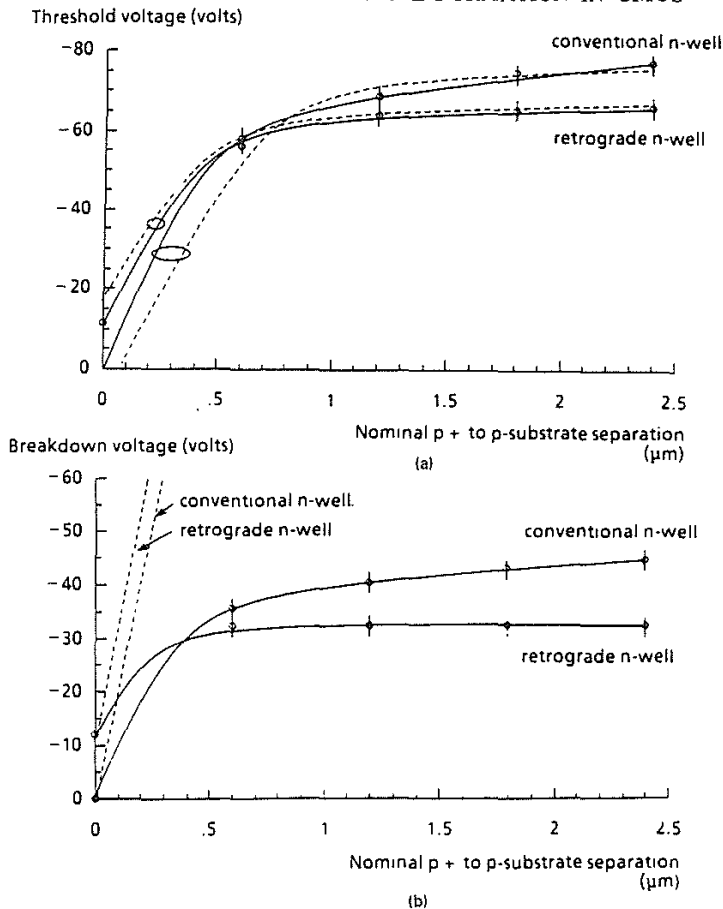


Fig. 8-17 Comparison of retrograde and conventional well structures for p^+ to p -well parasitic FETs: (a) field threshold voltage; and (b) p^+ -to- p substrate breakdown voltage. Solid lines: experimental data, dashed lines: modeled results, breakdown voltage is modeled by punchthrough only.³⁷ (© 1987 IEEE).

The measured and simulated $V_{T\text{FISO}}$ s and breakdown voltages (all defined at a drain current in the parasitic FET test structure of 10 pA) for n^+ -to- n -well parasitic transistors in both the conventional and retrograde aluminum-gate CMOS test structures are shown in Figs. 8-16a and 8-16b. At large n^+ -to- n -well separations, avalanche breakdown at the well-substrate junction occurs before lateral punchthrough occurs. The lower value of avalanche breakdown in the retrograde structures is attributed to higher well doping. As described earlier, shallower retrograde wells should exhibit less propensity to degradation

caused by short-channel effects (e.g., to threshold lowering and DIBL). These predictions are confirmed from the experimental V_{TFISO} and punchthrough data.

Figure 8-17a gives the V_{TFISO} values for lateral p^+ -to- p -substrate parasitic devices. The absolute values are much larger than those for the n^+ -to- n -well devices. These higher values are due in part to the greater field doping required in the n -well to guaranty sufficient compensation of the boron. The larger values are also partly caused by the fixed charge at the Si-SiO₂ interface, which is generally positive and thus increases the magnitude of the threshold voltage of the p -channel transistors. It is also evident that short-channel effects are much less severe in these p -channel parasitic devices, especially for the retrograde structures; the thresholds do not decrease until the channel length approaches 0.5 μm . This effect is caused by the higher channel (n -well) doping in these parasitic transistors, and to the drain (substrate) doping being lower than the channel (n -well) doping.

Figure 8-17b depicts the breakdown voltages exhibited by the same p^+ -to- p -substrate devices, and comparison with Figs. 8-17a and b again shows the superior performance of these structures. In these types of parasitic FETs, the p -substrate, which is considered to be the drain, is not only deep but wraps around under the source (see Fig. 6-45). Thus, both lateral and vertical leakage paths may possibly exist between the p^+ source and the p -substrate. The limiting values of breakdown voltage for retrograde n -well structures is set by well to substrate avalanche breakdown, as is the case of the n^+ -to- n -well transistors. In conventional wells, however, the limit is set by p^+ -to- p -substrate vertical punchthrough breakdown. Thus, the maximum breakdown voltage shown on Fig. 8-17b for these devices is not the same as that shown in Fig. 8-16b, even though the drain-channel junction is in fact the same in both cases.

The application of retrograde wells for both the n - and p -wells to a twin-tub CMOS process has also been reported. The approach, as described in the previous section, was called *quadruple-well CMOS* to indicate that for each well type (i.e., n - or p -), two wells were needed, one in the active area, and another, shallower one in the field region (see Fig. 8-18). The two shallow wells actually function as self-aligned channel stops, in which the surface concentration of the well dopants is tightly restricted on each side of the well boundary. That is, as shown in Fig. 8-19 the reduction of the net impurity concentration is restricted to a small distance from the boundary of the well edge. Specifically, the surface doping concentration near the well boundary is decreased significantly only within a few tenths of a micron from the boundary, instead of a few microns, as occurs in a conventional twin-well structure (see Fig. 6-47). Experimental data shows that at a 4 μm n^+ -to- p^+ separation, subthreshold leakage in the isolation region is <1 pA for gate voltages between -15V and +20V (Fig. 8-20). Simulation data indicate that it is possible to obtain adequate isolation with a 2 μm n^+ -to- p^+ separation.

A CMOS process based on high energy implantations without epi-layers or trenches that allows n^+ -to- p^+ spacings as small as 1.5 μm has been reported. To realize such small isolation spacing, the n -well was formed with 700 keV, $1 \times 10^{13} \text{ cm}^{-2}$ phosphorus and 1.5 MeV, $3 \times 10^{13} \text{ cm}^{-2}$ boron implantations. A unique feature of the well structure in this report is that these same two implantations simultaneously form not only the two retrograde wells and the n -well and p -well channel stops, but the boron implant also

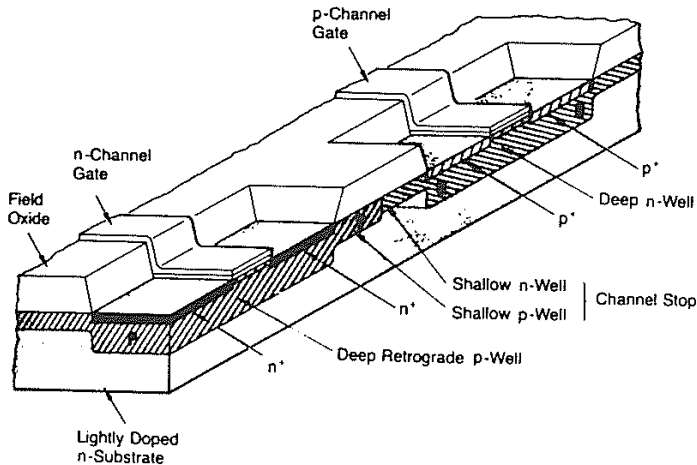


Fig. 8-18 Cross section of a quadruple-well CMOS device.³⁶ Reprinted with permission of VLSI Design.

formed a deep, buried p^+ layer under the phosphorus n -well profile. The p^+ layer eliminates the need for substrates with an epi layer, since the low resistivity substrate required to suppress latchup is created by the high energy implantation. It was also reported that the high energy implantation did not induce any excess leakage or measurably degrade MOS transistor characteristics.

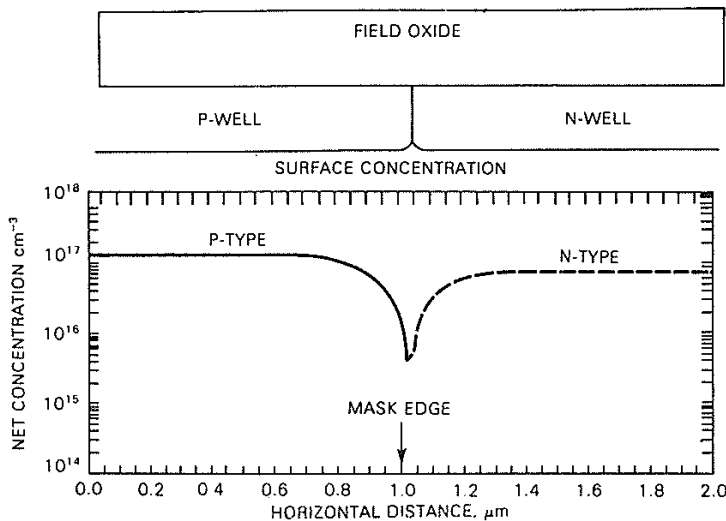


Fig. 8-19 Surface impurity concentration at the boundary of the p - and n -retrograde wells in quadruple-well CMOS.²⁹ (© 1986 IEEE).

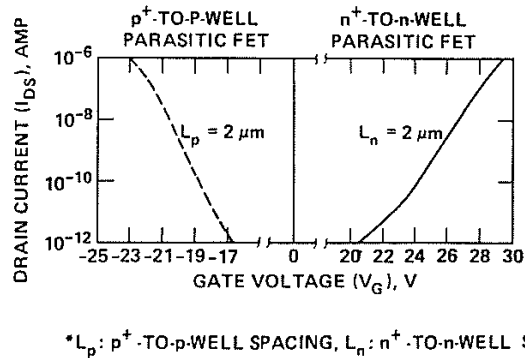


Fig. 8-20 Subthreshold characteristics of symmetric parasitic FETs in quadruple-well CMOS.²⁹ (© 1986 IEEE).

8.4 FUTURE TRENDS IN WELL FORMATION

It seems that high-energy ion implantation will find more application for generating profiled wells. Profiled wells are achieved by multiple implants at varying energies. They offer the ability to independently control several MOSFET characteristics, including threshold voltage, hot-carrier generation, punchthrough voltage, junction capacitance, latchup susceptibility, soft-error immunity, and the body-effect coefficient. The optimization of well profiles for improving device performance has been termed "well engineering."³¹

Unlike a conventional well, which has an almost flat impurity distribution, a profiled well can have an impurity profile tailored to independently optimize a variety of device characteristics. Figure 8-21 shows the impurity profile of a multiply-implanted well and the relation of the device characteristics to the profile.³²

In profile-implanted wells formed with high-energy implants, the wells are implanted after LOCOS isolation. In the *n*-well, multiple high-energy implants of phosphorus ions form a profiled well for active as well as isolation regions (with a thick photoresist serving as the *p*-well mask). Subsequently, low-energy implants are performed to adjust the PMOS V_T . A profiled *p*-well is next produced in the same way.

The deepest implant of the well in this example is fabricated with a 1 MeV implant. These implanted dopants form a heavily doped buried layer (10^{18} - 10^{19} cm⁻³) at a depth of 2-4 μ m that serves to reduce latchup susceptibility. Such a *p*⁺ buried layer is more effective than an epitaxial layer in reducing substrate resistance and suppressing latchup, thus eliminating the need for an epi layer. The ability to use non-epi substrates when wafer sizes reach 150 mm and 200 mm represents a major cost savings, as the cost of 200 mm epitaxial wafers are ~\$100 US higher than bulk-CZ wafers. The replacement of epitaxial wafers with MeV implantation processing also alleviates problems with autodoping control and backside sealing.³⁵

The use of MeV implantation avoids the long (3-21 hour), high-temperature (1150-1200°C) well-formation drive-in step of the conventional process. This minimizes

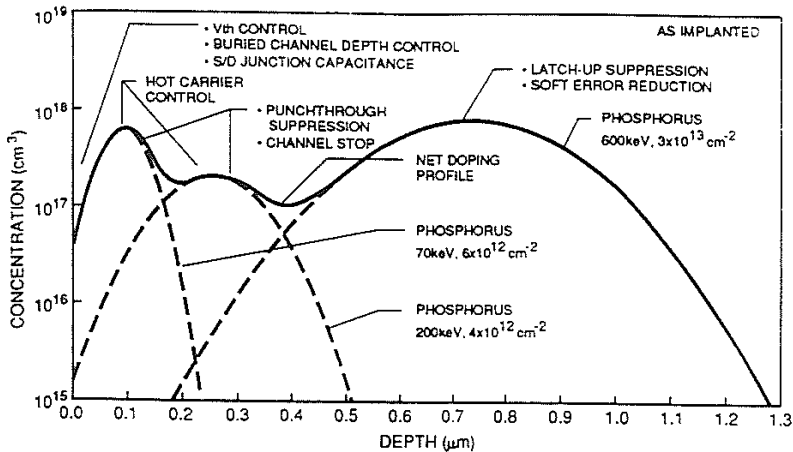


Fig. 8-21 The relation between device characteristics and the impurity concentration at various depths in multiple implanted well.³² Reprinted with permission of Solid State Technology.

thermally induced stress in large wafers that can cause wafer warpage, bow, and flatness degradation, as well as offering simplified processing with shorter fabrication times. Process simplification also results from the fact that the retrograde-well and field-isolation regions are formed by the same implant (after field oxidation). As noted in section 8.3.1, implantation through field oxide can also improve active device isolation characteristics. Employing MeV implantation to dope the bottom of the well more heavily leaves more room to tailor the doping profile at the top of the well in the channel area with lower energy implants.

Concern has existed that the high-energy implants may cause damage that will degrade device and circuit performance. Increased junction leakage is observed if the dose or energy are too low (i.e., $<3 \times 10^{14} \text{ cm}^{-2}$ and $\sim 1\text{--}1.5 \text{ MeV}$, respectively). Recent studies, however, indicate that if the dose and energy exceed $3 \times 10^{14} \text{ cm}^{-2}$ and 1.5 MeV , respectively, leakage due to implant damage is dramatically reduced with both boron and phosphorus implantation. In fact, when an optimum combination of dose, energy, and annealing process is used, *less* leakage is observed than in control devices having no buried layer. This behavior is believed to be caused by gettering of electrically active damage or impurity defects away from the depletion regions of active transistors by secondary defects near the concentration peak of the buried layer. The effect is called "self-gettering." For all of the above reasons it appears that high-energy implantation will become more widely used to form wells in device generations at and below $0.5 \mu\text{m}$.

8.5 AFTER-GATE IMPLANTATION (AGI)

An alternative technique which exploits high-energy implantation to reduce the complexity of CMOS process called *after-gate implantation* (AGI) has been suggested. That is, by performing all implantations after the gate electrodes are patterned (wells, channel-stop doping, and source/drain), the number of mask steps can be reduced.

Figure 8-22 illustrates this approach, in which only seven masks are required to form a single *p*-well CMOS structure.³³ The LOCOS-isolation field-oxide regions are formed first, then the polysilicon gates are deposited and patterned. Using a mask that obscures the PMOS regions, the NMOS *p*-well implant, multiple-channel implants, and the n^+ source/drain implants are carried out. Then a mask is used to cover the NMOS areas, and the PMOS-channel and p^+ source/drain implants are performed. Since the well and channel implantations must penetrate the field oxide and the polysilicon gate, they must be carried out at higher energies. The *p*-well implant energy is 600 keV (at a dose of $1 \times 10^{13} \text{ cm}^{-2}$), and the NMOS devices receive low-dose channel implants of 135 keV, 175 keV, and 215 keV. The channel of the PMOS devices receives two low-dose, higher-energy phosphorus implants (at 220 keV and 440 keV). Such multiple-channel implants are used in both NMOS and PMOS devices to avoid unacceptable V_T variations caused by polysilicon gate thickness variations encountered in processing (and to improve the punchthrough resistance of the device structures). Studies also indicate

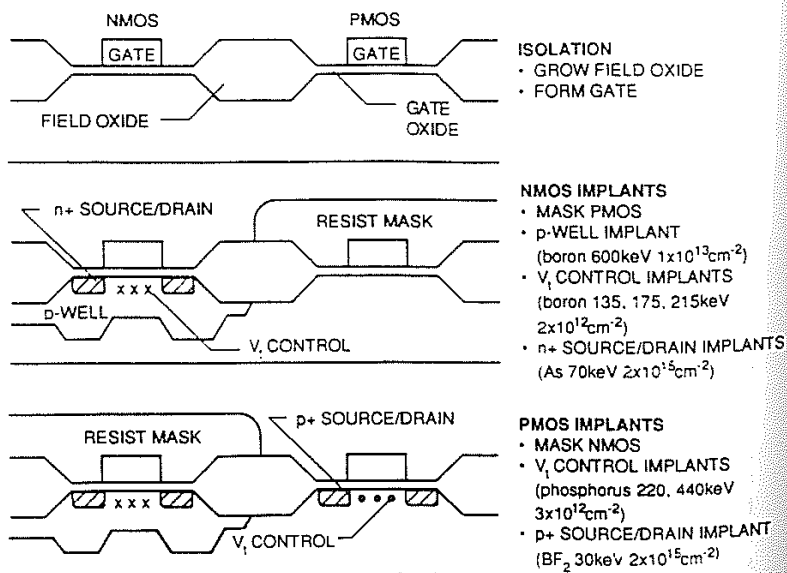


Fig. 8-22 After-Gate Implant (AGI) process sequence.³³ (© 1986 IEEE).

that the high energy AGI does not degrade the gate oxide integrity³² nor the MOS transistor characteristics.³⁴

The AGI process thus offers the benefits of implantation processing freedom and flexibility, as well as reduced CMOS process complexity. The low doses required for the well and channel implantations can be performed with a medium-current implanter which may also have a higher energy implantation capability.

REFERENCES

1. J.Y. Chen, "CMOS - The Emerging Technology," *IEEE Circuits and Devices Magazine*, March 1986, p. 16.
2. J.Y. Chen, *CMOS Devices and Technology for VLSI*, Prentice-Hall, Englewood Cliffs, N.J., 1990.
3. A.G. Lewis et al., *IEEE Trans. Electron Dev.*, **ED-34**, 2156, (1987).
4. T. Yamaguchi et al., *IEEE Trans. Electron Dev.*, **ED-31**, p. 205 (1984).
5. L.C. Parrillo "CMOS Active and Field Device Fabrication," *Semiconductor International*, April 1988, p. 64.
6. W. Wijaranakula et al., *J. Electrochem. Soc.*, December 1988, p. 3113.
7. H. Tsuya et al., *Japanese J. Appl. Phys.*, **22**, L16 (1983).
8. A. G. Lewis et al., "Vertical Isolation in Shallow *n*-Well CMOS Circuits," *IEEE Electron Dev. Lett.*, EDL-8, p. 107, 1987.
9. K. Tanno, F. Shimura, and T. Kawamura, *J. Electrochem. Soc.*, **128**, p. 395 (1981).
10. J. Manoliu and J.O. Borland, "A Submicron Buried Layer Twin-Well for High Density CMOS," *Tech. Dig. IEDM*, 1987, p. 20.
11. R. de Werdt et al., *Tech. Dig. IEDM*, 1987, p. 532.
12. L.C. Parrillo et al., "Twin-Tub CMOS-II," *Tech. Dig. IEDM*, 1982, p. 706.
13. R.A. Chapman et al., "A 0.8 μm CMOS Technology," *Tech. Dig. IEDM*, 1987, p. 362.
14. M.-L. Chen et al. *Tech. Dig. IEDM* 1986, p. 256.
15. S.J. Hillenius and L.C. Parrillo, U.S. Patent No. 4,554,726, Nov. 25, 1985.
16. R.A. Chapman et al., "High Performance Sub-Half Micron CMOS Using Rapid Thermal Processing," *Tech. Dig. IEDM* 1991, p. 101.
17. Ch. Zeller, C. Mazure, & M. Kerber, "Field-Implant-Free Isolation by Double-Well Split Drive-In," *IEEE Electron. Dev. Letts.*, p. 215, May 1990.
18. R. Martin and J.Y. Chen, "Optimized Retrograde *n*-well for 1 μm CMOS, *Proc. Custom Integ. Circuits Conf.*, 1985, p. 199.
19. R.D. Rung, C.J. Dell'Oca, and L.G. Walker, "A Retrograde *p*-well for High Density CMOS," *IEEE Trans. Electron Dev.*, **ED-28**, p. 1115, 1981.
20. D.M. Brown et al., "Trends in Advanced CMOS Process Technology," *Proceedings of the IEEE*, December 1986, p. 1646.
21. S.R. Coombs, "Scalable Retrograde *p*-well CMOS Technology," *IEEE Trans. Electron Dev.*, **ED-28**, p. 346, 1981.
22. Y. Taur et al., *IEEE J. Solid-State Circuits*, **SC-20**, p. 123, 1985.

23. G.J. Hu et al., "Design and Fabrication of p -channel FET for 1- μ m CMOS technology," *IEDM Tech. Dig.* 1982, p. 710.
24. A. Stolmeijer, "A twin-well CMOS process employing high-energy ion implantation," *IEEE Trans. Electron Dev.*, **ED-33**, April, 1986, p. 450.
25. A. Shida et al., 3.3V BiNMOS Technology Using NPN Transistors without Buried Layers," *Tech. Dig. IEDM*, 1991, p. 93.
26. Y. Mochizuki, "Special Report: High Energy Ion Implantation Technology," Nikkei Microelectronics, (Dec. 1991), p. 92.
27. M. Moslehi and K. Saraswat, "Thermal Nitridation of Si and SiO₂ for VLSI," *IEEE Trans Electron Dev.* **ED-32**, p. 116, January 1985.
28. P. Fahey, R.W. Dutton, and M. Moslehi, "Effects of Thermal Nitridation Processes on Boron and Phosphorus Diffusion in <100> Silicon," *Appl. Phys. Letts.*, **43**(7), p. 683 July, 1983.
29. J.Y. Chen, "Quadruple-well CMOS for VLSI Technology," *IEEE Trans. on Electron Dev.*, **ED-31**, p. 910, 1984.
30. A.G. Lewis, R.A. Martin, and J.Y. Chen, "Retrograde and conventional n -well CMOS technologies: A Comparison," Internal Report, EIL-87-17, Xerox Palo Alto Research Center, Oct. 1986.
31. K. Tsukamoto et al., *Nucl. Instrumentation & Methods*, B59/60, p. 584 (1991).
32. K. Tsukamoto et al., "High Energy Ion Implantation for ULSI: Well Engineering and Gettering," *Solid State Technology* P. 49, June 1992.
33. H. Mikoshiba et al., "A Novel CMOS Process Utilizing After-Gate Implantation Process," *Dig. of Tech. Papers, 1986 Symp. VLSI Tech.*, p. 41.
34. J. Pfister et al., "A Self-aligned Elevated Source/Drain MOSFET," *IEEE Trans. Electron Dev.*, September 1990, p. 365.
35. J.O. Borland and R. Koelsch, "MeV implantation technology: Next generation manufacturing with current generation equipment," *Solid-State Technology*, December 1993, p. 28.
36. J.Y. Chen, *VLSI Design* July 1984, p. 78.
37. A.G. Lewis et al., "Device Isolation in High Density LOCOS-Isolated CMOS," *IEEE Trans. Electron Devices*, **ED-34**, p. 1337, June 1987.

PROBLEMS

1. An n -well n^+ polysilicon-gate CMOS process starts with a substrate doping of $3 \times 10^{15}/\text{cm}^3$. The well doping near the surface of is approximately constant to a level of $3 \times 10^{16}/\text{cm}^3$. The gate oxide thicknesses are both 40 nm, and $Q_{\text{tot}} = 3 \times 10^{10}/\text{cm}^2$.
 - (a) Calculate the threshold voltages of the n - and p -channel transistors.
 - (b) Calculate the boron doses needed to shift the NMOS threshold to +1V and the PMOS threshold to -1V. Assume that the threshold shifts are achieved through shallow ion implantations.
2. In a p -well CMOS technology, the n -substrate has a doping concentration of $1 \times 10^{15}/\text{cm}^3$, and a p -well concentration near the surface of $5 \times 10^{16}/\text{cm}^3$. Calculate the source/drain junction capacitances (cm^{-2}) and body effect coefficients for both n - and p -channel MOSFETs assuming the gate oxide is 20 nm thick in both devices.

3. Describe the split well-drive process, and the advantages it offers compared to the conventional well-drive process.
4. Comparing retrograde well to conventional well, what is the most important difference in process technology and finished device structure? How does it affect device isolation?
5. In a (p^+ -diffusion)- n -well-(p -on- p^+ substrate)structure, the n -well is uniformly doped at $1 \times 10^{16} \text{ cm}^{-3}$ and is $1 \text{ } \mu\text{m}$ deep, the p -epi thickness is $4 \text{ } \mu\text{m}$ and is uniformly doped at $1 \times 10^{15} \text{ cm}^{-3}$, and the p^+ diffusion is $0.4 \text{ } \mu\text{m}$ deep. If a substrate bias generator is needed and the p^+ diffusion and n -well are both biased at 5V , what is the maximum (in absolute value) substrate voltage that can be applied before punchthrough occurs?

