Epitaxial Growth and Fabrication of MOS-Gated Ge_xSi_{1-x} / Si High Hole Mobility Transistors

Peter Michael Garone

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of Electrical Engineering
June 1992

This work was supported by NSF(ECS-865-7227), ONR(N00014-90-J-1316) NJCST, and DOE

Acknowledgements

Without the help and encouragement of a large number of people this thesis would not have been possible. Dr. Steven Schwarz (SIMS) and Dr. Barry Wilkens (RBS) provided invaluable help in analyzing our epitaxial films. am grateful to Venkataraman for his help with Hall measurements and thinking though the MOS-HHMT concept together on a paper towel. Peter Schwartz insured that everyone in the Sturm group had good material to work with due to his ceaseless efforts to keep the PRT-CVD system running and for the great work he did setting up the IR temperature measurement will not forget the (many) long nights spent working alongside Erwin Prinz before conference deadlines and his ability to plow though and correct the early versions of this thesis (if you think it's bad now. A thanks to my readers Dr. Dan Tsui and Dr. Antoine Kahn for reviewing a thesis that just seemed to keep growing. I would also like to acknowledge my many friends at Princeton who encouraged and supported me in various ways.

Despite all the efforts mentioned above it would have been impossible for me to finish if it were not for the encouragement and support of my advisor Dr. Jim Sturm and my wife Debbie. Jim finally convinced me that if you work on a problem hard enough it can be overcome. He also seemed to have an unlimited supply of ideas and suggestions when the chips were down. Debbie suffered through the Ph.D. process every step of the way and loved me anyway.

Most of all I would like to thank God for giving me strength to continue and hope to allow me to persevere.

Introduction

1. Attraction of Ge_xSi_{1-x} Epitaxy

The development of strained layer epitaxial growth of Ge_xSi_{1-x} on silicon by Kasper[1] and later by Bean[2] opened up a whole new realm of possibilities for silicon-based devices. A relatively large bandgap change with respect to silicon can be obtained for germanium fractions as low as 0.2 (180 meV). Ge_xSi_{1-x} forms a random alloy over the whole range of compositions allowing the bandgap to be tailored for a specific purpose with proper choice of the germanium fraction. Most of the heterojunction devices pioneered in band gap engineering with the GaAs /AlGaAs system have been successfully implemented using the Ge_xSi_{1-x} /Si material system - such as Heterojunction Bipolar Transistors[3], MODFETs[4], Resonant Tunneling Diodes[5] and Infrared detectors[6]. The Ge_xSi_{1-x} /Si material system is particularly attractive since Ge_xSi_{1-x} /Si heterojunction devices will be more amenable to silicon processing technology than III-V materials.

2. Thesis Overview

In this thesis the growth of strained Ge_xSi_{1-x} films by Rapid Thermal Chemical Vapor Deposition is explored and the application of the Ge_xSi_{1-x} /Si heterojunction to an improved pMOS device is considered.

In chapter 1 the two key aspects of strained Ge_xSi_{1-x} films, the strain and the bandgap offset, are reviewed.

In chapter 2 the growth of these films by Rapid Thermal Chemical Vapor Deposition (RT-CVD) is discussed. Low temperature silicon epitaxy is reviewed and the growth and doping kinetics of silicon and Ge_xSi_{1-x} epitaxial films using RT-CVD are shown.

In chapter 3 an improved pMOS device using a Ge_xSi_{1-x} / Si heterojunction is introduced, the MOS-gated High Hole Mobility Transistor (MOS-HHMT). The operation of such a device is analyzed and demonstrated experimentally.

In chapter 4 the current-voltage characteristics of the MOS-HHMT are measured and models for the inversion layer hole mobility and the drain conductance are developed. Remaining questions regarding the operation and viability of the MOS-HHMT device structure are then raised and commented upon.

---

[1] E. Kasper and H.J. Herzog, J. Applied Physics **8** (1975), p.199.

[2] J.C. Bean, L.C. Feldman, A.T. Fiory, S. Nakahura and .K. Robinson, "*Ge_xSi_{1-x} /Si Strained Layer Superlattice Grown by Molecular Beam Epitaxy*", J. Vacuum Science and Technology A2 (2), April-June 1984, p.436

[3] C.A. King, J.L. Hoyt and J.F. Gibbons, "*Bandgap and Transport Properties of Si_xGe_{1-x} by Analysis of Nearly Ideal Si/Si_xGe_{1-x} /Si Heterojunction Bipolar Transistors*", IEEE Transactions on Electron Devices, vol.36 (1989), p.2093.

[4] T.P. Pearsall and J.C. Bean, "*Enhancement and Depletion Mode p-Channel Ge_xSi_{1-x} Modulation Doped FETs*", IEEE Electron Device Letters vol.7, no.5 (May 1986), p.308

[5] S.S. Rhee, J.S. Park, R.P.G. Karunasiri, Q. Ye and K.L. Wang, "*Resonant Tunneling Through a Si/Ge_xSi_{1-x} /Si Heterostructure on a GeSi Buffer Layer*", Applied Physics Letters **53** (3), 18 July 1988, p.204.

[6] T.P. Pearsall, H. Temkin, J.C. Bean and S. Luryi, "*Avalanche Gain in Ge_xSi_{1-x} /Si Infrared Waveguide Detectors*", IEEE Electron Device Letters vol.7, no.5 (May 1986), p.330.

Chapter 1 : Strained Ge_xSi_{1-x} Review

1.1. The Ge_xSi_{1-x} /Si Heterojunction

Germanium and silicon both have a diamond crystal structure and form a solid solution across the whole range of compositions from germanium to silicon The two semiconductors have approximately a 400 meV difference in bandgap making the Ge/Si material system attractive for bandgap engineering. Unlike the GaAs/AlAs material system, which is almost lattice matched, there is a 4% lattice mismatch between the relaxed lattice constants of silicon and germanium. The large lattice mismatch presents limitations in growing these heterojunctions because of the strain in the films and the corresponding misfit dislocations that can result from the strain and changes to the Ge_xSi_{1-x} band structure. Most of the material properties of these strained Ge_xSi_{1-x} (and silicon) layers are still the subject of active investigation and theory.

In this chapter two of the most important aspects of the Ge_xSi_{1-x} /Si material system, the lattice mismatch (section .2) and the band structure (section 1.3) are reviewed. The discussion will focus on Ge_xSi_{1-x} films grown on Si(100) substrates since all the work in this thesis was done under these conditions. The methods outlined in this section are also applicable to other substrate orientations and substrate compositions.

1.2. Lattice Mismatch

The lattice constant of germanium is 4% larger than that of silicon so unstrained Ge_xSi_{1-x} alloys will have a larger lattice constant than silicon. The lattice constant of relaxed Ge_xSi_{1-x} films (a_o) can be approximated using Vegard's Law as

Eq. 1.1 $$a_o = x\, a_{Ge} + (1-x)\, a_{Si}$$

where x is the germanium fraction and a_{Ge} and a_{Si} are the lattice constants of germanium and silicon. Experimental measurements of the relaxed lattice constants of Ge_xSi_{1-x} alloys can be found in Reference [1]

The misfit (f) between the epitaxial layer and the substrate is defined as

Eq. 1.2 $$f = \frac{(a_s - a_o)}{a_o}$$

where a_s is the lattice constant of the substrate and a_o is the relaxed lattice constant of the epitaxial layer.

The misfit between the epitaxial layer (Ge_xSi_{1-x}) and the substrate (silicon) can be accommodated in two ways

1. By elastically straining the material (Figure 1.1a).

   $\varepsilon \equiv$ misfit accommodated by strain

2. By forming misfit dislocations along the interface (Figure 1b)

   $\delta \equiv$ misfit accommodated by dislocations

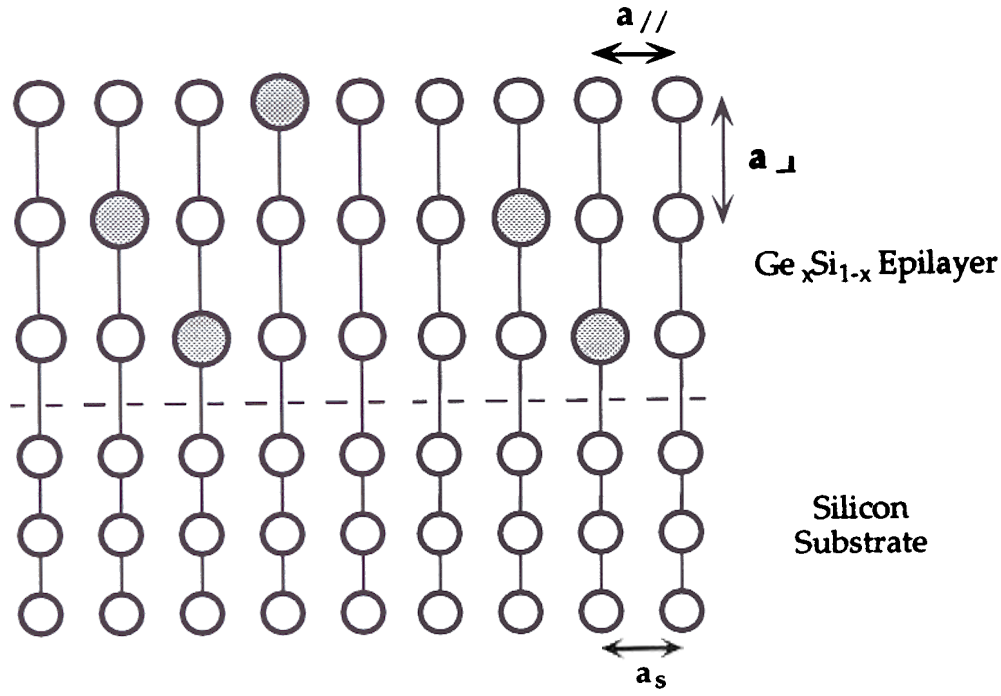
The total misfit will be equal to the sum of these two components such that

Eq. 1.3 $$f = \varepsilon + \delta$$

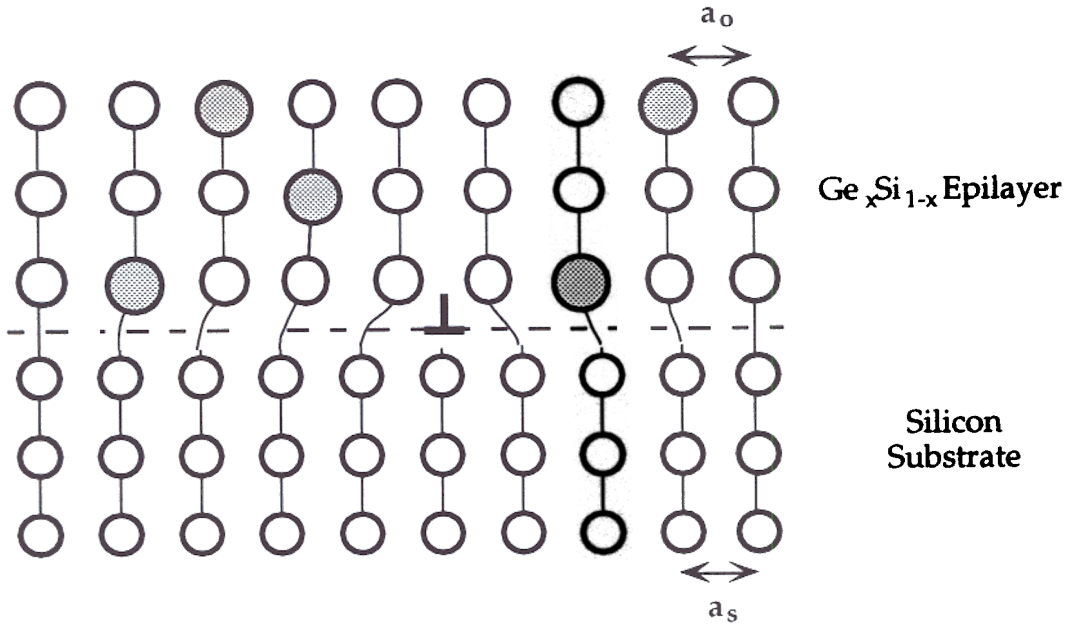
1.2.1 Equilibrium Critical Thickness

The Matthews/Blakeslee model[2] uses the energy balance of strain energy due to misfit dislocations (E_δ) and the elastic strain energy (E_ε) to determine what the equilibrium spacing of misfit dislocations in the overlying film will be. There is a strain energy associated with both the elastic deformation of the epitaxial film and the presence of dislocations (which disrupt the crystal periodicity). The elastic strain energy is given by

Eq. 1.4 $$E_\varepsilon = \varepsilon^2 B h$$



misfit accomodated by elastic strain



mismatch accomodated with misfit dislocations

Figure 1. Schematic showing two means of accommodating misfit, [a] formation of misfit dislocations and [b] epitaxial layer is coherently strained.

where ε is the misfit accommodated by elastic strain, h is the epitaxial layer thickness, and B is a factor depending on the elastic constants C_{ij} and the direction cosines of the normal to the plane relative to the cubic axis The energy associated with a misfit dislocation between two crystals is approximately

Eq. 1.5 $$E_\delta = \frac{1}{2}\xi\, b\left[\ln\left(\frac{R}{b}\right)+1\right] \quad \text{where} \quad \xi \quad \frac{G_o G_s b}{\pi(G_o+G_s)(1 \quad \nu)}$$

where G_o is the shear modulus of the epitaxial layer, G_s is the shear modulus of the silicon substrate, ν is Poisson's ratio of the strained layer, b is the Burgher's vector, and R is the radius of the dislocation strain field.

The Matthews-Blakeslee model is particularly useful for determining the maximum coherently strained epitaxial layer thickness (no misfit dislocations) that is thermodynamically stable for a given misfit (i.e. germanium fraction). The elastic strain for a given epitaxial layer of thickness h, is given by :

Eq. 1.6 $$\varepsilon = \left(\frac{D}{2\,B\,h}\right)\left[\ln\left(\frac{h}{b}\right)+1\right]$$

If the calculated value for the elastic strain (Eq. 1.6) equals or exceeds the misfit the epitaxial film will be coherently strained, i.e. not have any misfit dislocations. The *critical thickness* refers to the maximum thickness an epitaxial film can have and still be coherently strained in thermodynamic equilibrium. The critical thickness occurs when the equilibrium elastic strain energy equals the misfit (ε = f). For Ge_xSi_{1-x} epitaxial layers the critical thickness will depend on the germanium fraction (x). Films with larger germanium fractions will have larger misfits and therefore correspondingly lower critical thicknesses.

Houghton et al use a modified version of the Matthews/Blakeslee model (described above) to determine the equilibrium critical thickness for buried strained layers and strained-layer superlattices (reference 3) They assume that the elastic constants of the Ge_xSi_{1-x} layers are the same as that of silicon and that the crystal is isotropic. The critical thickness (h^*), in nm, calculated in this manner for a single Ge_xSi_{1-x} layer with no silicon cap is

Eq. 1.7 $$h^* = \frac{0.023}{f} \ln\left(\frac{4h^*}{b}\right)$$

If the epitaxial Ge_xSi_{1-x} layer has a silicon layer on top the critical thickness is increased since any relief of the elastic strain by forming dislocations in the Ge_xSi_{1-x} layer will add strain energy in the overlying silicon layer. The resulting modification to the critical thickness formula is

Eq. 1.8 $$h^* = \frac{0.023}{f} \ln\left(\frac{4(h^* + t_{sp})}{b}\right)$$

where t_{sp} is the thickness of the silicon cap layer.

The calculated critical thickness is plotted versus germanium fraction in Figure 1.2 for a single Ge_xSi_{1-x} layer with no silicon cap layer and with a 200 nm silicon cap layer. A single $Ge_{0.2}Si_{0.8}$ layer with no silicon cap has an equilibrium critical thickness of 14.4 nm. The equilibrium critical thickness for a $Ge_{0.2}Si_{0.8}$ layer increases to 22.4 nm with the addition of a 200 nm silicon cap layer. All combinations of germanium fraction and Ge_xSi_{1-x} thickness that fall on or below the critical thickness curve will be coherently strained while all combinations above the curve will either be relaxed with misfit dislocations or, if coherently strained, will be metastable.

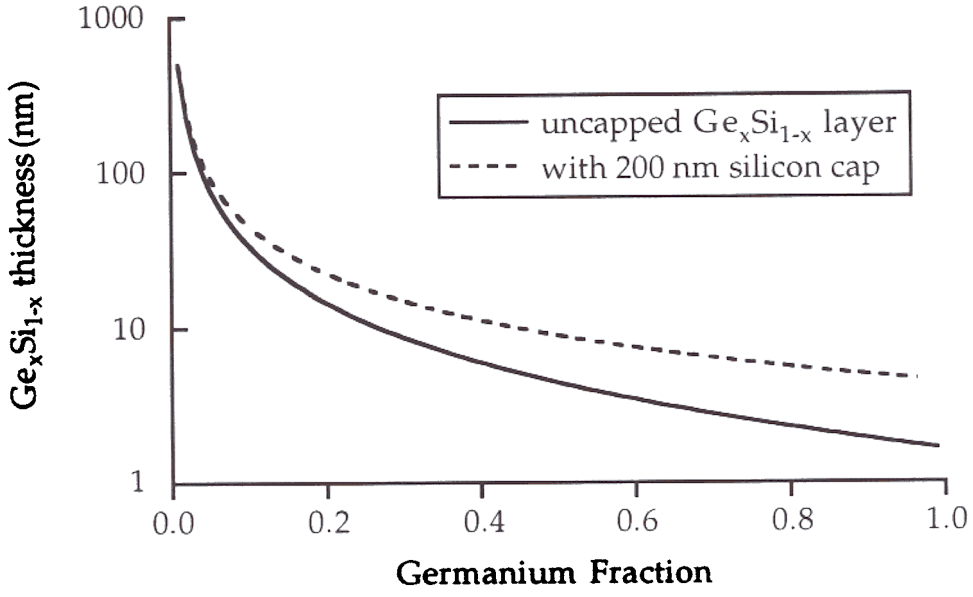


Figure 1.2 Critical thickness versus germanium fraction for a single GexSi1-x layer with no silicon cap and with a 200 nm silicon cap.

It is important to note that this is a thermodynamic calculation. It is often possible with low temperature epitaxial growth processes to grow coherently strained Ge_xSi_{1-x} films that exceed the equilibrium critical thickness [4,5] Metastable films are possible because of the large activation energy associated with the homogeneous nucleation of the misfit dislocations (≈100 eV [6]).

1.2.2 Misfit Dislocation Nucleation and Propagation

Much experimental[3,4,7] and theoretical[4,8,9] work has been directed toward understanding the relaxation kinetics of metastable strained layers (i.e. strained layer thicknesses above the critical thickness). This work is motivated by the ability of low temperature epitaxial techniques to grow coherently strained films above the critical thickness. It is often desirable to consider Ge_xSi_{1-x} structures which are metastable in order to obtain the largest possible bandgap offset and maintain a certain layer thickness. Many of the devices described in the literature fall into this regime[10,11] Because of the

desire to have thicker Ge_xSi_{1-x} films and larger germanium fractions it is important to consider whether metastable Ge_xSi_{1-x} layers will remain coherently strained during the subsequent thermal cycles which may include growth of additional epitaxial layers and device processing. The two key components in studying the relaxation of metastable strained layers are the dislocation nucleation and the dislocation propagation. Dislocation multiplication processes are also important at high dislocation densities, however for devices used in integrated circuits even small dislocation densities are unacceptable, so dislocation multiplication is omitted from this discussion.

Epitaxial growth of strained layers will always begin with a film thickness below the critical thickness, i.e. the film thickness is zero. As the film continues to grow the strain energy will continue to increase and eventually when the film exceeds the equilibrium critical thickness it will be energetically favorable to have misfit dislocations relieving the stress rather than to have a coherently strained epitaxial layer. There is a large kinetic barrier (≈100 eV) for the formation (nucleation) of homogeneous dislocations, such a large barrier that it is thought that most of the misfit dislocations are the result of heterogeneous nucleation [6,12]. The dependence of dislocation nucleation on heterogeneous sources has led to a large variation in the maximum metastable epitaxial Ge_xSi_{1-x} layer thicknesses and a great deal of difficulty in modeling kinetics of nucleation.

The effective stress is the unbalanced force responsible for the nucleation and propagation of misfit dislocations in strained layers that exceed the equilibrium critical thickness. The effective stress (τ_{eff}) is defined as the difference between the elastic stress in the layer and the stress related to the extension of a 60° dislocation segment[13] For a Ge_xSi_{1-x} epilayer of thickness h, in nm, and germanium fraction x, the effective stress (in GPa) is estimated to be [9]

Eq. 1.9
$$\tau_{eff} = 3.88\left[x - \frac{0.55}{h}\ln(10\,h)\right]$$

The film will begin to form misfit dislocations when the effective stress is greater than zero.

A semi-empirical expression for the dislocation nucleation rate (dN/dt), which depends on the effective stress (τ_{eff}), the heterogeneous nucleation site density (N_o) and the activation energy for nucleation (E_{an}) has been put forth by Houghton [9]

Eq. 1.10 $$\frac{dN(t)}{dt} = \chi N_o \left(\frac{\tau_{eff}}{G}\right)^n e^{-\left(\frac{E_{an}}{kT}\right)}$$

where χ and n are fitting parameters.

The dislocation velocity (V) in semiconductors with a diamond lattice structure has been extensively studied[14,15] and is expressed in a semi-empirical form as

Eq. 1.11 $$V = V_o \left(\frac{\tau_{eff}}{G}\right)^m e^{-\left(\frac{E_{av}}{kT}\right)}$$

where V_o is a material constant , E_{av} is the activation energy for dislocation glide, and the stress exponent m is found to vary between and .5 9.

A quantitative estimate of the dislocation density for a given metastable Ge_xSi_{1-x}/Si heterostructure and thermal cycle can be made using experimental data to fit the equations for dislocation nucleation (Eq 10) and dislocation velocity (Eq. 11). This was done by Houghton for MBE grown films in reference 9.

1.3. Band Structure

The motivation for investigating the Ge_xSi_{1-x} /Si material system is the substantial bandgap difference between silicon, which has a bandgap of 1.12 eV at 300 K, and germanium, which has a bandgap of 0.66 eV at 300 K The magnitude of the bandgap offset for Ge_xSi_{1-x} alloys has a strong dependence on the strain.

1.3.1 Influence of Strain on the bandgap

The indirect bandgap of strained Ge_xSi_{1-x} alloys on silicon substrates has been calculated by People using a full 6×6 strain Hamiltonian[16] The indirect band gap has four contributions

1. The band gap of the unstrained Ge_xSi_{1-x} alloy as measured by Braunstein et al[17]
2. Shifts of the band gap due to hydrostatic compression (change in volume).
3. Uniaxial splitting of the indirect conduction band edges, and
4. Uniaxial splitting of the degenerate valence band edge at k = 0.

In Figure 1.3 the bandgap versus germanium fraction is shown for unstrained Ge_xSi_{1-x} alloys (measurements of Braunstein) and coherently strained Ge_xSi_{1-x} alloys (calculated). In the unstrained alloys the dependence of band gap on germanium fraction is fairly weak for alloys with germanium fractions up to 0.8, with the conduction band minima being silicon-like. For an unstrained $Ge_{0.2}Si_{0.8}$ alloy the bandgap change from that of silicon is only about 110 meV. For germanium fractions higher than 0.8 the bandgap changes more rapidly with germanium fraction and the conduction

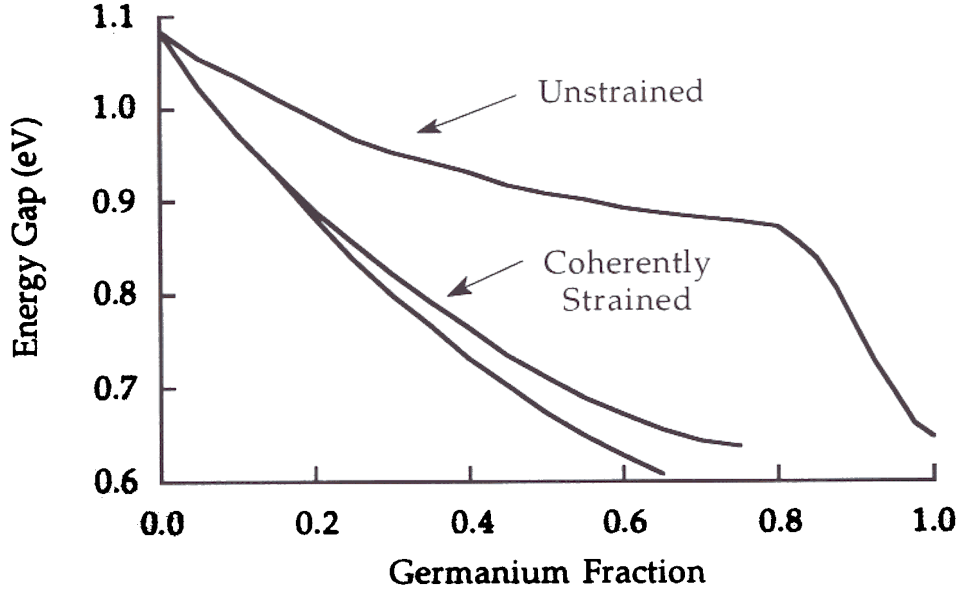


Figure 3 Bandgap energy versus germanium fraction for coherently strained and relaxed Ge_xSi_{1-x} films. The two line for the bandgap of the coherently strained Ge_xSi_{1-x} reflects the uncertainty in the deformation potentials used in the calculations. (After Reference 16)

band minimum is germanium-like. Strain causes the bandgap of Ge_xSi_{1-x} alloys to shrink more rapidly. For a coherently strained $Ge_{0.2}Si_{0.8}$ alloy the bandgap change from that of silicon is approximately 180 meV. It is apparent that maintaining a coherently strained layer is desirable not only to avoid misfit dislocations, but also to achieve a larger bandgap offset.

The effective density of states in both the conduction and valence bands is reduced because of the band splittings caused by the uniaxial strain. A schematic of the band lineup and splitting for germanium on a Si(100) substrate is shown in Figure 4. In the valence band the fourfold degenerate J=3/2 state (J ≡ angular momentum) and the twofold degenerate J=1/2 are separated in energy. In a coherently strained $Ge_{0.2}Si_{0.8}$ alloy the J=3/2 states

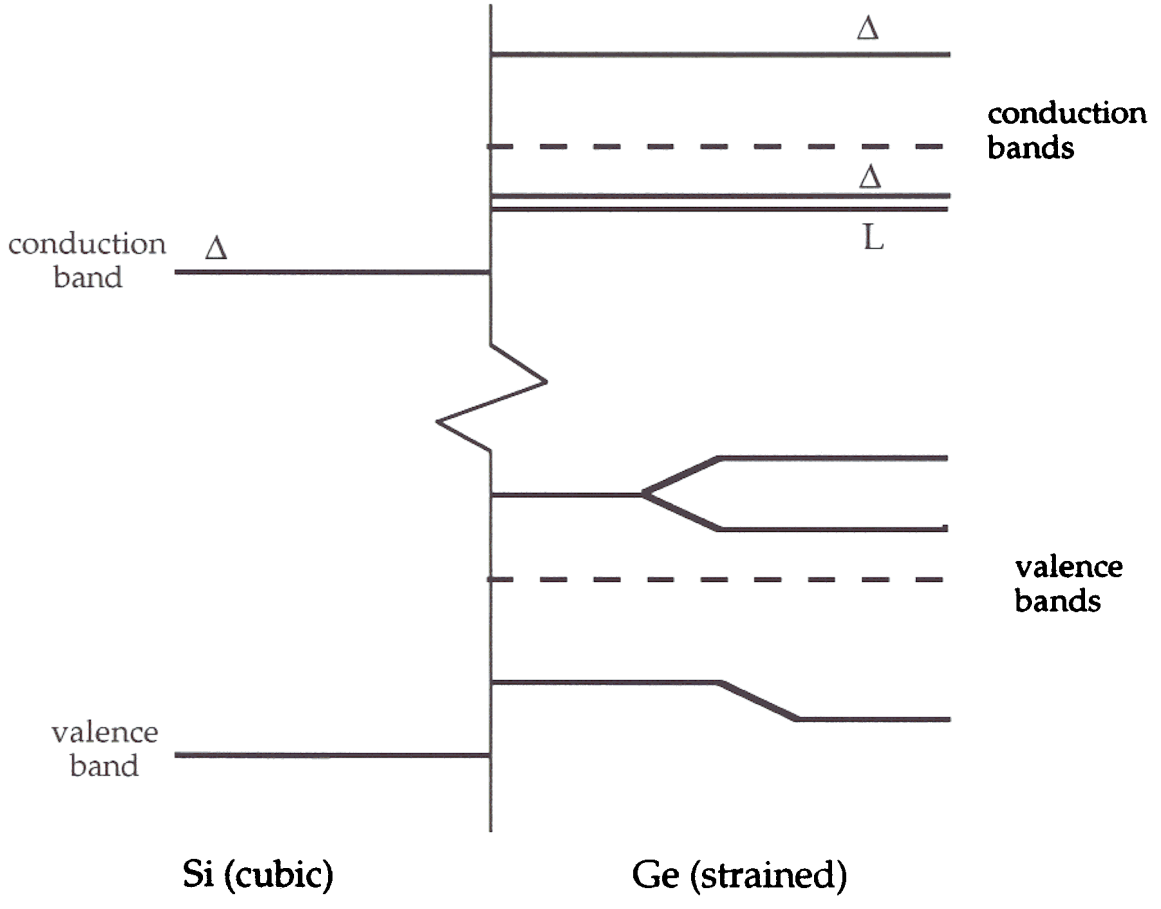


Figure 1.4 Splitting of the valence and conduction band degeneracy in a strained germanium layer on a silicon substrate. (After Reference 18)

are separated by 34 meV and the J=1/2 state is another 95 meV lower. In the conduction band the six conduction valleys are split according to the **k** vector. The four in-plane valleys [(100), (010), (010) and (100)] move down in energy and the two valleys perpendicular to the plane [(001) and (001)] move up.

1.3.2 Heterojunction Band Alignment

Van de Walle and Martin did self-consistent calculations based on the local density functional and ab initio pseudopotentials in order to study the properties of a coherently strained Si/Ge interface[18] In particular they calculated the heterojunction band lineups for germanium on silicon

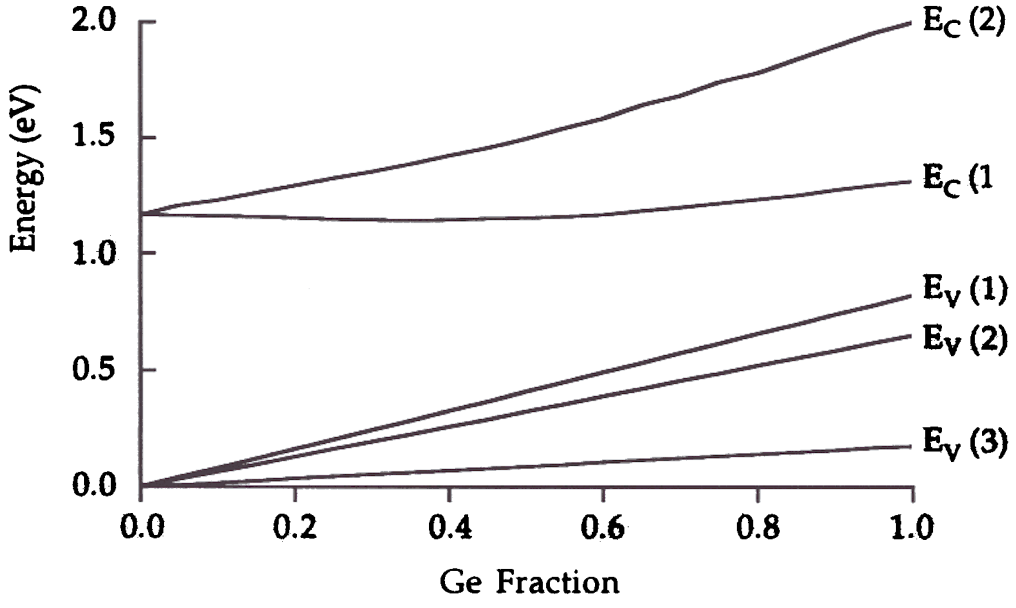
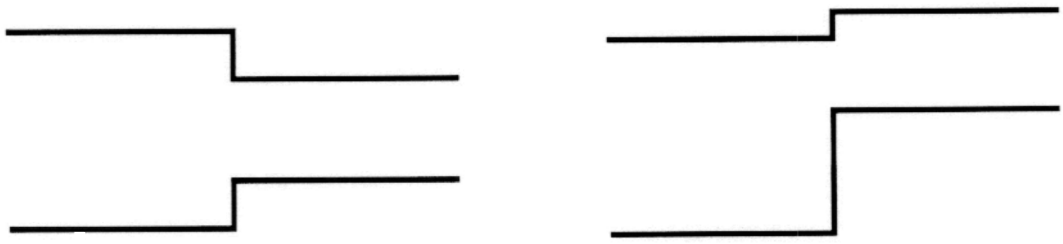


Figure 5 The offset of the conduction and valence bands, relative to the top of the silicon valence band, versus germanium fraction for coherently strained Ge_xSi_{1-x} films on Si(100) substrates. After Reference [18].

(100), (111) and (110) substrates as well as the band lineups for silicon on a germanium substrate.

The valence and conduction band offsets for strained Ge_xSi_{1-x} alloys matched to a Si(100) substrate are shown in Figure 1.5. The bands offsets are all referenced to the top of the valence band in silicon. Most of the bandgap offset for Ge_xSi_{1-x} films on silicon substrates is in the valence band. For a $Ge_{0.2}Si_{0.8}$ alloy, which has a 180 meV bandgap difference, there is a 164 ± 20 meV offset in the valence band and only a 16 meV offset in the conduction band. The band alignment is predicted to be type I for germanium fractions of up to around 0.6 and type II thereafter (see Figure .6 for pictorial representations of the band alignments).



Type I Type II

Figure .6 Pictorial definition of type I and type II band alignment. Type I - the smaller bandgap lies entirely within the bandgap of the larger. Type II - the conduction and valence band edges of the smaller bandgap lie above the respective conduction and valence band edges of the larger.

The discussion and examples discussed in this section relate to coherently strained Ge_xSi_{1-x} films on Si(100) substrates. If the substrate orientation or composition (i.e. germanium fraction) is changed there will be changes in the bandgaps and band lineups.

1.4. Material Parameters for Ge_XSi_{1-x}

Most of the properties of Ge_xSi_{1-x} alloys are estimated from their values for germanium and silicon using Vegard's Law (linear interpolation between the bulk values

for germanium and silicon). The tables on the next page include some of the more important material parameters.

Table 1.1 Band structure parameters for silicon and germanium at 300 K. After Reference [19].

		germanium
Bandgap	126 eV	0.66 eV
Effective density of states		
conduction band	2.8×10^{19} cm^{-3}	1.04×10^{19} cm^{-3}
valence band	1.04×10^{19} cm^{-3}	6.0×10^{18} cm^{-3}
Effective Mass (m^*/m_O)		
electrons	$m_l^* = 0.98$	$m_l^* = 1.64$
	$m_t^* = 0.19$	$m_t^* = 0.082$
holes	$m_{lh}^* = 0.16$	$m_{lh}^* = 0.044$
	$m_{hh}^* = 0.49$	$m_{hh}^* = 0.28$

Table 1.2 : Physical properties of silicon and germanium at 300 K. After Reference [19].

	silicon	germanium
Lattice Constant (nm)	5.431	5.646
Poisson Ratio	0.279	0.273
Dielectric Constant	11.7	16.0

J.P. Dismukes, L. Ekstrom and R.J. Paff, "*Lattice Parameter and Density in Germanium-Silicon Alloys*", Journal of Physical Chemistry vol.68, no.10 (1964), p.3021

[2] J.W. Matthews, "*Epitaxial Growth of Silicon- Part B*", Chapter 8: Coherent Interfaces and Misfit Dislocations, Academic Press (1975).

[3] D.C. Houghton, D.D. Perovic, J.-M. Baribeau and G.C. Weatherly, "*Misfit Strain Relaxation in Ge_xSi_{1-x} /Si Heterostructures: The Structural Stability of Buried Strained Layers and Strained Layer Superlattices*", Journal of Applied Physics vol.67, no.4, 15 February (1990), p.1850.

[4] J.C. Bean, L.C. Feldman, A.T. Fiory, S. Nakahura and I.K. Robinson, "*Ge_xSi_{1-x} /Si Strained Layer Superlattice Grown by Molecular Beam Epitaxy*", J. Vacuum Science and Technology A2 (2), April-June 1984, p.436

[5] S.R. Stiffler, J.H Comfort, C.L. Stanis, D.L. Harame, E. de Frésart and B.S. Meyerson, "*The Thermal Stability of SiGe Films Deposited by Ultrahigh-Vacuum Chemical Vapor Deposition*", J. Applied Physics vol.70, no.3, August 1991, p.1416.

[6] D.J. Eaglesham, E.P. Kvam, D.M. Maher, C.J. Humphreys and J.C. Bean, "*Dislocation Nucleation Near the Critical Thickness in GeSi/Si Strained Layers*", Philosophical Magazine A, vol.59, no.5 (1989), p.1059.

[7] R. Hull, J.C. Bean, and C. Buescher, "*A Phenomenological description of strain relaxation in Ge_xSi_{1-x}/Si(100) Heterostructures*", Journal of Applied Physics vol.66 , no.12 (1989), p.5837.

B.W. Dodson and J.Y. Tsao, "*Relaxation of Strained-Layer Semiconductor Structures via Plastic Flow*", Applied Physics Letters vol.51, no.17, 26 October 1987, p.1325.

[9] D.C. Houghton, "*Strain Relaxation Kinetics in $Si_{1-x}Ge_x$ / Si Heterostructures* J.Journal of Applied Physics vol.70, no.4, 15 August 1991, p.2136.

[10] C.A. King, J.L. Hoyt and J.F. Gibbons, "*Bandgap and Transport Properties of Si_xGe_{1-x} by Analysis of Nearly Ideal Si/Si_xGe_{1-x} /Si Heterojunction Bipolar Transistors*", IEEE Transactions on Electron Devices, vol.36 (1989), p.2093.

[11] J.C. Sturm, E.J. Prinz and C.W. Magee, "*Graded Base Si/SiGe/Si Heterojunction Bipolar Transistors Grown by Rapid Thermal Chemical Vapor Deposition with Near-Ideal Electrical Characteristics*", IEEE Electron Device Letters vol.12, no.6 (June 1991), p.303.

[12] C.J. Gibbings, C.G. Tuppen and M. Hockly, "*Dislocations Nucleation and Propagation in $Si_{0.95}Ge_{0.05}$ layers on Silicon*", Applied Physics Letters 54 , no. 2, 9 January 1989, p.148.

[13] C.G Tuppen and C.J. Gibbings, "*A Quantitative Analysis of Strain Relaxation by Misfit Dislocation Glide in $Si_{1-x}Ge_x$ / Si Heterostructures*", J. Applied Physics 68 , no. 4, 15 August 1990, p.1526.

[14] J.R. Patel, L.R. Testardi, and P.E. Freeland, "*Electronic Effects on Dislocation Velocities in Heavily Doped Silicon*", Physical Review B 13 (1976), p.3548.

[15] A.R. Chaudhuri, J.R. Patel, and L.G. Rubin, "*Velocities and Densities of Dislocations in Germanium and Other Semiconductors*", J. Applied Physics 33 , no.8 (1962), p. 2736.

[16] R. People, "*Indirect Bandgap of Coherently Strained GeSi bulk Alloys on <001> silicon substrates*", Physical Review B, vol. 32, no. 2 (1985), p.1405.

[17] R. Braunstein, A.R. Moore, and F. Herman, "Intrinsic Optical Absorption in Germanium-Silicon Alloys", Physical Review 109 (1958), p.695.

[18] C.G. Van de Walle and R.M. Martin, "*Theoretical Calculations of Heterojunction Discontinuities in the Si/Ge system*", Physical Review B , vol.34, no.8 (1986), p.5621

[19] S.M. Sze, *"Physics of Semiconductor Devices, 2nd edition"*, Wiley-Interscience (1981).

Chapter 2 : Low Temperature Epitaxial Growth and Kinetics

2.1 Drive to Low Temperature Processes

Great strides have been made in reducing device dimensions with improved lithography techniques. In the 1960s typical feature sizes were measured in mils (1 mil ≈ 25 μm) and integrated circuits contained only a few hundred transistors[1] Today line widths are already less than one micron in production and 16Mbit DRAM cells are going into production. Vertical device dimensions have not been shrinking as rapidly. One of the limiting factors in scaling vertical device dimensions is diffusion during the high temperature processing steps.

In the past the high temperatures required for good quality silicon epitaxy limited the ability to grow thin epitaxial layers because of inter diffusion and autodoping. In bipolar junction transistors a heavily doped buried layer is placed below in the collector to reduce the parasitic series resistance. A more lightly doped epitaxial layer is grown on top to reduce the collector/base junction capacitance, reduce the Early effect (depletion layer will be mostly in the more lightly doped collector), and to achieve a high breakdown voltage. Several microns of epitaxy was needed on top of a buried layer in a bipolar transistor to contain the dopant out diffusion. If epitaxy were done at much lower temperatures the vertical dimensions could be reduced considerably and the entire vertical structure could be grown epitaxially (emitter, base and collector).

The stability of SiO_2 at low temperatures has traditionally precluded low temperature epitaxial growth. As a result epitaxial growth temperatures in excess of 1000° C have been used to avoid the formation of hillocks and stacking faults which result from stable oxides forming on the surface.

The development of silicon MBE and UHV-CVD lowered the partial pressures of oxygen and water vapor sufficiently to allow the growth temperature to be reduced without

running into the problems of forming oxide related defects. Meyerson[2] used the equilibrium data of Ghidini and Smith for the $Si/O_2/SiO_2$ system[3] and the $Si/H_2O/SiO_2$ system[4] as the criteria for partial pressures of oxygen and water vapor below which a stable oxide would not form (see Figures 2.1a and 2.1b). On the basis of on this data and assuming water vapor and oxygen impurities of at least 1ppm in the source gases he proposed that by lowering the system pressure to less than 1 mtorr during low temperature epitaxial growth (T<850° C) the partial pressures of H_2O and O_2 would be low enough to avoid the formation of SiO_2 For example, if the source gases contain ppm of water vapor and growth is performed at 100 torr the partial pressure of water vapor is 10^{-4} torr and SiO_2 will be stable for any growth temperature below 1050° C. If the pressure during growth is reduced to 1 mtorr then the partial pressure of water vapor will drop to 10^{-9} torr and SiO_2 will not form at temperatures above 700° C.

These equilibrium $SiO_2/H_2O/O_2$ calculations proved to be overly pessimistic because kinetics, rather than thermodynamics, tend to dominate the epitaxial growth process at low temperature. If an epitaxial system's growth chamber is kept under vacuum and the source gases are pure, good quality silicon epitaxial layers can be grown at low temperature even at atmospheric pressure[5]. The ability to grow epitaxial silicon at low temperatures under non-UHV conditions is aided by the moving growth front which does not allow equilibrium conditions to be established.

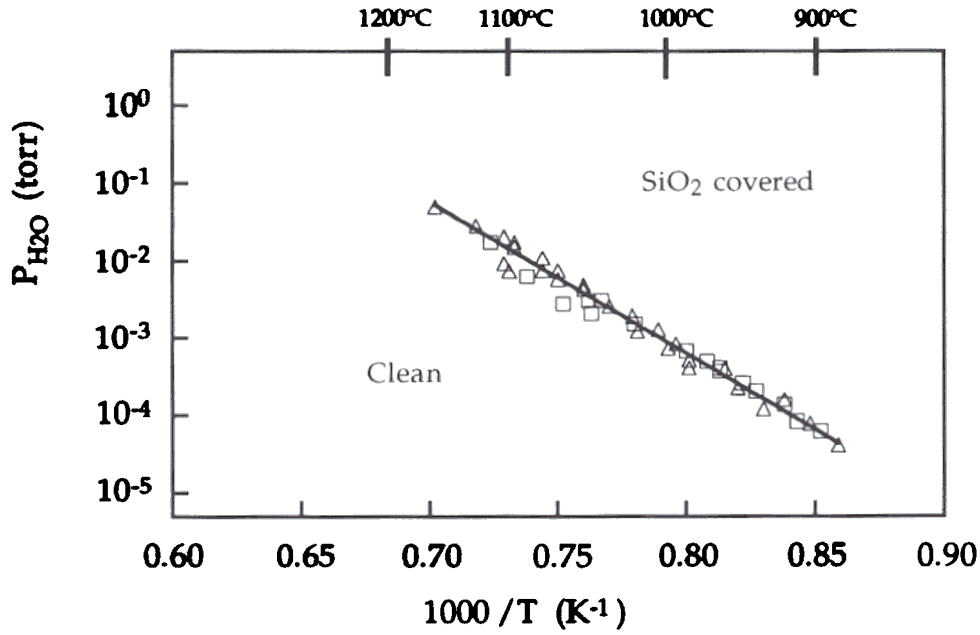


Figure 2.1a The critical conditions for growth of SiO_2 for the reaction of oxygen (O_2) with a Si(100) surface. All combinations of pressure and temperature above the line result in the formation of SiO_2. All combinations that are below the line result in a clean - oxide free - surface. After reference [3].

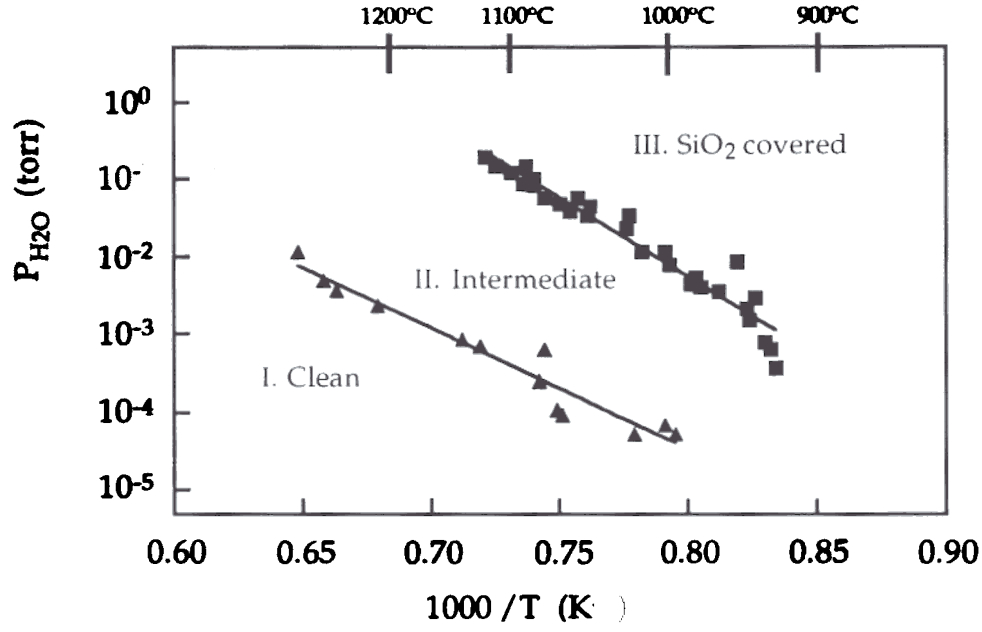


Figure 2.1b The critical conditions for growth of SiO_2 for the reaction of water vapor (H_2O) with a Si(100) surface. All combinations of pressure in area III have a stable SiO_2 layer. All combinations that are area I result in a clean - oxide free - surface. Area II is an intermediate region. After reference [4].

2.2 Princeton Rapid Thermal Chemical Vapor Deposition Reactor

2.2.1 System Description

The Princeton Rapid Thermal Chemical Vapor Deposition system is modeled after the Limited Reaction Processing reactor developed by Gibbons et al [6]. The reactor is a cold wall epitaxial reactor which heats the wafer using a bank of high intensity tungsten-halogen lamps. The wafer temperature can be changed very rapidly since the wafer provides the only thermal mass and the lamp power can be changed very quickly. For example, the wafer can be heated from room temperature to over 1000° C in less than 5 seconds. Cooling is predominantly radiative at high temperatures, so cooling from 1000° C to ≈600° C is rapid (a few seconds). Below ≈600° C the wafer is cooled by mainly conduction from the gas flow, a much slower process at low pressure. The rapid heating and cooling of the wafers minimizes the total thermal cycle that a wafer sees during a high temperature process.

Diffusion processes depend on both the time and temperature through the product of the time and the diffusivity (D), which is exponentially dependent on temperature. For example, the impurity distribution resulting from diffusion from a constant source[7] is

Eq. 2.1
$$N(z,t) = N_o \, \mathrm{erfc}\left(\frac{z}{2\sqrt{Dt}}\right)$$

where N(z,t) is the impurity concentration at a time t and a distance z from the source, and N_o is the impurity source concentration. Limited reaction processing reduces the time spent heating and cooling the wafer and thus reduces the thermal cycle to a minimum for a desired growth process.

Figure 2.2 shows a drawing of the reactor chamber in which the wafer rests on quartz pins inside a quartz tube through which the process gases flow. A

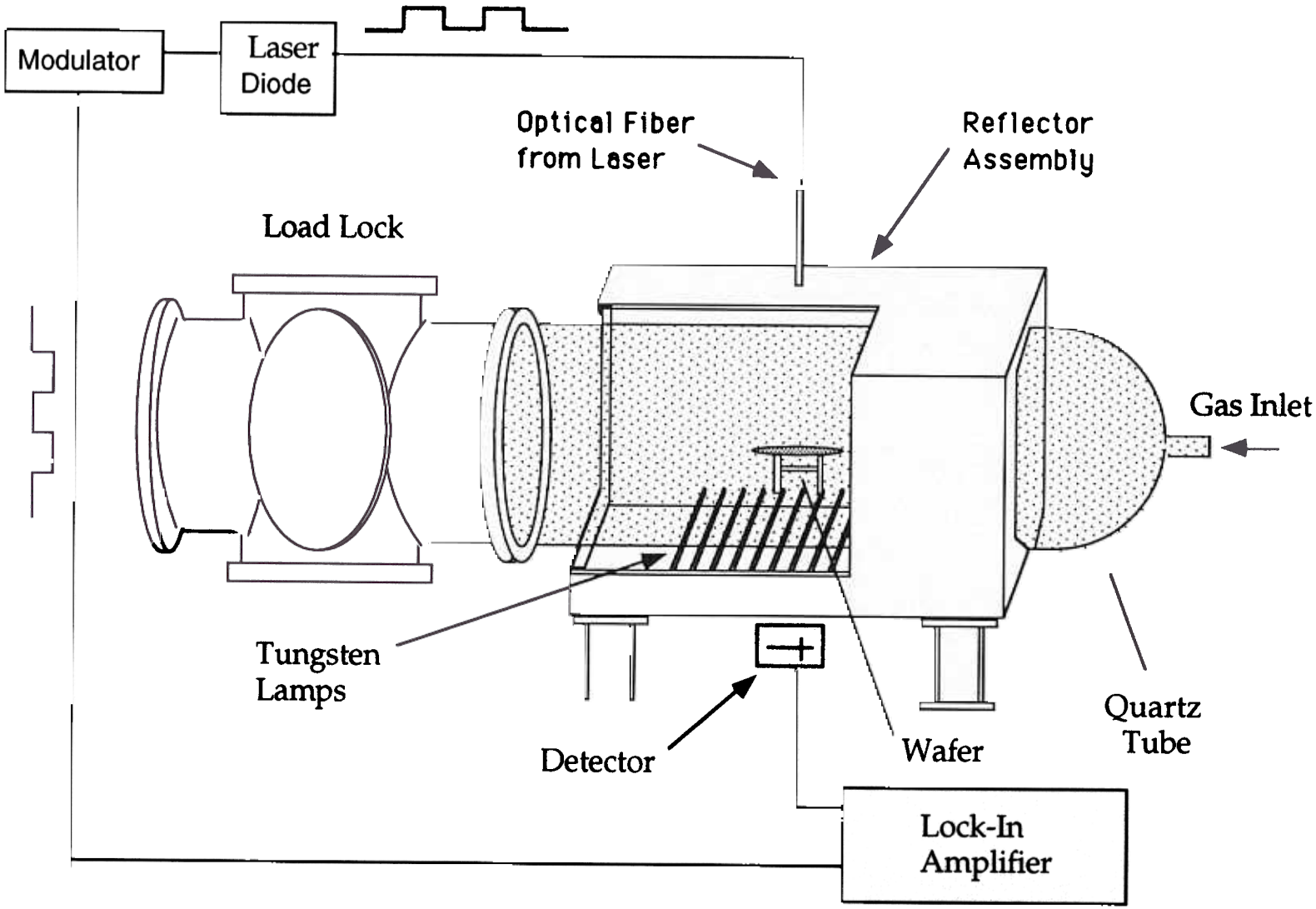


Figure 2.2 : Drawing of the Princeton Rapid Thermal CVD reactor
The growth chamber, load lock and the IR temperature measurement setup are shown. Courtesy of P.V. Schwartz.

gold reflector assembly encloses the quartz tube and houses the tungsten-halogen lamps, which lie in rows underneath the tube. Figure 2.2 also shows the load lock, which is used to minimize the introduction of oxygen and water vapor to the growth chamber while exchanging wafers, and the infrared temperature monitoring system which is described in section 2.3.3.

Figure 2.3 shows a flow diagram of the different systems of the Princeton Rapid Thermal Chemical Vapor Deposition Reactor which include the pressure control system, the temperature control system, the gas flow controls and the exhaust. The source gases are enclosed in vented cabinets and the flow into the chamber is metered using mass flow controllers. Hydrogen carrier gas, which makes up the bulk of the gas flow during growth, is passed through a palladium diffuser to remove any residual impurities (especially oxygen and water vapor). The flow of the various source gases is directed to either a vent line that bypasses the growth chamber or the main line. The flow rates of the source gases are typically stabilized in the vent line and then rapidly diverted using five-ported valves to the growth chamber to initiate growth. Thus growth cycles can be controlled by gas switching. A capacitive manometer measures the pressure and feeds the value into a pressure controller which throttles the gas flow with a butterfly valve to maintain a predetermined pressure setting. Vacuum is provided by a rotary vane pump which enables us to obtain pressures down to a few torr during growth. Temperature is monitored by the infrared temperature measurement setup and fed into a controller which varies the lamp power. Finally the reaction byproducts pass through a burn box, which oxidizes them before they are exhausted to the atmosphere.

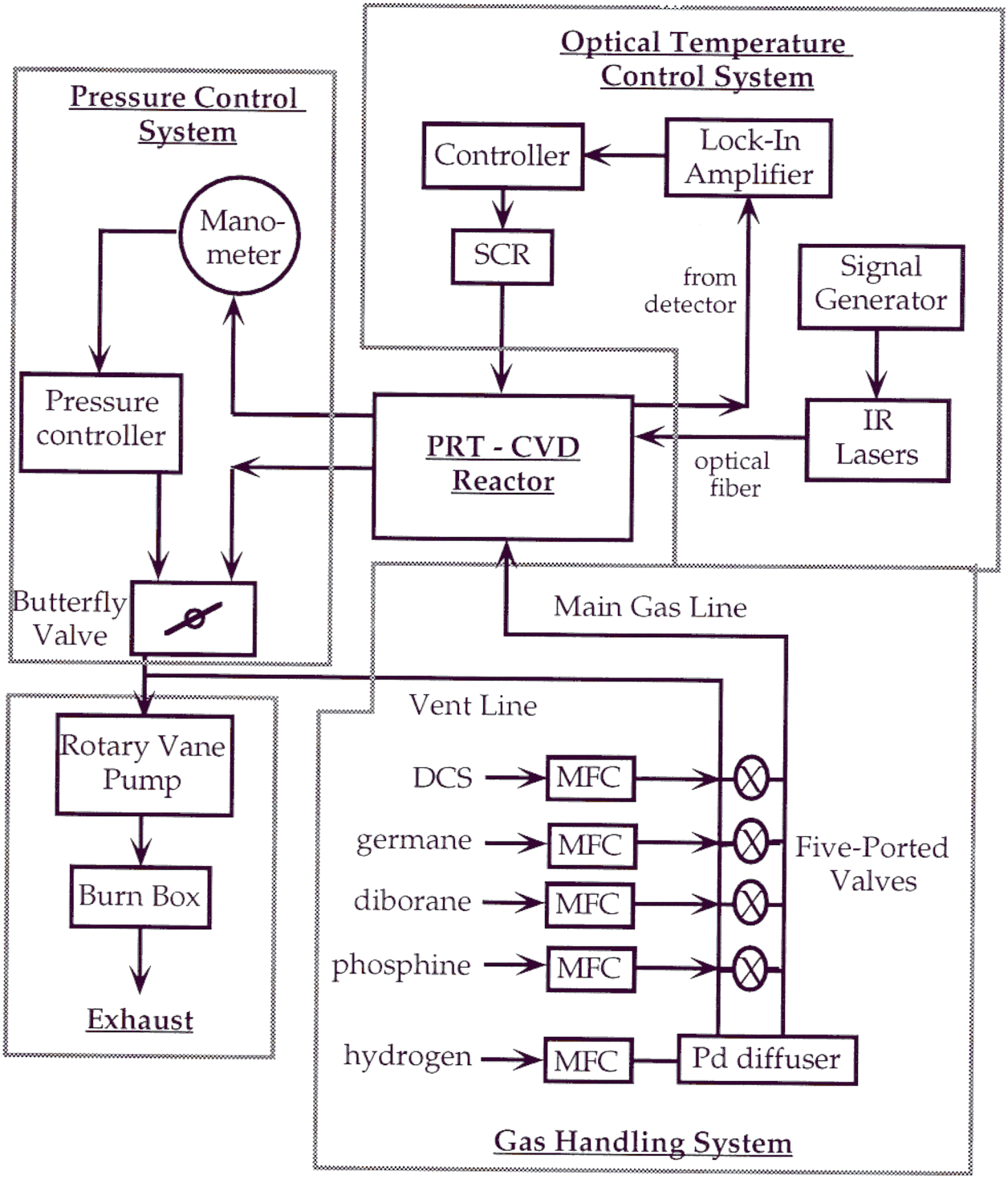


Figure 2.3 Schematic of the complete Princeton Rapid Thermal CVD reactor system.

2.2.2 Infrared Transmission Temperature Monitoring

Aside from temperature non uniformity, the greatest problem plaguing rapid thermal processes is temperature measurement. Pyrometry is typically used to measure temperature but variation in the wafer emissivity, due to the presence of other material such as oxides and nitrides, the wafer backside polish and doping, the measured wavelength, and temperature[8], make it an unreliable technique[9]. The infrared transmission temperature monitoring scheme developed at Princeton[10] avoids the problems associated with pyrometry and provides a highly accurate way of monitoring temperature.

The infrared absorption technique relies on the strong temperature dependence of the absorption of sub-bandgap radiation in the temperature range from 400° to 850° C At room temperature a silicon wafer is practically transparent to .3 μm and .5 μm radiation which have a photon energies of 0.96 eV and 0.83 eV respectively. When the temperature reaches 400° C absorption from free carriers and indirect band to band transitions begin to become significant. As the temperature increases the silicon bandgap narrows and the concentration of free carriers increases leading to an increase in the absorption from both of these mechanisms. In Figure 2.4 the normalized transmission for .3 μm and 1.5 μm radiation through a 450 μm thick Si wafer is plotted as a function of the wafer temperature. As the temperature increases, the transmission of the .3 μm laser, which has a photon energy closer to the band edge, drops off more rapidly. At 625° C the 450 μm thick silicon wafer is essentially opaque to the .3 μm laser. The .5 μm laser extends the range of the technique to almost 850° C before the transmission drops to zero. Simple absorption models indicate that bandgap absorption is the dominant mechanism for the .3 μm laser and free carrier

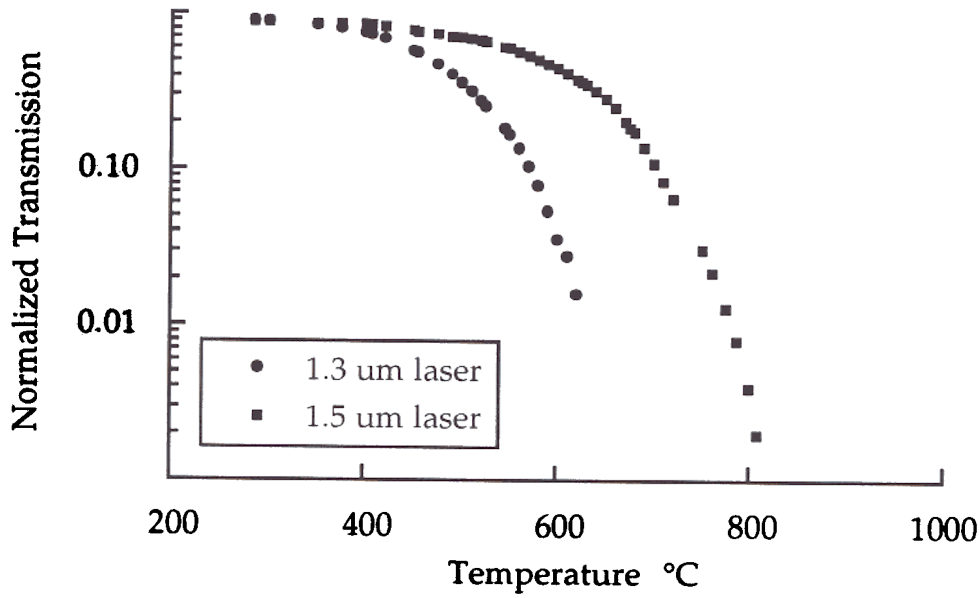


Figure 2.4 Normalized transmission through a 450 μm thick silicon wafer for a .3 μm and a 1.5 μm laser versus temperature.

absorption is the dominant mechanism for the .5 μm laser[1] The transmitted intensity at a given temperature, I(T), is given by

Eq. 2.2 $$I(T) = L_o I_o e^{-\alpha(T) t_s}$$

where I_o is the incident intensity, $\alpha(T)$ is the absorption coefficient at temperature T, t_s is the wafer thickness and L_o is the fraction of radiation that is not lost to reflection and scattering from the wafer backside roughness. Theses losses are assumed to be independent of temperature so that L_o cancels out if the transmission is *normalized* by dividing the transmission at a given temperature by the transmission at room temperature. The normalized transmission is

Eq. 2.3 $$N(T) = \frac{I(T)}{I(300)} = \frac{L_o I_o e^{-\alpha(T) t}}{L_o I_o e^{-\alpha(300\,K) t}} = e^{(-\alpha(T) + \alpha(300)) t}$$

Since the absorption at room temperature is essentially zero for the .3 μm and .5 μm radiation the normalized transmission reduces to

Eq. 2.4 $$N(T) \approx e^{-\alpha(T) t_s}$$

The transmission depends exponentially on thickness so the wafer thickness must be known to obtain accurate results The normalized transmission for a desired wafer thickness t_x can be determined from the normalized transmission of a reference wafer of thickness t_s using

Eq.2.5 $$N_x(T) = N_S(T)^{\left(\frac{t_x}{t_s}\right)}$$

where $N_x(T)$ and $N_S(T)$ are the normalized transmission at temperature T for the new wafer and the reference wafer respectively

The thickness for a 100 mm silicon wafer is required to be 525 μm ± 25 μm by SEMI standards[12]. This thickness variation will result in a ± 5 C° error in the temperature control if no correction is made The change in thickness of the wafer during epitaxial growth is insignificant since the maximum epitaxial layer thickness is typically less than 2-3 μm which corresponds to a thickness change of less than 0.6%

Thick films of Ge_xSi_{1-x} can be a problem with this technique since the bandgap of Ge_xSi_{1-x} is closer to the photon energies of the .3 μm and .5 μm lasers. Many of the Ge_xSi_{1-x} films of interest are very thin because of critical thickness considerations so, despite the higher absorption coefficient of the Ge_xSi_{1-x} films, the transmission will only change slightly because the optical path length through the Ge_xSi_{1-x} is small. In order to see the magnitude of the transmission change an epitaxial Ge_xSi_{1-x} layer was grown on a wafer with a thermocouple bonded to it The temperature was controlled using the thermocouple and the transmission was monitored as the Ge_xSi_{1-x} film grew After growing a 200 nm $Ge_{.3}Si_{.7}$ film only a 12% reduction in the normalized transmission at 700° C was seen This corresponds to an error in temperature of less than 5° C At 750° C where the bandgap is closer to the laser photon energy an error in the infrared temperature measurement of 20 C° was seen A thickness of 200 nm is over twenty-five times the equilibrium critical thickness for a $Ge_{.3}Si_{.7}$ film (h^*=7.6 nm)

A pure germanium layer was also grown on the thermocouple bonded wafer A film thickness of less than 50 nm was enough to reduce the infrared transmission to zero so

some care must be taken with this technique when growing thick Ge_xSi_{1-x} films or Ge_xSi_{1-x} films with germanium fractions larger than ≈0.5.

When used carefully the infrared technique is capable of maintaining the temperature within 1-2° C of the target value over the temperature range from 500° C to 800° C [17]

2.2.3 Source Gases for Epitaxial Growth

Dichlorosilane ($SiCl_2H_2$) was used as the silicon source and germane (GeH_4) was used as the germanium source. Dichlorosilane is a liquid at room temperature but the vapor pressure is high enough to use without a bubbler psi). Germane was diluted in hydrogen to a mole percent ranging from 0.6 to 0.9 %

The dopants were supplied by diborane (B_2H_6) and phosphine (PH_3) diluted in hydrogen. The diborane was typically diluted to 10 ppm and the phosphine to 70 ppm.

Hydrogen was used as the carrier gas. Since hydrogen constituted the bulk of the gas flow it was passed through an in-line palladium diffuser to remove any trace impurities (especially oxygen and water vapor).

2.3 Epitaxial Growth Rates

2.3.1 Thin Film Analysis

Basic information about the epitaxial film quality and composition was obtained using a number of different analytical techniques. The two techniques used most heavily in this work were Secondary Ion Mass Spectroscopy (SIMS) and Rutherford Backscattering (RBS). Rutherford Backscattering was used to do quantitative analysis of the Ge_xSi_{1-x} compositions and to check the crystallinity.

Almost all of the growth rate data presented in this thesis and all the boron doping data was determined from SIMS analysis. A Ge_xSi_{1-x} calibration standard with three different

germanium fractions was grown and measured by RBS and boron and phosphorous implants were used as dopant standards. All depth calibrations were done by measuring the crater depth The germanium compositional analysis was accurate to within approximately ± 3% and the doping concentrations agreed with results of electrical measurements within a factor of two

2.3.2 Silicon and Germanium Epitaxial Growth

The growth rate of silicon using a flow of 26 standard cubic cm (sccm) of dichlorosilane (SiH_2Cl_2) diluted in 3.0 liters per minute (lpm) of hydrogen at 6.0 torr was measured over the range of growth temperatures from 1000° C to 625° C An Arrhenius plot of the results is shown in Figure 2.5 The transition from the mass flow limited growth regime to the reaction rate limited growth regime occurs at a

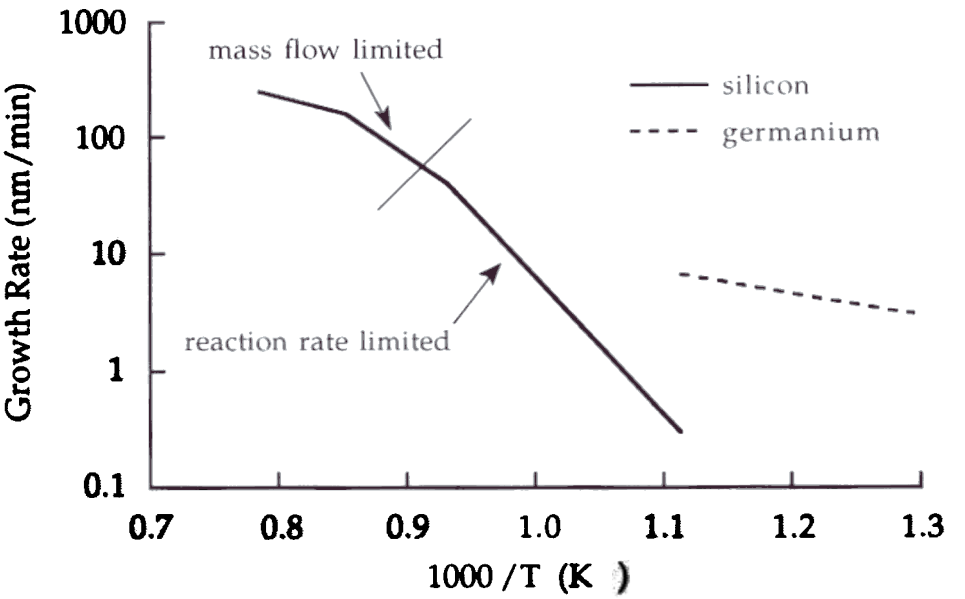


Figure 2.5 Arrhenius plot of the growth rate of silicon with a flow of 26 sccm of dichlorosilane and of germanium with a flow of 3.0 sccm. The growth was performed at 6.0 torr with a 3.0 lpm flow of hydrogen.

temperature of approximately 800° C [13]. In the reaction rate limited regime the slope of the Arrhenius plot indicates an activation energy of 1.9 eV. Below 650° C the growth rate of silicon is less than 0.5 nm/min.; this is too slow to be practical for a single-wafer reactor and leads to a practical lower limit for the epitaxial growth temperature with SiH_2Cl_2.

The growth rate of germanium using 3.0 sccm of germane was measured over the temperature range of 500° C to 700° C and plotted in Figure 2.5. The apparent activation energy for germanium growth is much smaller than that of silicon (0.38 eV). It has been postulated that germanium growth in this temperature range may be mass flow limited[14]

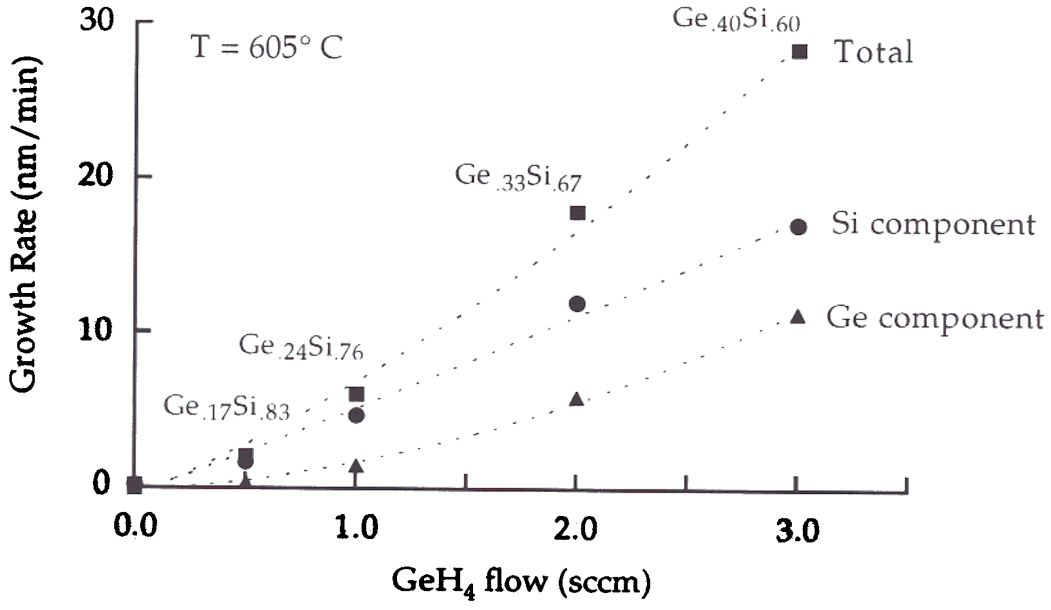


Figure 2.6 Growth rate of Ge_xSi_{1-x} epitaxial films as a function of germane flow at 605° C. The growth rate is also split up into its germanium and silicon components according to Eq. 2.16a and Eq. 2.16b.

2.3.3 Ge_xSi_{1-x} Epitaxial Growth

As small flows of germane (1-3 sccm) are added to the SiH_2Cl_2 flow (26 sccm) the growth rate increases dramatically in the reaction rate limited growth regime. At 605° C

the growth rate rises from less than 0.2 nm/min with no added germane to 28.4 nm/min with the addition of 3.0 sccm of germane. This is an increase of over 100 times and can't be attributed solely to the additional growth of germanium. In Figure 2.6 the growth rate at 605° C is plotted versus the germane flow. The total growth rate is also broken up into the germanium and silicon growth components using the germanium fraction measured by SIMS.

Eq. 2.6a $$R_G^{Si} = (1-x) \cdot R_G^{total}$$

and

Eq. 2.6b $$R_G^{Ge} = x \cdot R_G^{total}$$

where R_G^n is the growth rate component for species n and x is the germanium fraction. As germane is added to the SiH_2Cl_2 flow the component growth rates of both germanium and silicon increase. The increased silicon growth rate, with no corresponding change in the dichlorosilane partial pressure, indicates that the addition of germane is catalyzing the silicon growth.

The catalytic effect of germane is also seen in the growth rates of Ge_xSi_{1-x} films at 625° C and 700° C as shown in Figure 2.7 and Figure 2.8 respectively. The size of the catalytic effect diminishes as the growth temperature increases. When a 2.0 sccm flow of germane is added to the 26 sccm flow of dichlorosilane a growth rate enhancement of 90 times is seen at 605° C. The growth rate enhancement drops to 65 times at 625° C and 12 times at 700° C. The catalytic effect is more significant at lower temperatures because the silicon growth rate is falling off exponentially. In the mass flow limited growth regime (T>800° C) no catalytic effect is seen.

As the growth temperature is lowered the films become increasingly germanium rich. This is because the growth rate of the germanium component is only weakly dependent on temperature, like the growth rate of pure germanium. The germanium growth rate component for a germane addition of 2.0 sccm to 26 sccm of dichlorosilane is 7.6 nm/min at 700° C and only drops to 5.9 nm/min at 605° C, a reduction of only 22%. The silicon

growth rate component, in the same films, drops from 27.4 nm/min at 700° C to only 12.0 nm/min at 625° C – a reduction of over 100%.

At higher temperatures the growth of Ge_xSi_{1-x} epitaxial layers tends to be three-dimensional (also referred to as islanding). The high levels of stress in the films make it energetically favorable for the growing Ge_xSi_{1-x} layer to ball

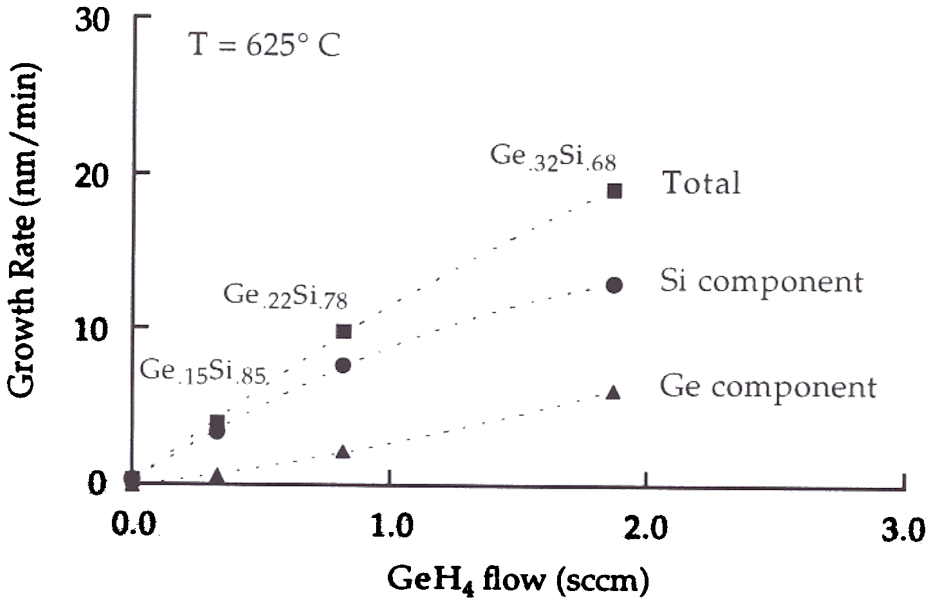


Figure 2.7 Growth rate of Ge_xSi_{1-x} epitaxial films as a function of germane flow at 625° C. The growth rate is also split up into its germanium and silicon components according to Eq. 2.16a and Eq. 2.16b.

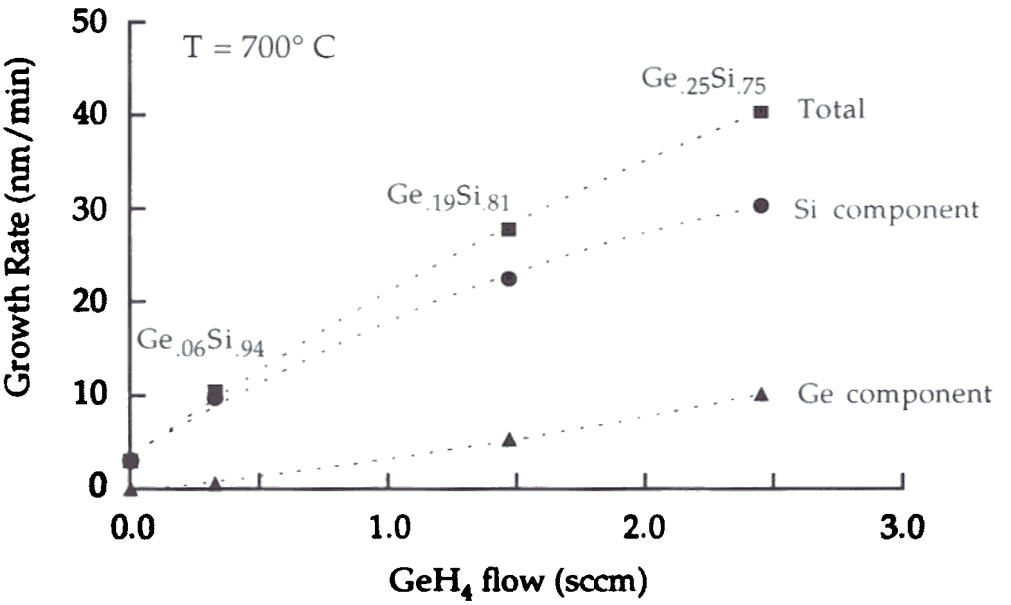


Figure 2.8 Growth rate of Ge_xSi_{1-x} epitaxial films as a function of germane flow at 700° C. The growth rate is also split up into its germanium and silicon components according to Eq. 2.16a and Eq. 2.16b.

up to reduce strain At higher growth temperatures, such as 700° C, the islanding is observable to the naked eye, giving the wafer a hazy appearance Three-dimensional growth is usually undesirable because any interface above the Ge_xSi_{1-x} layer will be very rough This interface roughness ruins the in-plane mobility (interface scattering) and makes the thickness of a Ge_xSi_{1-x} layer non uniform, broadening the energy levels of quantum wells At growth temperatures of 625° C and below, the surface appears perfectly flat even under a microscope using Nomarski (phase contrast microscopy)

Rough surfaces can also result from too high a partial pressure of oxygen or water vapor in the reactor The roughness in this case results from stacking faults and hillocks

2.3.4 Molecular View of Epitaxial Growth

The epitaxial growth process consists of a series of steps leading to the incorporation of the reactant species into the growing film. Figure 2.9 shows a schematic representation of the epitaxial growth process[15] which can be thought of as consisting of five major steps

1. Gas phase reactions.
2. Adsorption to the crystal surface.
3. Surface diffusion of the adsorbed species.
4. Dissociation of the species and incorporation into film.
5. Desorption of the reaction byproducts.

When modeling this complicated sequence of events it is usual to assume that one of these steps will be much slower than the others and will determine the reaction rate. If this is true, then all the steps prior to the rate determining step can be assumed to reach equilibrium and all the subsequent steps can be assumed to happen instantaneously.

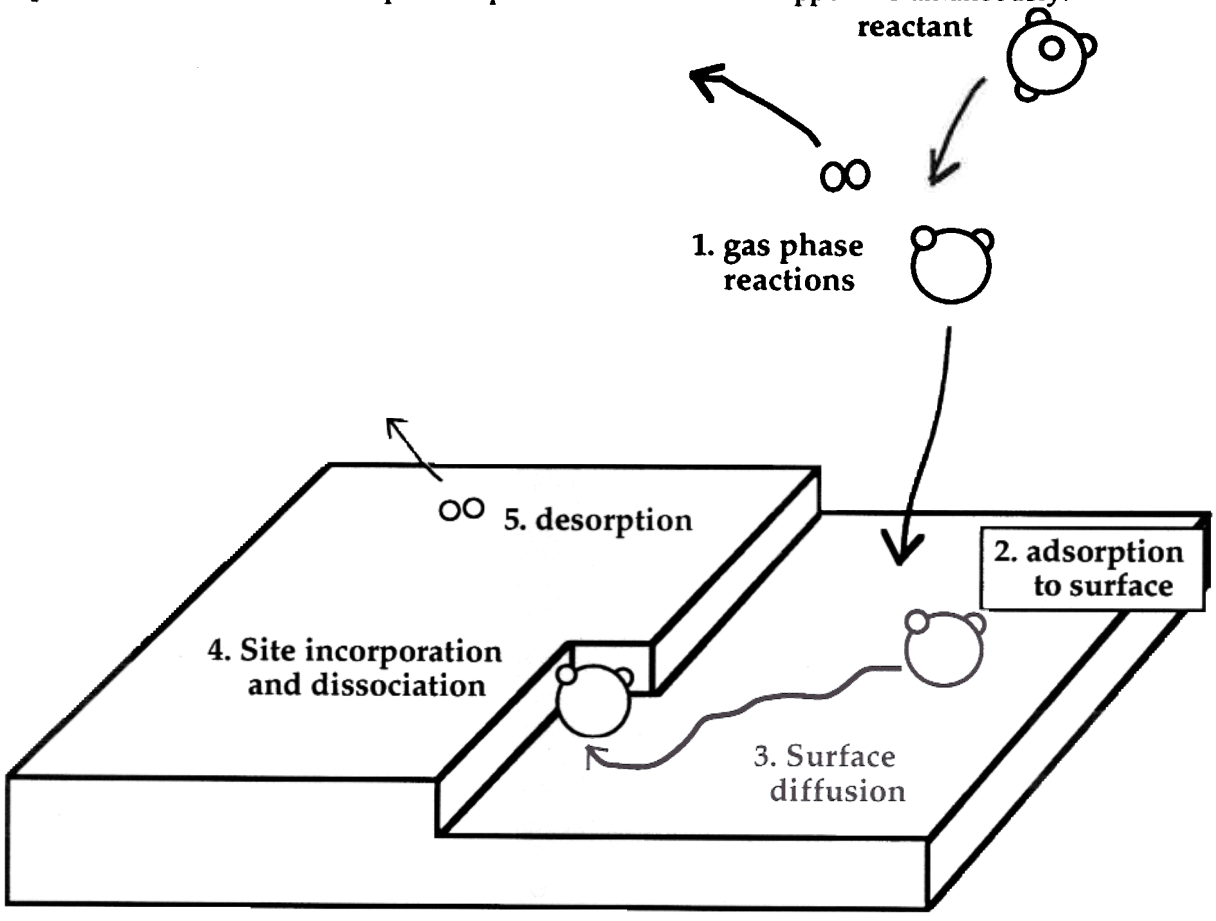


Figure 2.9 Schematic of a sequence of molecular reactions that lead to epitaxial growth from the vapor phase. After Reference [15].

2.3.5 Hydrogen Desorption Limited Growth Model

Epitaxial growth with silane and germane has a similar growth rate enhancement[16] to that of dichlorosilane and germane described earlier. The growth enhancement with silane is smaller - only a 25 times increase in the growth rate at 550° C with a 10% addition of germane to silane. The catalytic effect on growth rate of germane additions to silane was found to be consistent with increased hydrogen desorption from germanium on the growth surface. An increased desorption rate of hydrogen from the surface will result in a lower coverage of surface adsorption sites by hydrogen. The lower hydrogen coverage of the surface allows for increased adsorption of the growth species and hence an enhanced growth rate.

The following model[17] attempts to account for the growth rate enhancement seen with additions of germane to dichlorosilane by assuming that hydrogen desorption limits the growth rate. The ad-sites on the silicon surface are assumed to be predominantly blocked with adsorbed hydrogen. The growth rate enhancement, seen with germane addition, is postulated to result from a decreased surface coverage of hydrogen due to the weaker Ge-H bond (Compare the $^{+}$90.79 kJ/mol enthalpy of formation for germane to the $^{+}$32.64 kJ/mol value for silane[5]).

Additional (simplifying) assumptions are as follows:

1. Only one type of adsorbing species is important for Si and Ge growth and these result from simple decomposition reactions which we will take to be

Eq.2.7 $$SiCl_2H_2 = SiCl_2 + H_2$$

and

Eq.2.8 $$GeH_4 = GeH_2 + H_2$$

Note : the exact mechanism is *not* critical to the derivation.

2. All the adsorbing species (GeH_2, $SiCl_2$, and H) compete for the same type of adsorption site.

3. All Si and Ge containing species that adsorb contribute to the growth rate - i.e.

Eq.2.9 $$R_G^{total} = \sigma \times R_{ads}$$

or

Eq.2.10 $$SiCl_{2\,(g)} \rightarrow SiCl^*_{2\,(ads)} \rightarrow Si_{(s)}$$

where R_G and R_{ads} are the growth and adsorption rates respectively and σ is a geometric factor.

Since the adsorption rate is assumed to be rate limiting, preceding reaction steps can be considered to be in equilibrium. For the gas phase decomposition reactions this means that

Eq. 2. 1a $$P_{SiCl_2} = K_{Si}\, P_{SiH_2C}$$

and

Eq. 2. 1b $$P_{GeH_2} = K_{Ge}\, P_{GeH_4}$$

where P_{SiCl2}, P_{GeH4}, etc., are the partial pressures of the respective gas species and K_{Si} and K_{Ge} are the equilibrium constants. The adsorption rate for each species is assumed to be of the nondissociative Langmuir form[18] which may be written as

Eq.2.12 $$R_{ads} = (1-\theta)\, e^{\left(E^x_{ads}/kT\right)} R_{collision}$$

where $(1-\theta)$ is the fraction of available sites, E^x_{ads} is the activation energy for adsorption of species x, and $R_{collision}$ is the collision rate of the adsorbing species with the surface. The collision rates are given by the kinetic theory of gases as

Eq.2.13 $$R_C^{Si} = Y_S\, P_{SiCl_2}$$

Eq.2.14 $$R_C^{Ge} \quad Y_G\, P_{GeH_2}$$

Eq.2.15 and $$R_C^{H} = Y_H\, P_{H_2}$$

where Y_S is $\left(2\pi m_{SiCl_2} kT\right)^{(1/2)}$ and Y_G and Y_H are similarly defined.

The Polanyi-Wigner equation[19] for the first order hydrogen desorption process is :

Eq.2.16 $$R^{H}_{desorption} = \nu N \theta e^{-(E_d/kT)}$$

where $\nu \equiv$frequency factor, N = # of sites per cm^2, $\theta \equiv$fraction of filled sites, and E_d is the activation energy for H desorption. Eq.2.16 is modified to account for the proposed enhanced desorption of H due to the presence of Ge on the surface, by replacing the desorption probability $e^{-Ed/kT}$ by

Eq.2.17 $$F(x) = x \cdot e^{-(E_d^{Ge}/kT)} + (1-x) \cdot e^{-(E_d^{Si}/kT)}$$

where E_d^{Ge} and E_d^{Si} are the activation energies of hydrogen desorption from Ge and Si sites on the surface. We assume that $E_d^{Ge} < E_d^{Si}$ because of the weaker Ge-H bond so that the desorption rate increases as the germanium fraction increases.

The hydrogen adsorption and desorption will be equal under steady state growth conditions. Using Eq 2.12, Eq. 2.15, Eq. 2.16 and Eq. 2.17 the fraction of open sites is found to be

Eq. 2.18 $$\theta) = \frac{\nu N F(x)}{\left[\nu N F(x) + Y_H (P_{H_2})^{1/2} e^{-(E_{ads}^{H_2}/kT)}\right]}$$

Qualitatively Eq. 2.18 shows that as the Ge fraction of the film increases the number of available sites increases through the dependence on F(x), the modified hydrogen desorption probability.

The component growth rates of Ge and Si may be written in terms of the fraction of available adsorption sites (1–θ) as :

Eq. 2.19 $$R_G^{Si} = \sigma Y_S P_{SiH_2Cl_2} (1-\theta) e^{-(E_{ads}^{Si}/kT)}$$

and

Eq. 2.20 $$R_G^{Ge} \quad \sigma Y_G P_{GeH_4} (1-\theta) e^{-(E_{ads}^{Ge}/kT)}$$

Since the growth rates both contain the common terms (1-θ) we may write

$$(1-\theta) = \frac{R_G^{Ge} e^{+(E_{ads}^{Ge}/kT)}}{\sigma Y_G P_{GeH_4}} = \frac{R_G^{Si} e^{+(E_{ads}^{Si}/kT)}}{\sigma Y_S P_{SiH_2Cl_2}}$$

or

Eq. 2.21
$$\frac{R_G^{Ge} P_{SiH_2Cl_2}}{R_G^{Si} P_{GeH_4}} = \left(\frac{Y_G}{Y_S}\right) e^{+\left(E_{ads}^{Si} - E_{ads}^{Ge}\right)/kT} \equiv J$$

GeH_4 flow (sccm)	partial pressure (mtorr)		component growth rates (nm/min)		"J"
	SiH_2Cl_2	GeH_4	silicon	germanium	
0.5	8.4	0.16	1.7	0.4	**12.4**
1.0	8.2	0.32	4.7	1.5	**8.2**
2.0	7.9	0.60	12.0	5.9	**6.5**
3.0	7.5	0.87	17.1	11.3	**5.7**

Table 2. **Calculated value of the dimensionless parameter "J" (Eq. 2.31) using Ge_xSi_{1-x} growth rate data at 625° C and a total pressure of 6.0 torr. If hydrogen desorption is limiting the growth rate "J" should be constant.**

GeH_4 flow (sccm)	partial pressure (mtorr)		component growth rates (nm/min)		"J"
	SiH_2Cl_2	GeH_4	silicon	germanium	
0.33	8.5	0.11	9.8	0.6	**4.9**
1.47	8.0	0.45	22.6	5.3	**4.2**
2.45	7.7	0.72	30.4	10.1	**3.6**

Table 2.2 **Calculated value of the dimensionless parameter "J" (Eq. 2.31) using Ge_xSi_{1-x} growth rate data at 700° C and a total pressure of 6.0 torr. If hydrogen desorption is limiting the growth rate "J" should be constant.**

The parameters on the left-hand side of Eq. 2.21 are known for a set of growth conditions and the expression on the right side should be constant at a given temperature. We define this constant value for a given growth temperature and total pressure as the dimensionless parameter "J". The value of "J" is calculated from the growth rate data measured at 605° C in Table 2. and from the growth rate data measured at 700° C in Table 2.2 For growth of Ge_xSi_{1-x} at 605° C it is evident that the value of "J" is not

constant as predicted by the hydrogen desorption limited growth rate model, but falls off with increasing P_{GeH4}. At 700° C the parameter "J" is more nearly constant suggesting that the hydrogen desorption limited growth model may be more reasonable under these growth conditions.

The data suggests that the growth rate using dichlorosilane and germane sources at 700° C may be limited by hydrogen desorption from the surface. At 605° C this is evidently not the case and the growth rate may be limited by some other factor such as gas phase kinetics, surface nucleation, oxygen desorption (growth is not under ultra high vacuum conditions), or the influence of the chlorine from the dichlorosilane.

2.4 Boron Doping Kinetics

The boron dopant incorporation in silicon as a function of diborane flow (11ppm in H_2) was investigated at 700° C and 800° C. As may be seen in Figure 2.10 the boron doping was linear with B_2H_6 flow. This linearity extended above the solid solubility of 2.0×10^{19} cm^{-3} for silicon[20] at 700° C suggesting a kinetic dependence for the boron incorporation rather than a thermodynamic dependence. Equilibrium would set an upper limit to the doping levels equal to the solid solubility. It is also evident from Figure 2.10 that the boron doping in silicon at 800° C Si is a factor of ten lower than the boron doping at 700° C for equal diborane flows. At 800° C the solid solubility of boron is 4.5×10^{19} cm^{-3}, more than twice the solid solubility boron at 700° C. Equilibrium considerations would lead one to believe that

the higher the solid solubility, the higher the doping incorporation. The growth rate of silicon at 800° C is 45.0 nm/min, a factor of 15 higher than the growth rate at 700° C. This suggests that the boron dopant incorporation is inversely proportional to the growth rate and, as mentioned previously, directly proportional to the diborane flow :

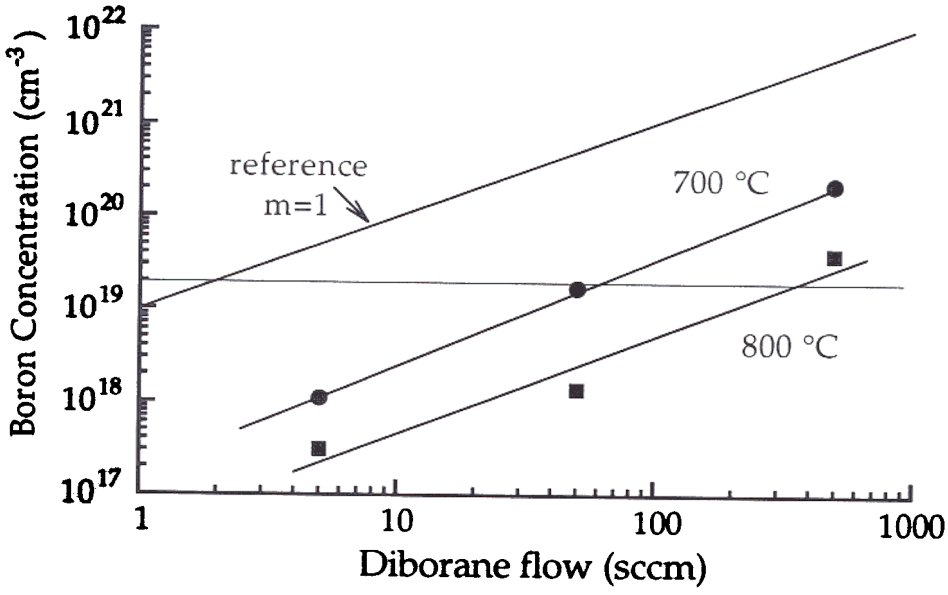


Figure 2.10 Boron concentration in silicon epitaxial films versus 10 ppm diborane flow rate for growth temperatures of 700° C and 800° C.

Eq. 2.22
$$C_{boron} \propto \frac{P_{B_2H_6}}{R_G}$$

The kinetic dependence of the boron dopant incorporation may be thought to result simply from the ratio of the diborane flux upon the sample to the silicon flux of the growing film. In other words the boron flux into the growing epitaxial film remains fixed for a given diborane flow but the boron concentration in the epitaxial film will be modulated by the rate of the concurrent epitaxial growth. In Ge_xSi_{1-x} films, where the growth rate depends on the germane flow, the boron dopant incorporation is also found to drop off linearly as the growth rate increases (see Figure 2 This further confirms the simple kinetic relation of Eq. 2.22.

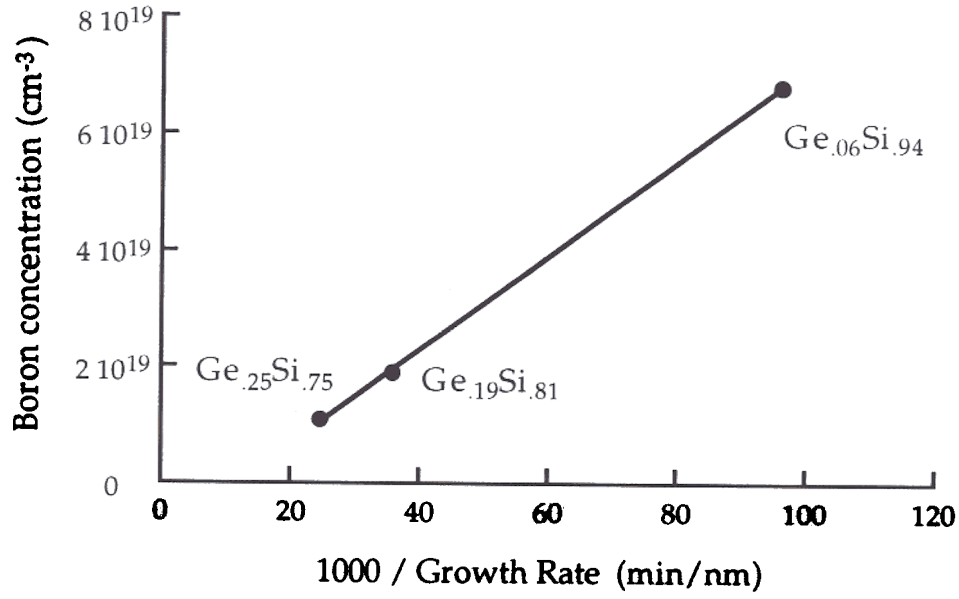


Figure 2.1 Boron concentration versus inverse growth rate for Ge_xSi_{1-x} films grown at 700° C.

2.5 Growth Summary

Good quality epitaxial silicon and Ge_xSi_{1-x} films can be grown at low temperatures in a low pressure epitaxial reactor (6 torr) if care is taken to reduce the partial pressures of oxygen and water vapor. Precise temperature control in the reaction rate limited growth regime (T<800° C) is essential because the growth rate at these temperatures is exponentially dependent on temperature. Infrared transmission has proven to be a reliable and highly accurate method for monitoring and controlling the temperature in this range.

It is found that the addition of germane to dichlorosilane catalyzes the epitaxial growth rate in the reaction rate limited growth regime allowing reasonable Ge_xSi_{1-x} growth rates to be achieved at temperatures down to 605° C.

In the reaction rate limited growth regime boron doping kinetics are determined by kinetics rather than by equilibrium considerations (solid solubility).

[1] S.M. Sze, *"VLSI Technology"*, 2nd edition, McGraw-Hill (1988).

[2] G.R. Srinivasan and B.S. Meyerson, *"Current Status of Reduced Temperature Silicon Epitaxy by Chemical Vapor Deposition"*, Journal of the Electrochemical Society vol.134, no.6 (1987), p.1518.

[3] F.W. Smith and G.Ghidini, "*Reaction of Oxygen with Si(111) and (100): Critical Conditions for the Growth of SiO_2*", Journal of the Electrochemical Society vol.129, no.6 (1982), p.1300.

[4] G. Ghidini and F.W. Smith, "*Interaction of H_2O with Si(111) and (100): Critical Conditions for the Growth of SiO_2*", Journal of the Electrochemical Society vol.131, no.12 (1984), p.2924.

[5] T.O. Sedgewick, M. Berkenblit and T.S. Kuan, "Low Temperature Selective Epitaxial Growth of Silicon at Atmospheric Temperature", Applied Physics Letters **54** (1989), p.2689.

[6] J.F. Gibbons, C.M. Gronet and K.E. Williams, "*Limited Reaction Processing : Silicon Epitaxy*", Applied Physics Letters **47**, no.7 (1985), p.721.

[7] S.K. Ghandhi, "*VLSI Fabrication Principles, Silicon and Gallium Arsenide*", Wiley-Interscience (1983), Chapter 4 - Diffusion, p.141.

[8] T. Sato, *"Spectral Emissivity of Silicon"*, Japanese Journal of Applied Physics **6** (1967), p.339.

[9] D.W. Pettibone, J.R. Suarez and A. Gat, Proceedings of the Material Research Society **52** (1986), p.209.

[10] J.C. Sturm, P.V. Schwartz, and P.M. Garone, *"Silicon Temperature Measurement by Infrared Transmission for Rapid Thermal Processing Applications"*, Applied Physics Letters **56** (1990), p.961.

[11] J.C. Sturm and C.M. Reaves, "*Physical Modeling of Non-Invasive Silicon Temperature Measurement by Infrared Absorption*", 1991 IEDM Technical Digest, p.895.

[12] Semiconductor Equipment and Materials Institute (SEMI) Mountain View, CA.

[13] A.S. Grove, "Physics and Technology of Semiconductor Devices - Chapter 1 : Vapor Phase Growth", Wiley (1967), p.10.

[14] T.I. Kamins and D.J. Meyer, *"Kinetics of Silicon-Germanium Deposition by Atmospheric-Pressure Chemical Vapor Deposition"*, Applied Physics Letters **59**, no.2, 8 July 1991, p.178.

[15] R.Reif, T. Kamins, and K.C. Saraswat, "*A model for dopant incorporation into Growing Silicon Epitaxial Films*", Journal of the Electrochemical Society **126** (1979), p.644.

[16] K.J. Uram and B.S. Meyerson, *"Cooperative Growth Phenomena in Silicon/Germanium Low-Temperature Epitaxy* Applied Physics Letters **53** (1988), p.2555.

[17] J.C. Sturm, P.V. Schwartz, S.A. Schwarz, and B. Wilkens, "*Silicon Vapor-phase Epitaxial Growth Catalysis by the Presence of Germane*" P.M. Garone, Applied Physics Letters 56, 1275 (1990).

[18] M.A. Morris. M. Bowker and D.A. King, *"Kinetics of Adsorption, Desorption and Diffusion at Metal Surfaces"*, Comprehensive Chemical Kinetics vol.19, Elsevier, Amsterdam (1984), p.65.

[19] M.A. Morris. M. Bowker and D.A. King, *"Kinetics of Adsorption, Desorption and Diffusion at Metal Surfaces"*, Comprehensive Chemical Kinetics vol.19, Elsevier, Amsterdam (1984), p.84

[20] S.K. Ghandi, "VLSI Fabrication Principles", Wiley-Interscience (1983), p.70.

Chapter 3 : MOS-Gated High Hole Mobility Transistors (MOS-HHMTs)

3.1 What is a MOS-HHMT and why is it relevant ?

3.1.1 Limitations of pMOS

The performance of CMOS circuits is largely limited by the low transconductance of pMOS transistors, resulting from the lower mobility of holes in silicon. The width of the pMOS devices must be increased (*sized*) in order to achieve symmetric rise/fall times in digital circuits. *Sizing* the pMOS transistors reduces the circuit density and increases the input capacitance of each logic gate. As a result improvements in pMOS performance will have a disproportionately large impact on CMOS circuit density and speed.

3.1.2 Inversion Layer Mobility

The mobility of carriers in the inversion layer of a MOSFET is significantly less than that of carriers in the bulk semiconductor. In addition to the phonon and ionized impurity scattering seen in bulk material, carriers in an inversion layer also suffer scattering induced by surface acoustic phonons [1], oxide fixed charge and the variation of the Si/SiO_2 interface potential caused by roughness of the oxide interface [2].

The inversion layer mobility is often considered by looking at the effective (field effect) mobility versus the effective transverse field caused by the gate potential (for example see reference [3]). The effective mobility (μ_{eff}) is extracted from the drain conductance curve of a MOSFET (see Figure 3.1) and is taken to be

Eq. 3.1 :
$$\mu_{eff} = \left(\frac{L}{W}\right)\left(\frac{I_D}{V_D}\right)\left(\frac{1}{C_{ox}(V_G - V_T)}\right)$$

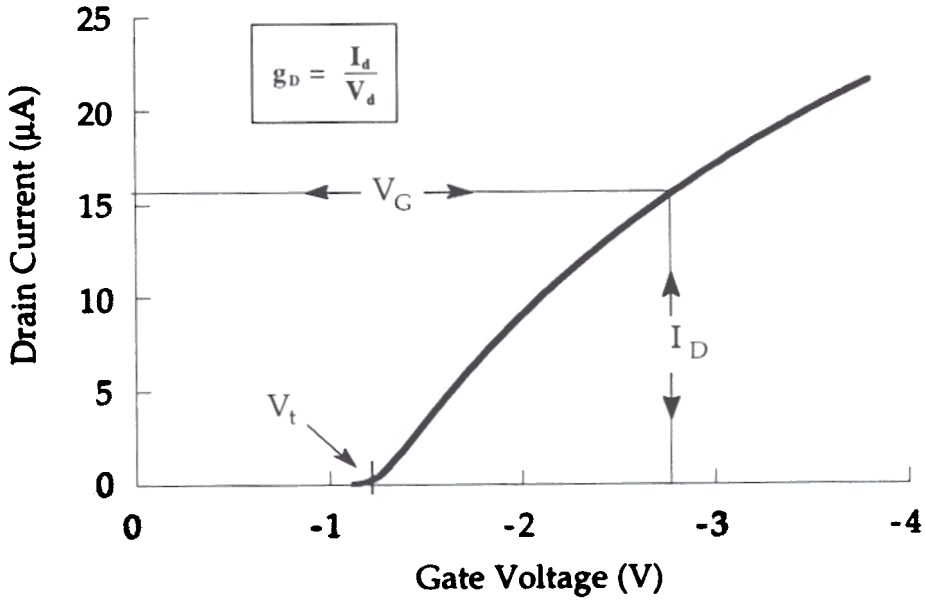


Figure 3 Typical drain conductance curve for a MOSFET.

where $\left(\frac{L}{W}\right)$ is the ratio of the gate length over the width, $\left(\frac{I_D}{V_D}\right)$ is a measure of the drain conductance at a given gate bias, and C_{ox} $(V_G - V_T)$ is the difference between the gate potential and the threshold voltage times the oxide capacitance, which is an estimate of the inversion charge density

The *effective* vertical field (E_{eff}) is defined as the average electric field resulting from the gate potential which is *felt* by the carriers in the inversion layer The effective vertical field is taken to be

Eq. 3.2
$$E_{eff} = \frac{(Q_{dep} + \eta Q_{inv})}{\varepsilon_{Si}}$$

where ε_{Si} is the dielectric constant of silicon, Q_{dep} is the depletion charge, Q_{inv} is the inversion charge, and η is a weighting factor that takes into account the partial screening of the gate field by other carriers The weighting factor

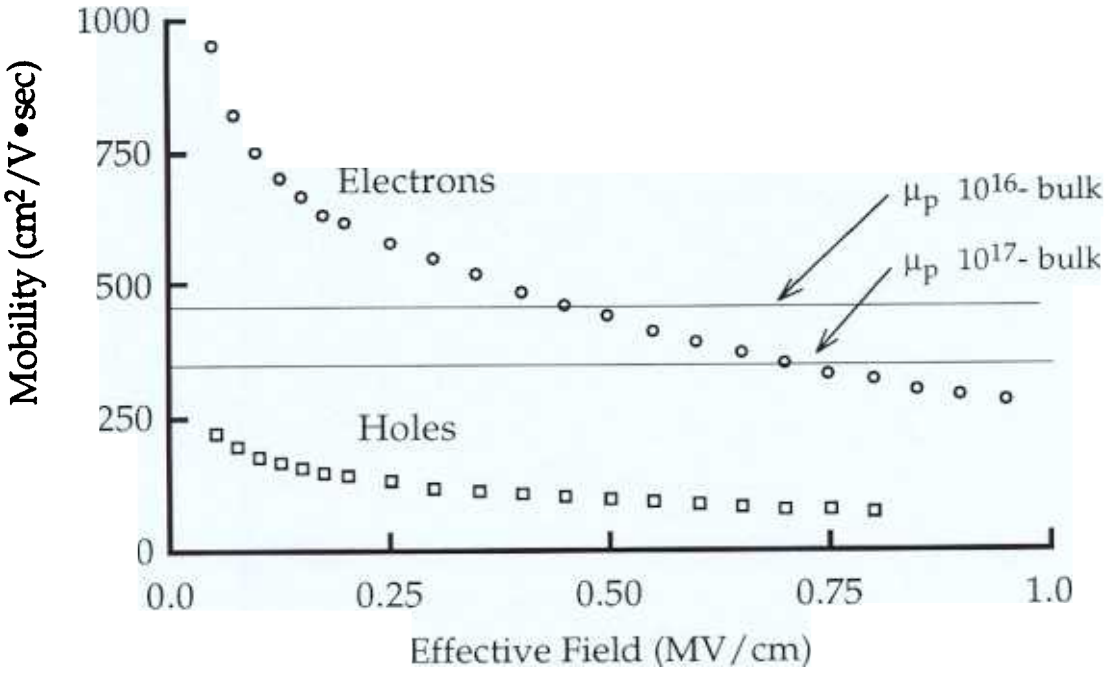


Figure 3. 2 An example of the effective mobility vs. effective vertical field plot for electrons and holes. Data taken from reference [6].

is usually taken to be $\eta = 1/2$ for electrons [4] and $\eta = 1/3$ for holes [5] on a Si (100) surface. The physical significance of the weighting factor is not understood theoretically. When the appropriate weighting factor is chosen the effective mobility will follow a "universal" curve of effective mobility vs. effective vertical field irrespective of oxide thickness, doping density, and substrate bias [5].

A typical example of a plot of the effective mobility versus effective vertical field, taken from data by Watt and Plummer [6], is shown in Figure 3.2. At low effective vertical fields (gate voltages just above threshold), the peak effective mobility is limited by ionized impurity scattering and interface fixed charge. After reaching a peak value, which depends on the doping level and the fixed charge in the oxide, the effective mobility decreases along a "universal" curve with increasing effective vertical field. The decrease in mobility at higher fields is a result of the carriers being confined more closely to the Si/SiO_2 interface

as the effective vertical field increases. As the carriers move closer to the Si/SiO_2 interface they experience increased interface scattering This can be qualitatively understood by looking at a simple picture showing the carrier potential at two different effective vertical fields and the corresponding wave functions At smaller effective vertical fields the potential formed by the gate field is relatively shallow and broad and the corresponding carrier probability distribution is spread out (Figure 3.3a) As the field gets stronger the potential becomes deeper and narrower forcing the carrier probability distribution closer to the interface (Figure 3.3b). As the wavefunction is drawn closer to the interface the average spacing of the carriers from the Si/SiO_2 interface (Z_{avg}) gets smaller, therefore interface scattering becomes more probable and drives the effective mobility down

It is obvious from Figure 3.2 that the effective mobility of holes is only about one third that of electrons. Thus the discrepancy in mobility between electrons and holes persists in an inversion layer The two horizontal lines shown in Figure 3.2 represent the bulk hole mobility for doping levels of 10^{16} cm^{-3} and 10^{17} cm^{-3}. The bulk hole mobility values are very similar to the electron effective mobilities across the range of the effective vertical field of the gate This raises the question, "*What if a pMOS device with a more bulk-like hole mobility could be designed?*". If this could be done then the difference in transconductance between the nMOS and pMOS devices would

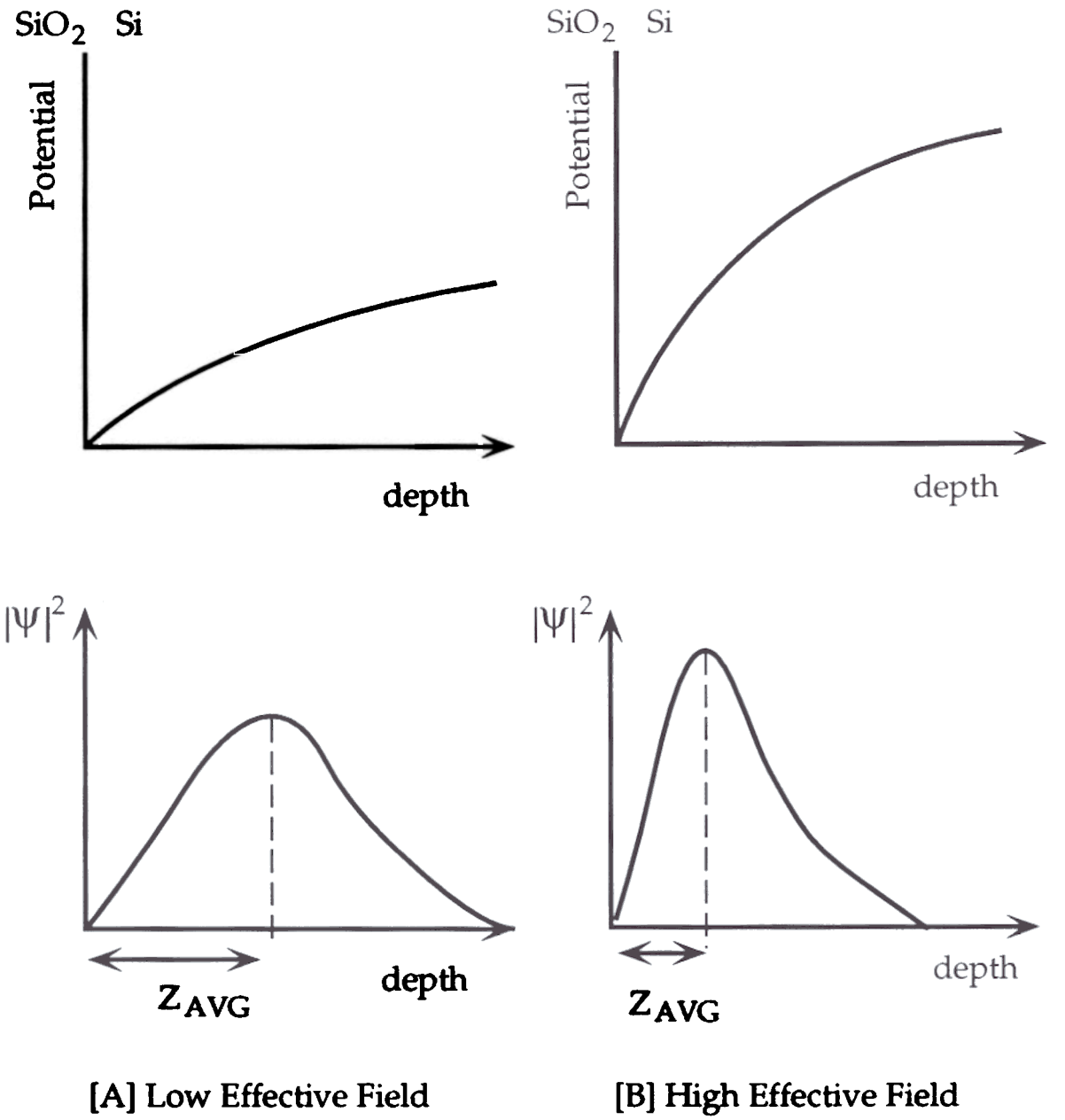


Figure 3. 3 Sketch of the hole potential and hole probability distribution underneath the gate of a MOSFET at both low and high effective vertical fields.

not be as large and a more symmetric CMOS circuit could be developed, reducing the size and input capacitance of the logic gates.

3.1.3 Introducing the MOS-HHMT

One structure that has been proposed [6,7,8] moves carriers away from the Si/SiO_2 interface by adding a buried Ge_xSi_{1-x} layer underneath the gate of a standard MOSFET structure. The Ge_xSi_{1-x} layer (*well*) is separated from the Si/SiO_2 interface by a thin (5.0 to 10.0 nm) layer of silicon. Because of the valence band discontinuity it is possible to form an inversion layer at the Ge_xSi_{1-x}/Si interface before the Si/SiO_2 interface also inverts. The silicon layer between the Ge_xSi_{1-x} and the oxide acts as a spacer separating the inversion layer in the Ge_xSi_{1-x} well from the scattering sites at the Si/SiO_2 interface.

The term MOS-HHMT (MOS-gated High Hole Mobility Transistor) is used in this thesis to refer to this device. MOS refers to the gate and HHMT is reminiscent of HEMT which also uses a semiconductor heterojunction to confine carriers.

In addition to forming a semiconductor heterojunction the Ge_xSi_{1-x} layer and silicon spacer change the internal device capacitances and threshold voltage of the MOS-HHMT compared to a standard MOSFET. Figure 3.4 depicts the cross-section of a MOS-HHMT device.

3.2 How Does it Work?

3.2.1 Band Diagrams

The operation of the MOS-HHMT can be understood by considering the band diagram with different applied gate biases. Figure 3.5a shows the band

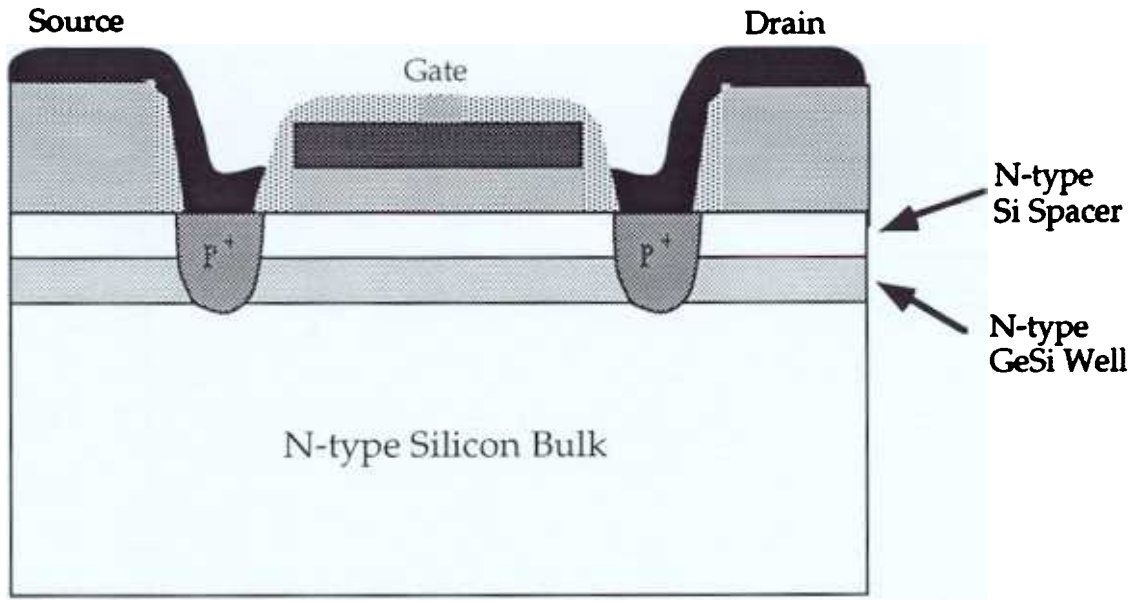


Figure 3. 4 Cross-section of a MOS-HHMT. The buried GexSi1-x layer is separated from the gate oxide by a thin 5.0-10.5 nm epitaxial silicon spacer layer. Implanted souce/drains contact both the spacer layer and the GexSi1-x layer.

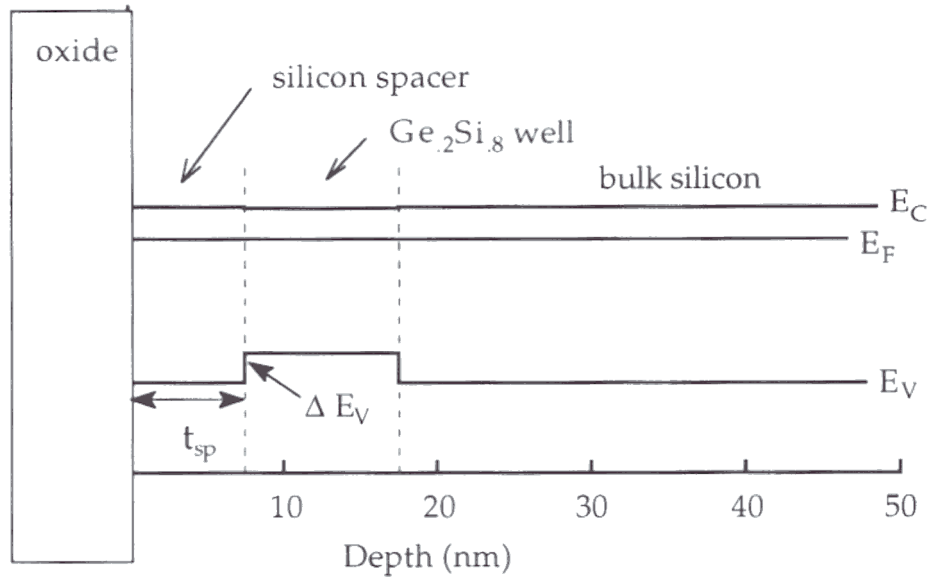


Figure 3.5a Energy band of MOS-HHMT near flatband for a device with a $Ge_{.2}Si_{.8}$ well and a 7.5 nm silicon spacer layer. Note the large valence band discontinuity at the $Ge_{.2}Si_{.8}$ /Si interface. (Oxide bandgap is not to scale.)

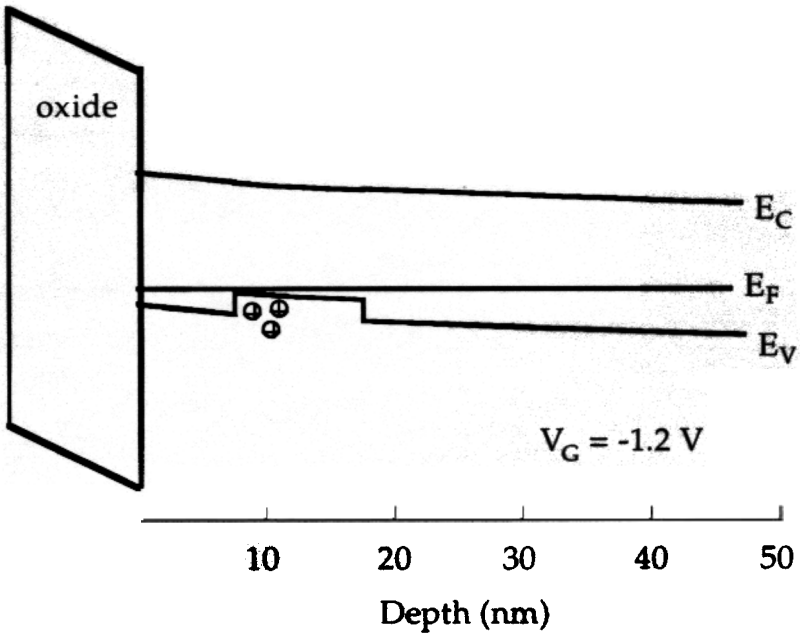


Figure 3.5b MOS-HHMT band diagram with an applied gate voltage of -1.2 volts. An inversion layer has formed in the $Ge_{.2}Si_{.8}$ well but not at the Si/SiO_2 interface.

diagram near flatband for a structure with a 7.5 nm Si spacer layer and $Ge_{.2}Si_{.8}$ well. Most of the bandgap offset at the Ge_xSi_{1-x}/Si interface is in the valence band ($\Delta E_V \approx 165$ meV) forming a potential well for holes and enabling the Ge_xSi_{1-x} well to invert before the Si/SiO_2 interface. When a small negative gate bias is applied (Figure 3.5b) an inversion layer first forms in the Ge_xSi_{1-x} well with no corresponding inversion layer at the Si/SiO_2 interface. However as the magnitude of the gate bias is increased an inversion layer does eventually form at the Si/SiO_2 interface (Figure 3.5c). One of the goals in optimizing the device structure is to maximize the number of carriers added to the Ge_xSi_{1-x} well where a higher -more bulk-like- mobility is anticipated.

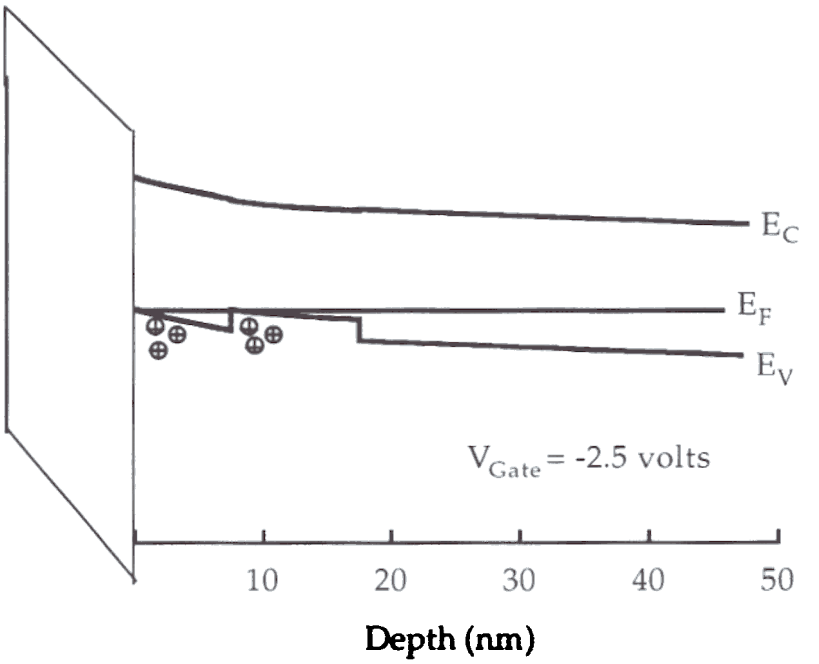


Figure 3.5c MOS-HHMT band diagram with an applied gate voltage of -2.5 volts. An inversion layer has now also formed at the Si/SiO_2 interface

3.2.2 Simple Capacitor Model

The operation of the MOSFET is largely determined by the capacitive control of the surface potential (ϕ_S) by the gate. Looking at a gate/substrate cross-section, the MOSFET input can be modeled as two capacitors in series, corresponding to the gate oxide (C_{ox}) and the depletion layer (C_{dep}) as shown in Figure 3.6. In a MOS-HHMT, charge can also build up at the Ge_xSi_{1-x}/Si interface (ϕ_G) so an additional capacitance associated with the silicon spacer layer (C_{SP}) must be included (Figure 3.6).

The most important consequence of the change in gate capacitance is that the capacitive coupling of the gate potential V_G to the potential in the Ge_xSi_{1-x} well ϕ_G, is weaker, $C_G^{-1} = C_{SP}^{-1} + C_{ox}^{-1}$, than the capacitive coupling of gate potential to the potential at the Si/SiO_2 interface ϕ_S, $C_G = C_{ox}$

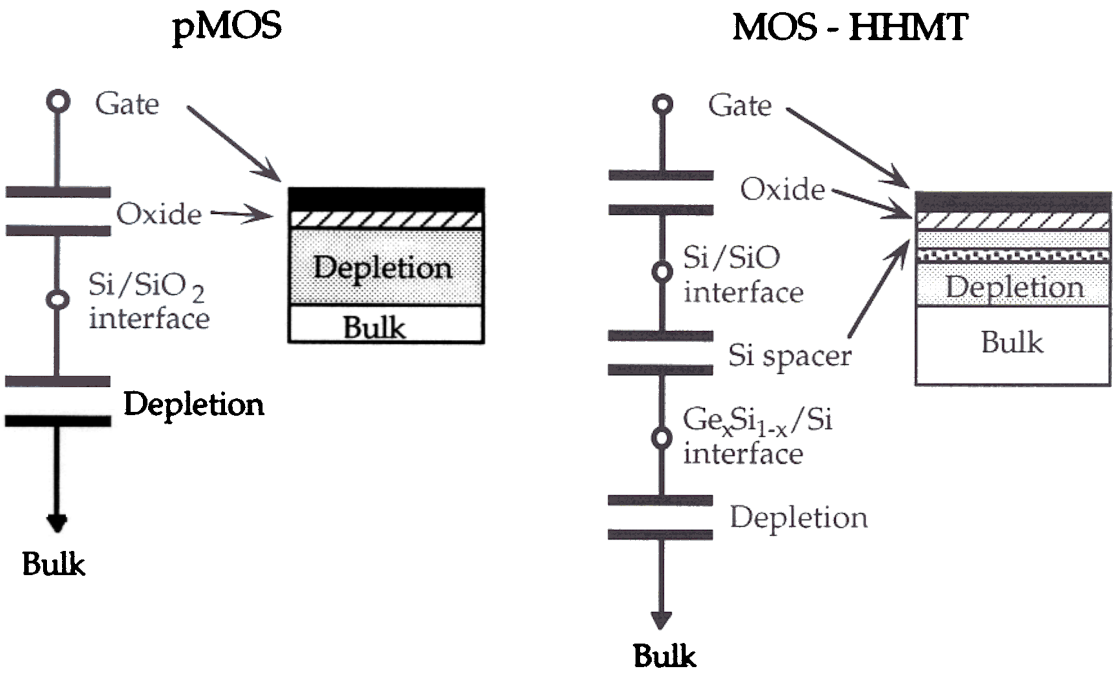


Figure 3. 6 : Simple capacitor model of the gate-subtrate connection. Comparison between a MOSFET and a MOSHHMT. The addition of the silicon spacer layer capacitance differentiates the two.

This means that the number of carriers in the inversion layer (Q_{inv}) of a MOS-HHMT at a given gate voltage above threshold (V_T) will be less than that in a silicon MOSFET since

Eq. 3.3 $$Q_{inv} \approx C_G(V_G - V_T)$$

The addition of the Ge_xSi_{1-x} layer also changes the depletion capacitance (discussed in section 3.2.6). These changes in the gate capacitance will also effect the threshold voltage (section 3.2.5) and the subthreshold swing (section 3.2.6).

3.2.3 One Dimensional Poisson Solver for a MOS capacitor

In order to better understand and predict the behavior of the MOS-HHMT a simple program was developed to find the one-dimensional electrostatic solution to Poisson's equation for an MOS capacitor at a given gate bias. Details of the semiconductor vertical composition and doping profiles as well as the gate oxide thickness, fixed charge and interface state density are input interactively. The gate voltage is also input and a trial

surface potential is selected by the program The choice of the surface potential determines the surface electric field and the solution to Poisson's equation for these initial conditions is propagated using the fourth order Runge-Kutta method. The surface potential is refined during successive iterations until convergence is reached when the potential in the bulk semiconductor goes to zero (within 1 mV). A constant quasi Fermi level in the thin Si and Ge_xSi_{1-x} layers equal to that deep within the bulk was assumed (i.e. no channel-substrate bias). Fermi-Dirac statistics are used to obtain carrier concentrations Information regarding the band offsets was taken from calculations by Van de Walle and Martin for coherently strained Ge_xSi_{1-x} on silicon substrates[9], the strain induced splitting of the band degeneracies is included The density of states for the individual bands was assumed to be one third that of silicon

Using this program the hole density in the Ge_xSi_{1-x} well and at the Si/SiO_2 interface can be simulated for a given epitaxial structure The simulated device structures had a gate oxide thickness of 10.0 or 12.5 nm, a 10.0 nm Ge_xSi_{1-x} well, and a thin Si spacer layer The silicon spacer thickness and germanium fraction in the well were varied. The doping was fixed at 10^{16} cm^{-3} n-type. More details on the routines used in the Poisson solver and assumptions made are found in Appendix II

3.2.4 Hole Density vs. Gate Voltage

In order to compare different MOS-HHMT structures the device operation must be understood in a quantitative way Therefore the dependence of

oxide capacitance. A kink in the slope of the total hole density curve due to the changing gate capacitance will therefore be seen at the onset of the Si/SiO_2 inversion. Inversion of the Si/SiO_2 interface effectively limits the number of holes in the Ge_xSi_{1-x} well by *pinning* the surface potential

In order to compare the effectiveness of different structures in confining holes to the Ge_xSi_{1-x} well a figure of merit which will be referred to as *crossover* is introduced Crossover is defined as the number of holes in the Ge_xSi_{1-x} well when the number of holes at the Si/SiO_2 interface equals in the number of holes in the Ge_xSi_{1-x} well Crossover is indicative of the maximum number of holes that can be put into the Ge_xSi_{1-x} well for a given structure

In Figure 3.8 the fraction of holes in the Ge_xSi_{1-x} well is plotted as a function of gate voltage for two structures with a $Ge_{.2}Si_{.8}$ well and different spacer layer thicknesses At typical operating voltages less than 40% of the carriers are found in the Ge_xSi_{1-x} well. It is also apparent that the structure with the larger spacer layer thickness (weaker capacitive coupling to the Ge_xSi_{1-x} well) has a smaller fraction of carriers in the Ge_xSi_{1-x} well. Only the carriers added to the Ge_xSi_{1-x} well will serve to enhance the transconductance of the FETs, thus one of the constraints in designing a MOS-HHMT structure is to have the maximum number of carriers in the Ge_xSi_{1-x} well

3.2.4.2 Dependence of the Hole Density in the Ge_xSi_{1-x} Well on the Structure.

A simple analytical formula for the hole density in the Ge_xSi_{1-x} well at crossover can be derived by considering the hole densities in the Ge_xSi_{1-x} well

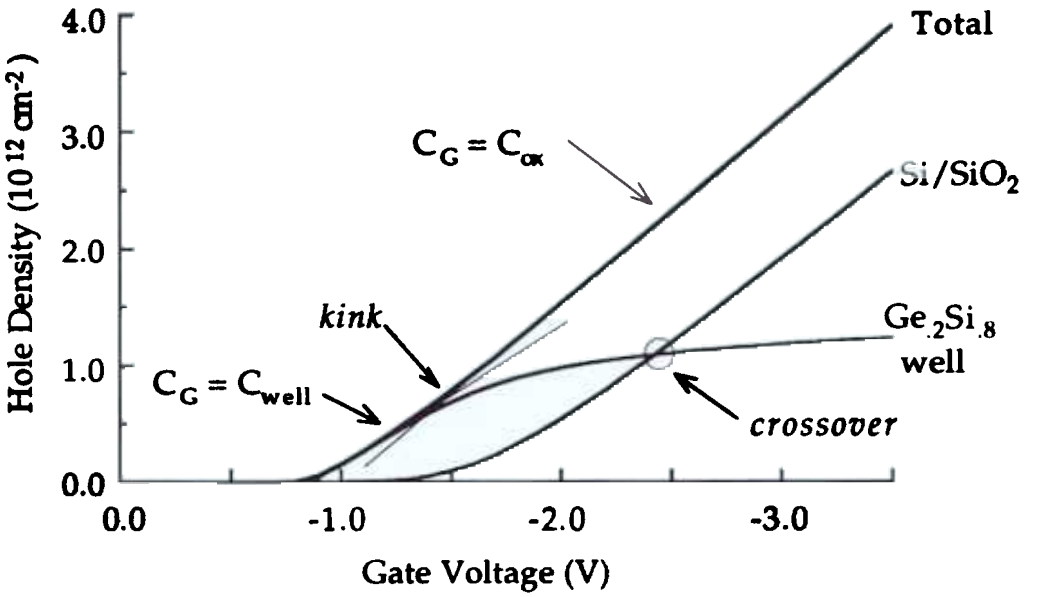


Figure 3. 7 Simulated hole density vs. gate voltage for a MOS-HHMT with a $Ge_{.2}Si_{.8}$ well and a 7.5 nm Silicon spacer layer (same as Fig.5a-c). The total hole density under the gate is shown as well as the hole density in the $Ge_{.2}Si_{.8}$ well and at the Si/SiO_2 interface.

the hole density in the Ge_xSi_{1-x} well and at the Si/SiO_2 interface as a function of the gate bias must be studied.

3.2.4.1 Description

The simulated hole density versus gate bias for a structure with a 7.5 nm silicon spacer and a $Ge_{.2}Si_{.8}$ well is shown in Figure 3.7 (band diagram in Figure 3.5) The total hole density under the gate is shown as well as the hole densities in the Ge_xSi_{1-x} well and at the Si/SiO_2 interface The Ge_xSi_{1-x} well inverts first, due to the valence band offset, so initially all the holes are being added to the Ge_xSi_{1-x} well The slope of the total hole density curve is proportional to the gate capacitance, which is equal to the series combination of the oxide and spacer layer capacitance when carriers are being added predominantly to the Ge_xSi_{1-x} well. When the Si/SiO_2 interface inverts most of the holes are subsequently added to the Si/SiO_2 interface so that the gate capacitance now equals the

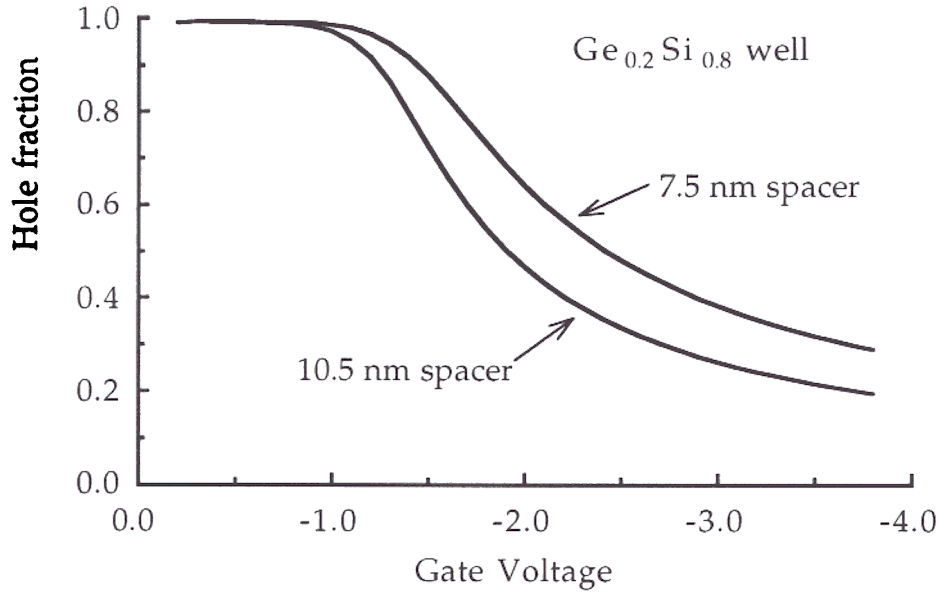


Figure 3. 8 Fraction of hole in the $Ge_{.2}Si_{.8}$ well of two MOS-HHMT structures (simulations)- one with a 7.5 nm silicon spacer layer and one with a 10.5 nm spacer layer.

and at the Si/SiO_2 interface to be equal when the tops of the bands at their respective heterojunctions are equidistant from the Fermi level (Figure 3.9). The voltage drop across the silicon spacer ($\Delta\phi_{sp}$) is then equal to the potential of the bandgap offset ($\Delta E_v/q$). If, for simplicity, it is assumed that the holes in the silicon spacer layer are all at the Si/SiO_2 interface and the ionized impurity density in the spacer is negligible (compared to Q_{inv} and Q_{dep}) then the electric field across the spacer(E_{SP}) will be constant and by Gauss' Law equal to:

Eq. 3.4

$$E_{sp} \quad \frac{Q_{Ge_xSi_{1-x}} + Q_{dep}}{\varepsilon_{Si}}$$

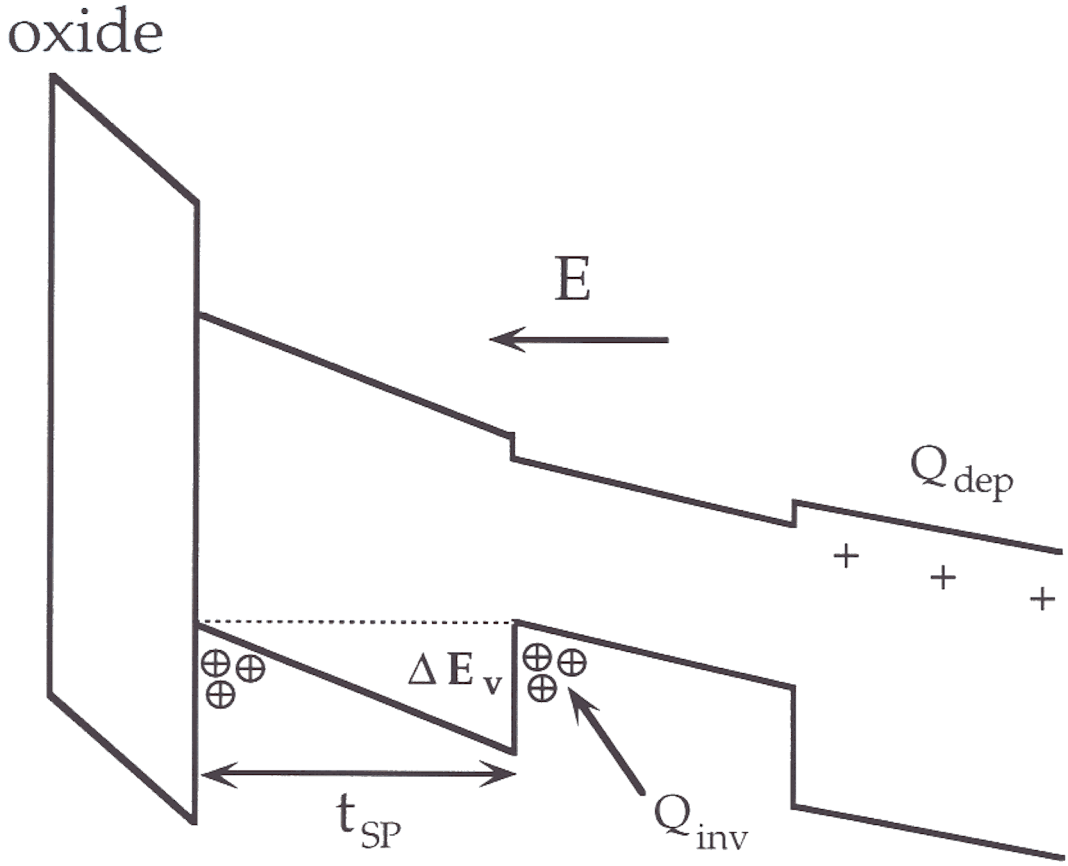


Figure 3. 9 : Schematic energy band diagram near crossover.

where Q_{dep} is the depletion charge, Q_{GeSi} is the charge from the holes in the Ge_xSi_{1-x} well, and ε_{Si} is the dielectric constant of silicon. The potential drop across the spacer layer can be written as $\Delta\phi_{sp} = E_{sp} \times t_{SP}$, so that at crossover

Eq. 3.5
$$\frac{\Delta E_V}{q} = \frac{Q_{Ge_xSi_{1-x}} + Q_{dep}}{\varepsilon_{Si}} \times t_{sp}$$

Eq. 3.5 may be rearranged to solve for the hole density in the Ge_xSi_{1-x} well at crossover as

Eq. 3.6
$$N_{Ge_xSi_{1-x}} \quad \frac{\varepsilon_{Si}\,\Delta E_V}{q^2 t_{sp}} \quad N_{dep}$$

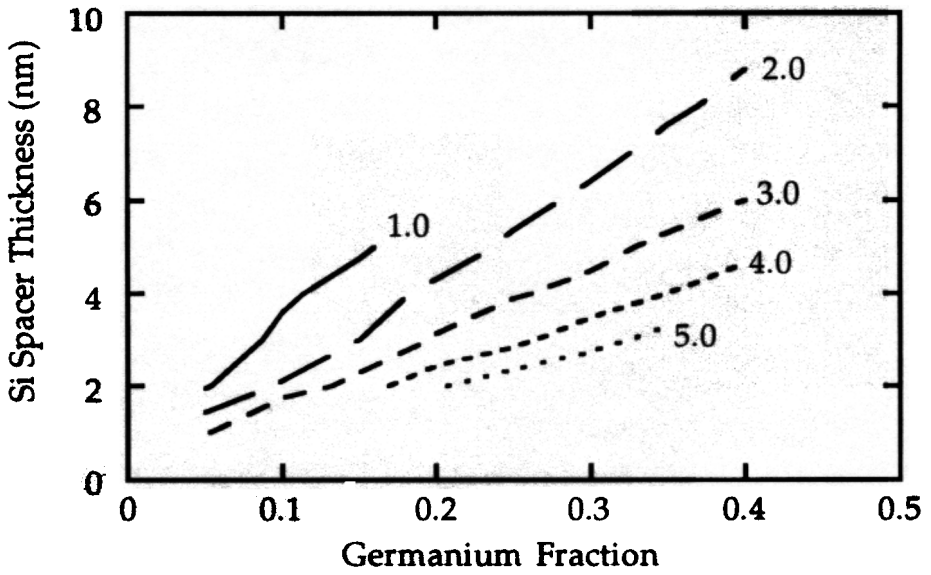


Figure 3.10 Lines of equal hole density in the GexSi1-x well at crossover ($\times 10^{12}$ cm^{-2}) from simulations of structures with various germanium fractions in the well and silicon spacer layer thicknesses.

Thus the density of holes in the Ge_xSi_{1-x} well (N_{GeSi}) will increase if the germanium fraction in the well is increased ($\Delta E_v \uparrow$), the spacer layer thickness is reduced ($t_{sp} \downarrow$), or the depletion charge is reduced ($Q_{dep} \downarrow$). In section 3.2.6 it is seen that the depletion density (N_{dep}) changes due to the bandgap offset at the Ge_xSi_{1-x} /Si interface. This effect is taken into account in calculations using Eq. 3.6.

The hole density at crossover was also extracted from simulations of structures with spacer layers thicknesses ranging from .0 to 7.0 nm, germanium fractions ranging from 0.1 to 0.4, a uniform doping of 1×10^{16} cm^{-3} and an oxide thickness of 10.0 nm. result of these simulations is shown in Figure 3.10 where the crossover density is plotted as a response of the spacer layer thickness and the germanium fraction in the well lines plotted are of equal hole density in the Ge_xSi_{1-x} well at crossover.

Spacer Width (nm)	Ge Fraction	Hole Density at Crossover (10^{12} cm^{-2})	
		analytic model	simulation
5.0	0.1	0.75	0.55
7.5	0.1	0.40	0.23
10.5	0.1	0.21	0.07
5.0	0.2	2.12	1.80
7.5	0.2	1.36	1.11
10.5	0.2	0.89	0.69
5.0	0.3	3.33	3.15
7.5	0.3	2.14	1.84
10.5	0.3	1.46	1.23

Table 3 Comparison of the hole density in the GexSi1-x well at crossover as predicted by the analytic model, Eq. 3.6 and to results from the Poisson solver for various spacer layer widths and germanium fractions in the Ge_xSi_{1-x} well.

In Table 3 the crossover density is calculated using Eq. 3.6 for a series of different spacer layer thicknesses and germanium fractions in the well and compared with results from the Poisson solver program (t_{ox} = 12.5 nm and $N_D = 1 \times 10^{16}$ cm^{-3} are fixed). The analytic model tends to overestimate the hole density in the Ge_xSi_{1-x} well at crossover This may result from the fact that the reduced density of states in the valence band of the Ge_xSi_{1-x} layer caused by strain splitting (section .3) is not taken into account in the analytic model

When designing the structures for a maximum hole density in the Ge_xSi_{1-x} well it is desirable to have a large germanium fraction and a small spacer thickness. In practice the germanium fraction will be limited by the critical thickness of the Ge_xSi_{1-x} layer (section .2.1). In chapter 4 the optimum spacer layer width (t_{SP}) is shown to be a tradeoff between increasing the hole density in the Ge_xSi_{1-x} well ($t_{SP} \downarrow$) and increasing the mobility

in the well (t_{SP} ↑). For device performance optimizing the transconductance of the MOS-HHMT is of primary importance and this depends on both the carrier density and the inversion mobility.

3.2.5 Threshold Shift

In simple MOSFET theory the transistor is considered to enter into strong inversion when the surface potential (ϕ_S) is twice the Fermi potential in the bulk ($2\phi_F$). For a non-degenerate semiconductor this is approximately :

Eq. 3.7
$$\phi_F \equiv \frac{(E_F - E_i)}{q} \approx \left(\frac{kT}{q}\right)\cdot \ln\left(\frac{N_D}{N_i}\right)$$

where E_F is the Fermi energy, E_i is the intrinsic Fermi energy, N_D is the donor impurity concentration and N_i is the intrinsic carrier concentration.

The threshold voltage (gate voltage at the onset of strong inversion) is then equal to:

Eq. 3.8
$$|V_t| = \frac{|Q_{dep}|}{C_{gate}} + 2\phi_F \approx \frac{\sqrt{2\,\varepsilon_{Si} N_D (2\phi_F)}}{C_{ox}} + 2\phi_F + \phi_{ms}$$

where Q_{dep}/C_G is the potential drop across the oxide resulting from the field caused by the depletion charge (Q_{dep}), $2\phi_F$ is the surface potential at inversion, and ϕ_{ms} is the difference in wirkfunction between the gate material and the semiconductor.

In a MOS-HHMT the Ge_xSi_{1-x} well inverts before the Si/SiO_2 interface so that threshold is determined by the potential at the Ge_xSi_{1-x}/Si interface (ϕ_G) rather than the surface potential (ϕ_S) The potential at the Ge_xSi_{1-x}/Si interface at the onset of strong inversion is lower than that of a typical MOSFET because of the valence band offset :

Eq. 3.9
$$\phi_G = \frac{(2\phi_F - \Delta E_V)}{q}$$

In a MOS-HHMT the gate capacitance when adding carriers to the Ge_xSi_{1-x} well is not equal to the gate oxide capacitance but rather the series capacitance of the silicon spacer layer and the gate oxide as discussed in section 3.2.2. Making these substitutions into Eq. 3.8 the threshold for inversion in the Ge_xSi_{1-x} well becomes:

$$\text{Eq. 3.10} \quad |V_t| \approx \frac{\sqrt{2q\varepsilon_{Si}N_D(2\phi_F - \Delta E_V)}}{\left(\frac{(C_{ox})(C_{SP})}{C_{ox}+C_{SP}}\right)} + (2\phi_F - \Delta E_V)$$

and the threshold shift of the MOS-HHMT relative to a silicon MOSFET with the same doping can be written as

$$\text{Eq.3} \quad \Delta V_t = \Delta E_v + \frac{\sqrt{2\varepsilon_{Si}qN_D(2\phi_F)}}{C_{ox}} \left(\frac{C_{ox}}{C_{Si}}+1\right) \times \left(1 - \frac{\Delta E_V}{2q\phi_F}\right)^{\frac{1}{2}}$$

The threshold shift, calculated using Eq. 3 1, is shown as a function of the germanium fraction for two spacer widths (5.0 and 5.0 nm) in Figure 3 1 The magnitude of the threshold shift is strongly dependent on the size of the valence band offset. Increasing the spacer layer thickness reduces the capacitive coupling of the gate to the Ge_xSi_{1-x} channel, partially offsetting the threshold shift caused by the valence band offset A positive threshold shift means that the magnitude of the threshold voltage is reduced (i.e. V_t approaches zero), while a negative threshold shift means that the Ge_xSi_{1-x} well does not invert before the Si/SiO_2 interface

In Table 3.2 the threshold shift calculated by this simple analytic model is compared to the threshold shift extracted from a linear fit of the hole concentration vs. gate voltage curve from simulations and from the drain