1. INTRODUCTION

1.1. Problem Statement

Silicon (Si) is by far the most dominantly used semiconductor material with over 97 % utilization in integrated microelectronics [1].

The speed limitation of electrical interconnect is to a first order proportional to the product of conductor resistance and capacitance to ground ([2], pg. 5). The persistent miniaturization trend in modern microelectronic manufacturing processes has the effect of increasing both the resistance of the thinner conductors and the capacitance of the more closely spaced conductors. This has the detrimental effect of decreasing the maximum Si CMOS intra-chip data transmission speed to about 3.5 GHz. The situation is worse for inter-chip data communication where the resistance and capacitance of printed circuit board (PCB) tracks and integrated circuit (IC) package pins are even higher. These fundamental speed limitations are becoming so problematic that other data and clock distribution means are desperately sought.

The practically proven high speed, space saving, noise immune, low power consumption, high capacity and economical advantages of fibre-optic data communication seem to suggest that optical clock and data distribution could also be employed in short-distance inter- and intra-chip applications.

While on-chip Si optical transmission [3], detection and manipulation elements are already practically achievable ([4] and [5]), a suitably efficient on-chip Si light source is not yet available. Although light emission from Si was observed as early as 1955 [6] and its high-speed capability ([7] and [8]) and long-term reliability ([9] - [11]) are established, the major reason for silicon's inherently weak light emission is that it is an indirect band-gap material ([12] and section 2.1.4). While alternative light-sources have been proposed ([4] and [13] - [21]), most require special manufacturing steps that are not easily implemented in the currently prevailing CMOS manufacturing processing technology.

Power efficient silicon-based light sources are also desirable in near-to eye (NTE) displays ([22] and [23]) and in bio-medical lab-on-chip applications [24].

The provision of CMOS-compatible light sources necessitates better understanding of light emission in Si ([25] - [37]) and resultantly improving the power efficiency of Si light sources to practically usable levels with, if possible, no or a minimum of additional manufacturing steps.

The Carl and Emily Fuchs Institute for Microelectronics (CEFIM) at the University of Pretoria (UP) has been developing Si light sources since 1992 ([38] - [46]). Si electroluminescence (EL) improvement research within the INSiAVA project at CEFIM focussed on promising light source configurations that include *inter alia* avalanche, punch-through (see section 2.1.9) and carrier-injection ([38], [39] and subsection 2.2.1.4) Si light sources.

1.2. Quantum Confinement

Enhancement factors of up to 30 in EL due to quantum mechanical confinement in ultra-thin single crystal Si compared to bulk devices were reported in [47] - [50].

References [47] and [48] reported a strong efficiency improvement in forward-biased SOI light-emitting diodes (LEDs) on a buried oxide (BOX) when the thickness of the regions shown in Figure 1.1 was reduced.

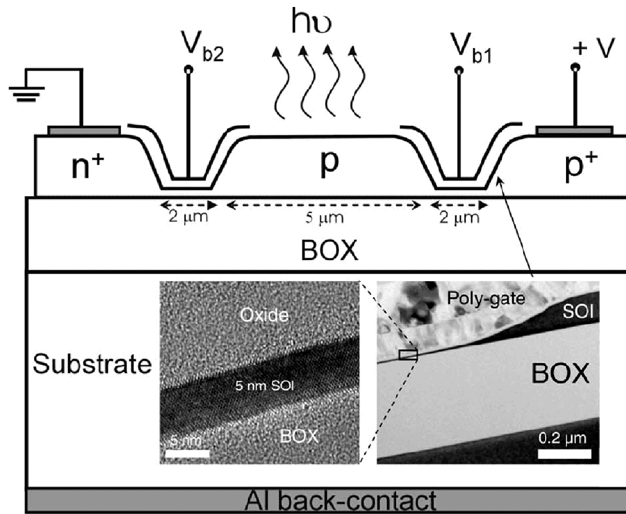


Figure 1.1 Cross-section of the SOI LED manufactured in [47] and [48].

The dramatic increase in integrated EL with reduced device layer thickness (Figure 1.2) is mostly attributed to the suppression of non-radiative recombination.

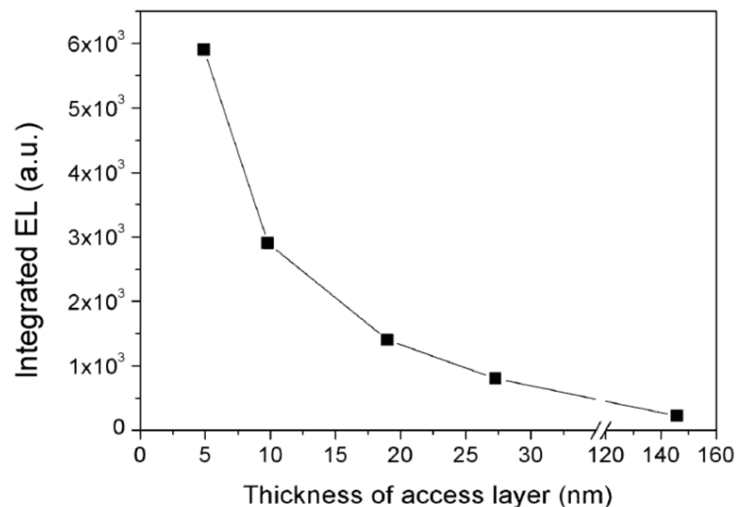


Figure 1.2 Integrated EL intensity against access layer thickness [48].

Abovementioned SOI light sources are only thin in one dimension (1D), their planar thickness, with device widths ranging between 20 and 60 µm. Their *pn*-junctions are also located outside the thinned area.

1.3. Objective

The main purpose of this work was to design and manufacture SOI light sources that would enable the investigation of quantum confinement effects in avalanche, punch-through and carrier-injection Si light sources on EL characteristics like external power efficiency and spectral emission.

Instead of just creating planar thin (1D quantum-confined) devices, the technical objective was to design and manufacture SOI light sources that are smaller in two dimensions, i.e. two-dimensional (2D) quantum-confined SOI light sources.

With reference to the definitions in Figure 1.3, the following finger junction dimensions were aimed at:

$$5\text{ nm} \le t \le 100\text{ nm}, \quad (1.1)$$

$$10\text{ nm} \le w \le 100\text{ nm and} \quad (1.2)$$

$$200\text{ nm} \le l \le 400\text{ nm}. \quad (1.3)$$

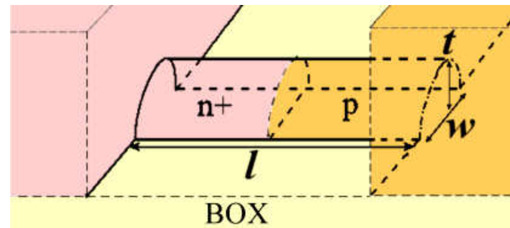


Figure 1.3 Generalized SOI finger junction dimension definitions.

Manufacturing larger rectangular Si structures and selectively oxidizing these created the desired thinner Si fingers.

Nanometre-scale Si wires had already been manufactured through oxidation ([51] - [52]), but *pn*-junctions had, to our knowledge, never been implemented inside such thin Si wires.

Without explaining their functional principles here, the three Si light source types that were selected for silicon-on-insulator (SOI) miniaturization in this research work are:

- Avalanche (n^+p or p^+n junctions),
- Punch-through (n^+pn^+ or p^+np^+ junctions) and
- Carrier-injection light sources.

1.4. Approach

Figure 1.4 displays the approach in terms of an activity flow diagram.

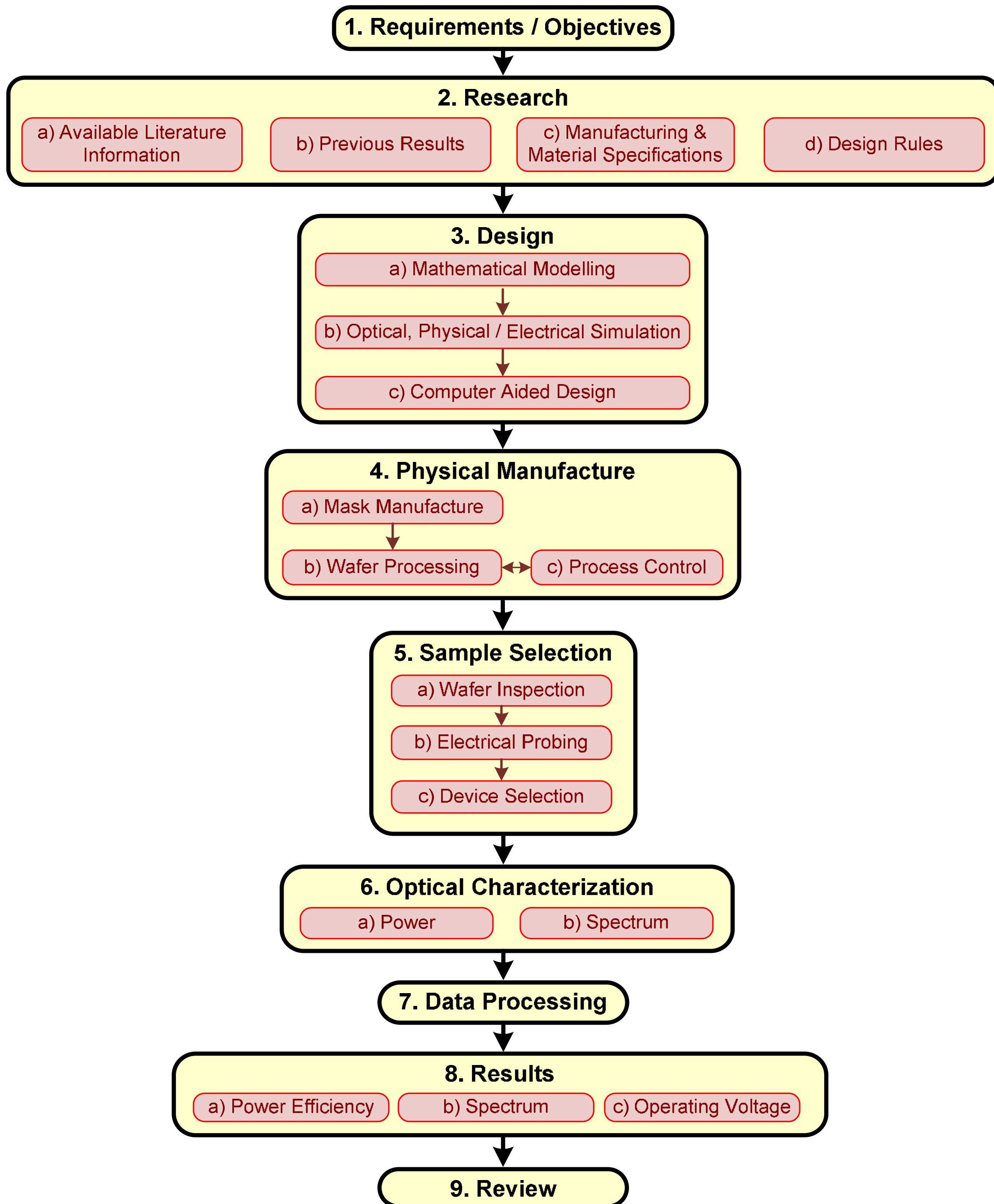


Figure 1.4. Activity flow diagram.

Although this work is mostly concerned with the design and manufacture of the nanometre-scale SOI light sources, some basic electrical and optical measurements were also conducted.

2. SILICON BACKGROUND

Before delving into the design and manufacture aspects of the nanometre-scale SOI light sources, some physical and optical properties of silicon that found application in this works are briefly highlighted in this chapter.

2.1. Semiconductor Properties

To estimate required physical parameters, like semiconductor acceptor and donor doping concentrations N_A and N_D, and resultant material and junction properties, like junction breakdown voltage V_{BD}, resistivity ρ ,sheet resistance R_S and depletion region width w_D, the relations in this section were used.

2.1.1. Resistivity

The relationship between a material's resistivity ρ, its sheet resistance R_S and thickness $\boldsymbol{t}$ can be stated as ([53], pg. 31)

$$\rho = R_S t. \tag{2.1}$$

Figure 2.1 illustrates the relationship between impurity doping concentration and resistivity of doped Si ([53], pg. 32).

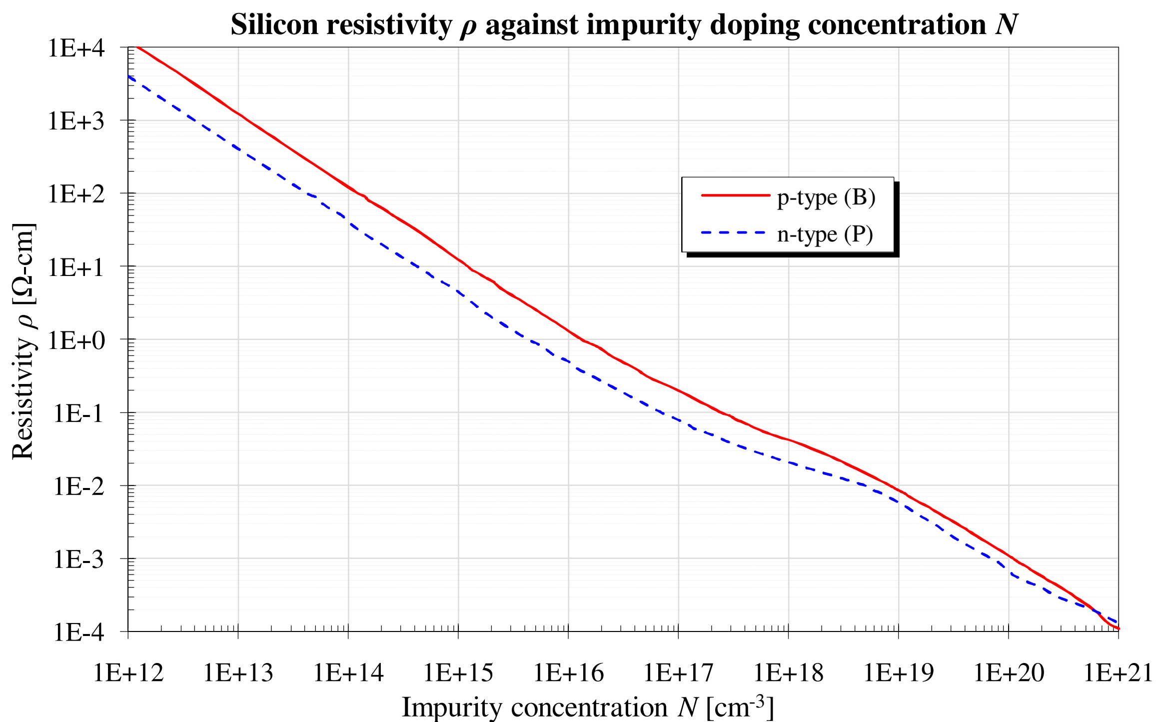


Figure 2.1. Silicon resistivity as a function of impurity doping concentration.

2.1.2. Intrinsic Carrier Concentration

Figure 2.2 shows how silicon's intrinsic carrier concentration n_i increases with impurity doping concentration above about 10^{16} cm^{-3} with respect to the weakly or undoped silicon intrinsic concentration $n_{i0} = 1.45 \cdot 10^{10}$ cm^{-3} at 300 K (physical constants listed in Addendum A).

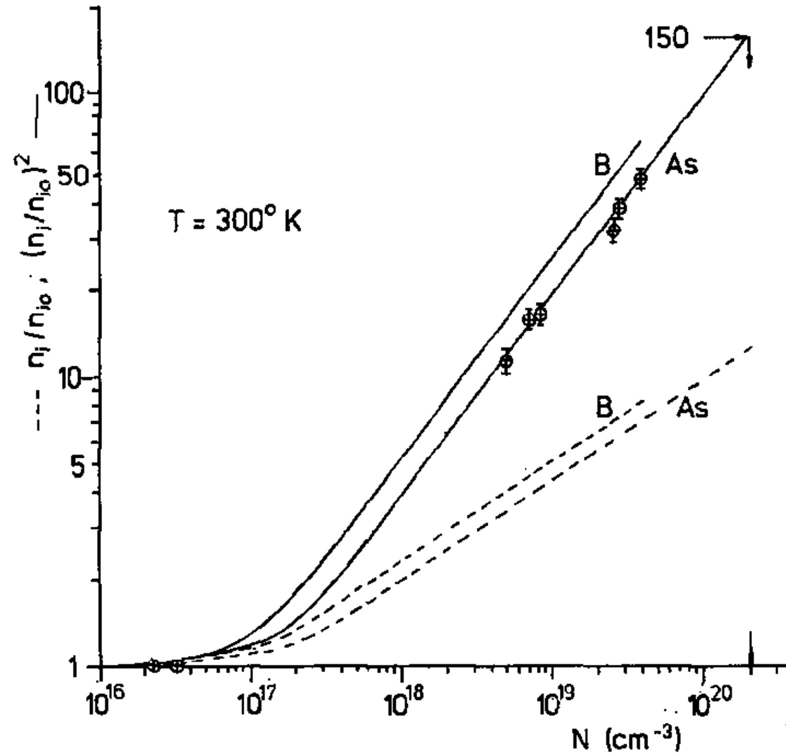


Figure 2.2. Silicon intrinsic concentration n_i variation with doping species and concentration [54].

An empirical expression modelling above band-gap narrowing in a simulation program is given in [55] pg. 347, but assuming that estimated ratios from above plot are sufficient shows that n_i in a n^+ material with Arsenic (As) concentration of 10^{19} cm^{-3} is about $4.1n_{i0} = 5.9 \cdot 10^{10}$ cm^{-3} and $2.3n_{i0} = 3.3 \cdot 10^{10}$ cm^{-3} for a Boron (B) concentration of 10^{18} cm^{-3}.

2.1.3. Junction Built-in Potential

The *pn*-junction built-in potential ψ_{BI} is defined as ([53], pg. 81)

$$\psi_{BI} \approx \frac{k_B T}{q} \ln\left(\frac{N_A N_D}{n_i^2}\right), \tag{2.2}$$

where k_B is Boltzmann's constant, T is the absolute temperature, q is the electron charge, n_i is silicon's intrinsic carrier concentration, N_A and N_D are the acceptor and donor impurity doping concentrations in the *p* and *n*-type regions respectively.

For $N_A = 10^{19}$ cm^{-3} and $N_D = 10^{17}$ cm^{-3}, $\psi_{BI} = 875$ mV.

2.1.4. Energy Band Structure

Figure 2.3 displays the energy band structure of an infinitely large ultra-pure perfect Si crystal at 300 K against the three basic wave-vector directions (k = Λ, Γ and Σ) from the centre (Γ) of silicon's first Brillouin zone ([53], pg. 11 – 13).

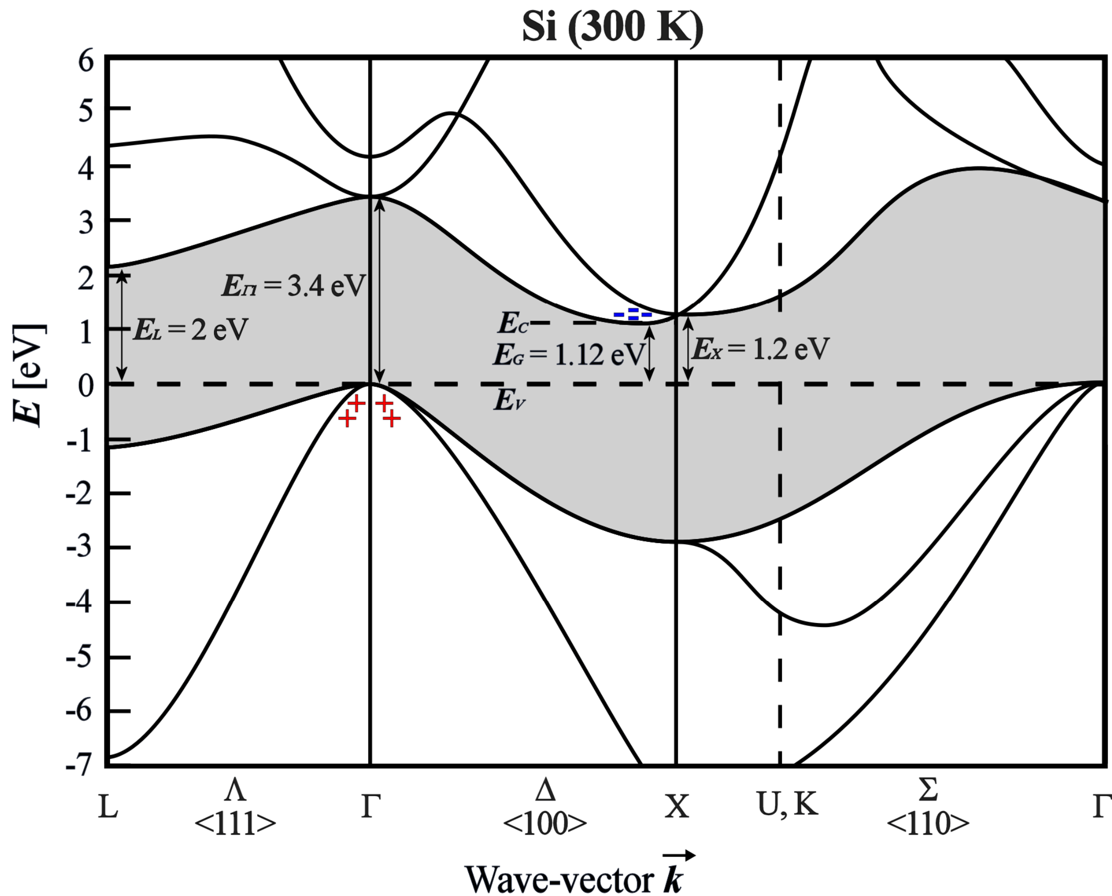


Figure 2.3. Energy band structure of Si at 300 K.

The bands below and above the gray "forbidden" zone are the valence and conduction bands respectively. Only the lowest-energy bands directly bordering on the band-gap are depicted. Although there are many more bands above and below the energy gap, these have almost no effect on the electronic properties of the material.

While electrons and holes can occupy any energy and position in the conduction and valence bands, they cannot remain in the gray "forbidden" zone for any appreciable time. The "-" and "+" symbols denote the locations of electrons and holes in the lowest-energy conduction and valence bands respectively.

The valence band is very similar for most crystalline materials and usually has a maximum at location $k = \Gamma$ (= 0), the centre of the Brillouin zone ([56], pg. 3). The maximum valence band energy is indicated with E_V (and is in this case set to $E = 0$ eV). The lowest conduction band energy occurs close to $k = X$ and is identified by E_C.

The energy band-gap E_G is the energy difference between the valence and conduction bands. $E_G = 1.12$ eV for pure Si at 300 K. This is the minimum energy associated with an electron-hole generation or recombination across the forbidden energy gap. Since silicon's lowest conduction-band energy valley and the highest valence peak do not coincide at the same wave-vector position within the Si crystal, Si, in contrast to GaAs, is an indirect band-gap material.

The energy band diagram in Figure 2.3 is only valid for an infinitely large, pure and perfect Si crystal at 300 K. Reducing the dimensions of the crystal, varying the temperature and the presence of impurities and crystal defects will alter the position and shape of the energy bands. All of these factors can therefore change the energy band-gap. This work attempts to increase the light emission efficiency of a Si light emitter by reducing the device dimensions, which should increase the band-gap energy and the wave-vector difference associated with it ([15], pg. 6, [16], pg. 629, [20], pg. 32, [21], pg. 29, [27], pg. 1044, [48], pg. 85 and [57], pg. 2).

2.1.5. Impact Ionization

In a sufficiently high electric field, a carrier can gain energy higher than the band-gap and excite additional electron-hole pairs through impact ionization. This carrier multiplication process is characterized by the ionization rate α that is defined by the number of electron-hole pairs generated by a primary electron (hole) carrier with velocity v_n (v_p) per unit distance travelled ([53], pg. 37)

$$\alpha_n = \frac{1}{nv_n}\frac{\mathrm{d}n}{\mathrm{d}t} \text{ and} \tag{2.3}$$

$$\alpha_p = \frac{1}{nv_p}\frac{\mathrm{d}p}{\mathrm{d}t}, \tag{2.4}$$

which lets the carrier density or current variation with distance x be expressed as

$$\frac{\mathrm{d}J_n}{\mathrm{d}x} = -\frac{\mathrm{d}J_p}{\mathrm{d}x} = \alpha_n J_n + \alpha_p J_p, \tag{2.5}$$

where the total current $J = J_n + J_p$ is constant at any position x.

The electron and hole ionization rates are highly dependent on the electric field and can be expressed in the form [53]

$$\alpha(\mathscr{E}) = \frac{q\,\mathscr{E}}{E_I}\,\mathrm{e}^{\frac{-\mathscr{E}_I}{\mathscr{E}\left(1+\frac{\mathscr{E}}{\mathscr{E}_P}\right)+\mathscr{E}_T}}, \tag{2.6}$$

where E_I is the high-field effective ionization threshold energy and is typically 3.6 eV for electrons and 5.0 eV for holes for Si. $\mathscr{E}_T$, $\mathscr{E}_P$ and $\mathscr{E}_I$ are thermal, phonon-optical and ionization threshold fields respectively.

Figure 2.4 shows that the ionization rate of electrons is always higher than that of holes.

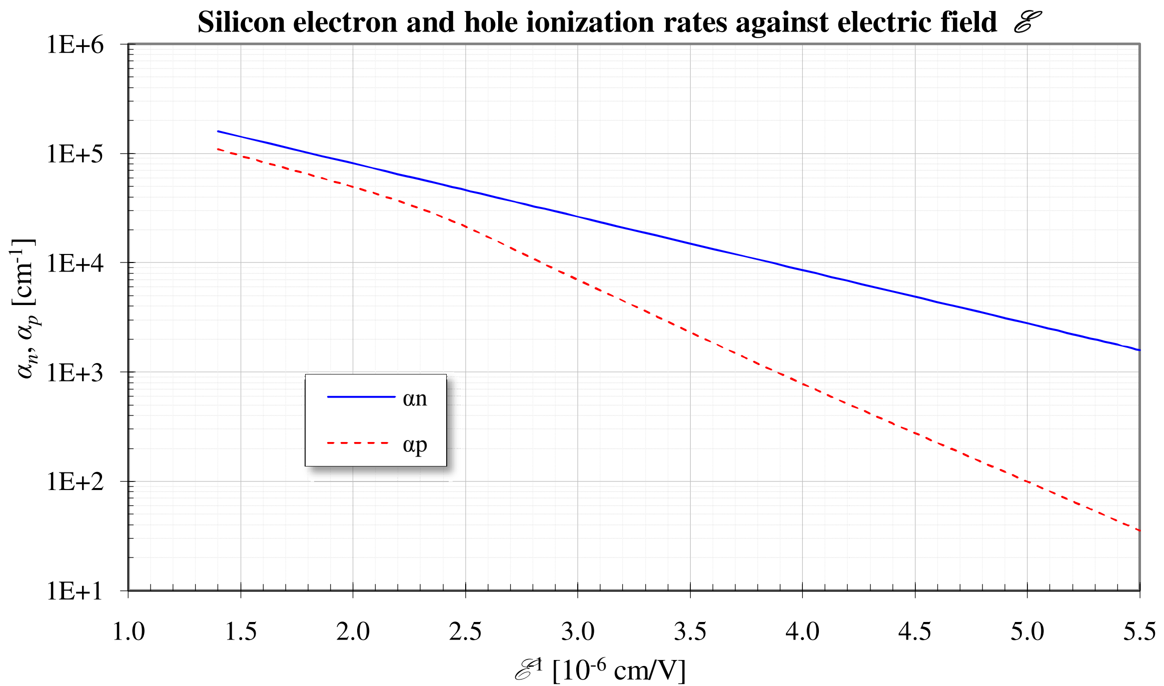


Figure 2.4. Silicon electron and hole ionization rates against electric field.

2.1.6. Critical Electric Field

As Figure 2.5 shows, the critical electric field $\mathscr{E}_C$ of an abrupt Si junction at breakdown can be expressed in terms of background doping concentration N [cm^{-3}] by ([53], pg. 107)

$$\mathscr{E}_C = \frac{4 \cdot 10^5 \text{ V/cm}}{1 - \frac{1}{3}\log_{10}\left(\frac{N}{10^{16}\text{ cm}^{-3}}\right)}. \tag{2.7}$$

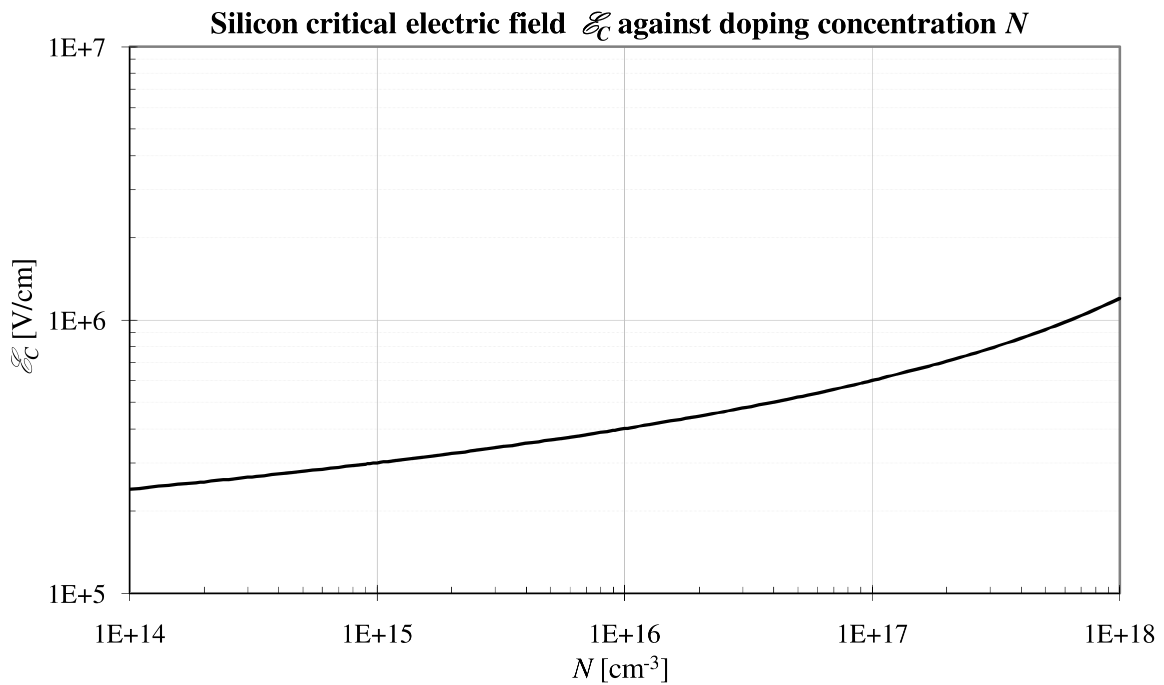


Figure 2.5. Si critical electric field $\mathscr{E}_C$ against background doping N.

For a background doping $N = 10^{17}$ cm^{-3} the critical electric field $\mathscr{E}_C \approx 6{\cdot}10^5$ V/cm.

2.1.7. Avalanche Breakdown

The planar avalanche breakdown voltage of an abrupt junction is expressible in terms of background doping concentration N and silicon's critical electric field $\mathscr{E}_C$ by

$$V_{BD} \cong \frac{\varepsilon_{Si}\mathscr{E}_C^{\,2}}{2qN} - \psi_{BI}, \qquad (2.8)$$

where silicon's dielectric constant $\varepsilon_{Si} = \varepsilon_0\varepsilon_{rSi}$ (Addendum A).

Above equation is plotted in Figure 2.6, but also indicates that (2.8) is only valid for impurity concentrations to the left of the stippled line. For higher concentrations, quantum mechanical tunnelling (Zener breakdown) will contribute and later dominate the breakdown mechanism.

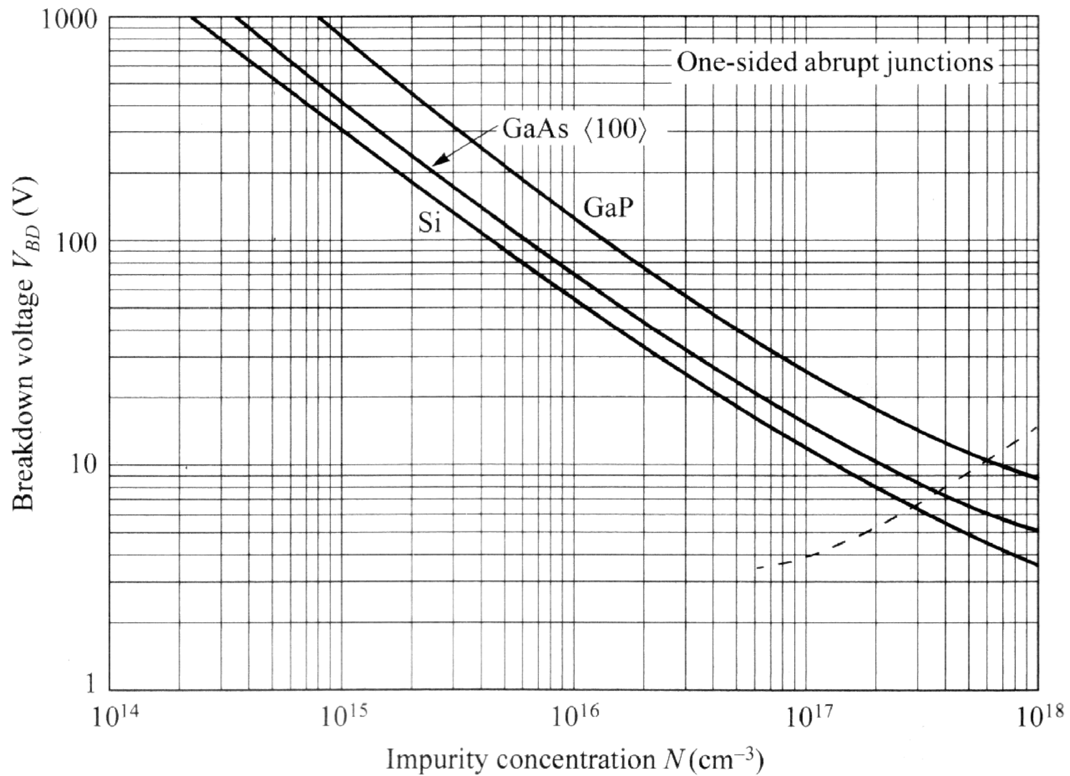


Figure 2.6. Planar breakdown voltage V_{BD} against impurity doping concentration N ([53], pg. 108).

Since tunnelling is not yet dominant at $N = 10^{17}$ cm^{-3}, the planar breakdown voltage of a *pn*-junction with $N_A = 10^{19}$ cm^{-3} and $N_D = 10^{17}$ cm^{-3} is approximately $V_{BD} \approx 11$ V.

2.1.8. Depletion Region Width

The depletion region width w_d of a *pn*-junction can be expressed as ([53], pg. 83)

$$w_d = \sqrt{\frac{2\varepsilon_{Si}}{q}\left(\frac{N_A + N_D}{N_A N_D}\right)\left(\psi_{BI} - 2\frac{k_B T}{q} - V_D\right)}, \tag{2.9}$$

where V_D is the voltage applied across the junction and is positive for forward bias.

For a junction with N_A = 10^{19} cm^{-3} and N_D = 10^{17} cm^{-3} biased at V_D = 0 V the thermal equilibrium depletion region width w_D = 104 nm.

As Figure 2.7 confirms, the same junction at breakdown (V_D = -V_{BD} = -11 V) has a depletion region width w_D = 389 nm.

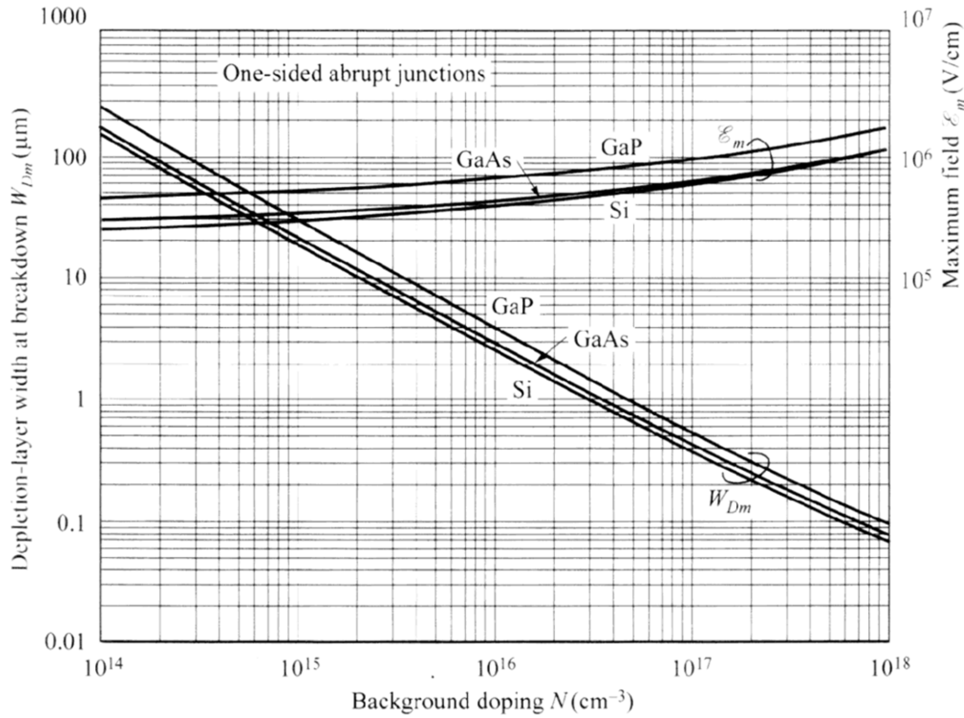


Figure 2.7. Breakdown depletion width and critical field against background doping [53].

The depletion region extensions x_n and x_p into the *n*- and *p*-doped sides of a junction respectively depend on the impurity doping concentrations on both sides as ([58], pg. 41)

$$x_n = w_d \frac{N_A}{N_A + N_D} \text{ and} \tag{2.10}$$

$$x_p = w_d \frac{N_D}{N_A + N_D}. \tag{2.11}$$

For $N_A = 10^{19}$ cm^{-3} and $N_D = 10^{17}$ cm^{-3} $x_n = 103$ nm and $x_p = 1.03$ nm at $V_D = 0$ V and $x_n = 385$ nm and $x_p = 3.85$ nm at breakdown ($V_D = -V_{BD}$). As expected from the fact that $N_A = 100N_D$, $x_n = 100x_p$.

2.1.9. Punch-through

Figure 2.8 shows two heavily doped p^+ end regions with equal acceptor concentration N_A spaced a distance W_P apart by a lower doped n-type drift region with donor concentration N_D.

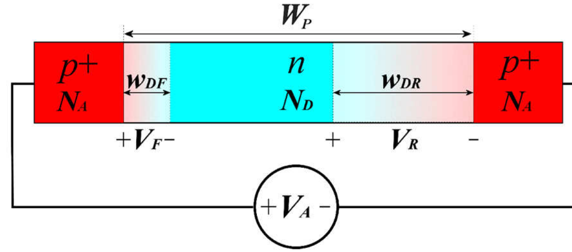


Figure 2.8. Punch-through voltage variable definitions.

Applying a voltage V_A across the two back-to-back diodes has the effect of forward biasing the left junction and reverse biasing the right junction. From equation (2.9) the depletion region widths of both junctions can be expressed as

$$w_{DF} = \sqrt{\frac{2\varepsilon_{Si}}{q}\left(\frac{N_A + N_D}{N_A N_D}\right)\left(\psi_{BI} - 2\frac{k_B T}{q} - V_F\right)} \text{ and} \tag{2.12}$$

$$w_{DR} = \sqrt{\frac{2\varepsilon_{Si}}{q}\left(\frac{N_A + N_D}{N_A N_D}\right)\left(\psi_{BI} - 2\frac{k_B T}{q} + V_R\right)}, \tag{2.13}$$

where the sum of the voltages across the forward and reverse biased junction, V_F and V_R respectively, equals the applied voltage V_A.

Under the condition that the reverse biased junction does not experience avalanche breakdown, the applied voltage V_A can be increased to the point where the expanding reverse biased depletion region reaches the forward biased depletion region. The applied voltage V_A at which $w_{DF} + w_{DR} = W_P$ is defined as the punch-through voltage

$$V_{PT} = \frac{qW_P^2}{2\varepsilon_{Si}}\left(\frac{N_A N_D}{N_A + N_D}\right) - W_P\sqrt{\frac{2q}{\varepsilon_{Si}}\left(\frac{N_A N_D}{N_A + N_D}\right)\left(\psi_{BI} - 2\frac{k_B T}{q} - V_F\right)} \tag{2.14}$$

$$\approx \frac{qW_P^2}{2\varepsilon_{Si}}\left(\frac{N_A N_D}{N_A + N_D}\right) - W_P\sqrt{\frac{2q}{\varepsilon_{Si}}\left(\frac{N_A N_D}{N_A + N_D}\right)\left(\psi_{BI} - 2\frac{k_B T}{q}\right)}. \tag{2.15}$$

Approximation (2.15) is not very accurate though as substitution for a n^+pn^+ punch-through structure with N_A = 10^{19} cm^{-3}, N_D = 10^{17} cm^{-3}, W_P = 400 nm and assumed V_F = 0.7 V shows: equation (2.15) renders a punch-through voltage $V_{PT} \approx 6$ V, which is almost half the punch-through voltage $V_{PT} \approx 10$ V predicted by equation (2.14).

Figure 2.9 shows the energy band diagrams of the p^+np^+ punch-through device for different biasing voltages: a) no applied bias (V_A = 0V), b) 0 V < V_A < V_{PT} and c) V_A = V_{PT}.

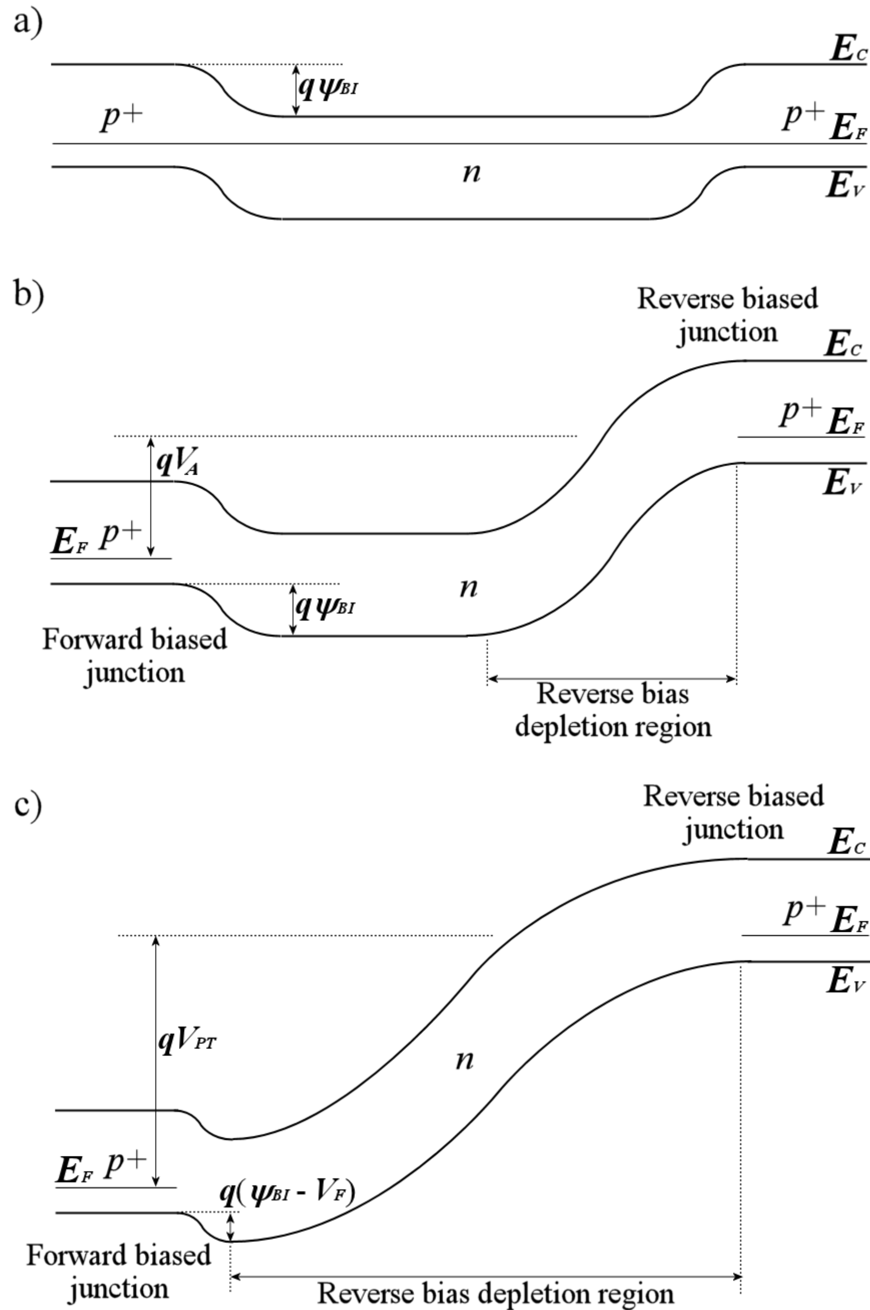


Figure 2.9. Energy-bands of the p^+np^+ junctions at a) V_A = 0 V, b) 0 V < V_A < V_{PT} and c) V_A = V_{PT}.

It is evident that when the reverse-bias depletion region reaches the forward-biased junction at $V_A = V_{PT}$, that it lowers the energy barrier at the forward-biased junction, which then injects a larger thermionic emission hole current.

Figure 2.10 shows the electron (n_{xy}) and hole (p_{xy}) carrier distributions in the p^+np^+ device for a) 0V < V_A < V_{PT} and b) $V_A = V_{PT}$.

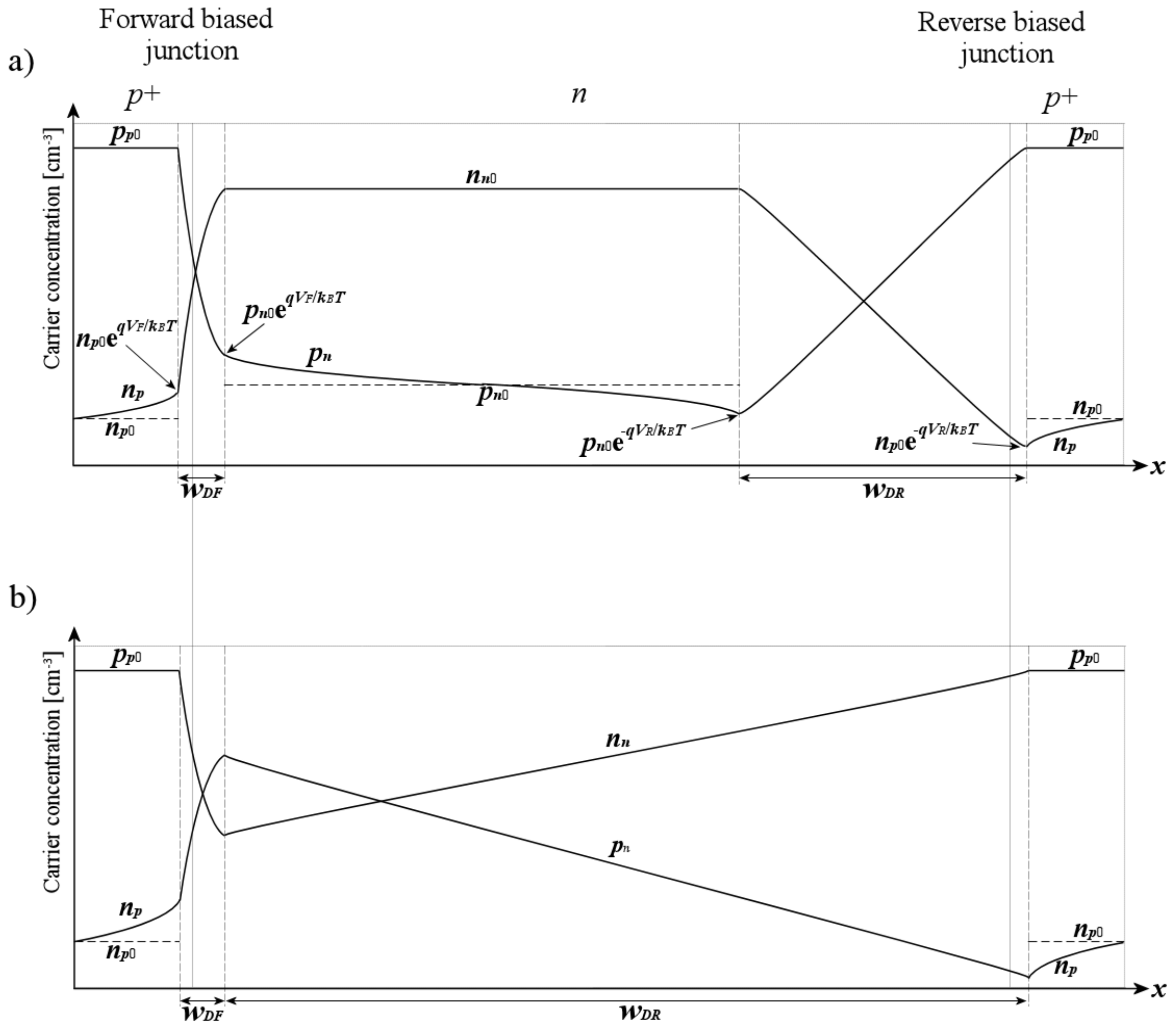


Figure 2.10. Back-to-back p^+np^+ junction carrier distribution for a) 0 V < V_A < V_{PT} and b) $V_A = V_{PT}$.

The barrier-lowering effect of the reverse-bias depletion region punching-through to the forward-biased junction has the effect of increasing the hole injection (p_n) into the intermediate *n*-type drift region. This makes more minority carriers in the reverse-bias depletion region available for avalanche multiplication and radiative recombination.

Since the radiative recombination rate is proportional to the *pn*-product {see equation (2.20)} a higher minority carrier concentration in the drift region will increase the probability of radiative recombination. The wider spread of high carrier concentrations in the intermediate *n*-region in Figure 2.10 b) also means the *pn*-product is larger for a wider region within the reverse biased depletion region, which also increases the probability of radiative recombination.

The fact that higher avalanche currents are achievable at lower terminal voltages also implies that the power efficiency of a punch-through light source is higher than the power efficiency of a single avalanching *pn*-junction.

Figure 2.11 shows the electric field distribution of the back-to-back diodes of Figure 2.8 at punch-through.

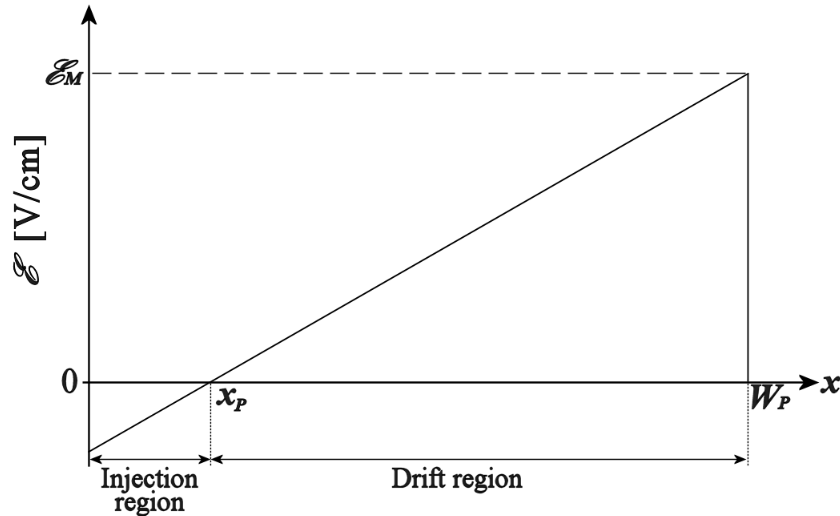


Figure 2.11. Electric field distribution in the punch-through device.

The linear electric field $\mathscr{E}$ is zero at $x = x_P$ (the point where the reverse-biased depletion region touches the forward-bias depletion region) and achieves its maximum field strength $\mathscr{E}_M$ at $x = W_P$ at the reverse biased junction. Since Figure 2.4 showed that the carrier ionization rates increase with electric field strength, it is expected that radiative recombination will predominantly occur in the depletion region close to the reverse biased junction at W_P.

This section illustrated the punch-through effect with a p^+np^+ device, but the higher impact ionization constant of electrons compared to holes (section 2.1.5) suggests that an n^+pn^+ punch-through light source would have a higher optical power efficiency than an equivalent p^+np^+ device.

2.1.10. Reach-through

While the punch-through device can increase radiative recombination by injecting additional minority carriers into the reverse-bias depletion region, its electric field decreases linearly from its maximum $\mathscr{E} = \mathscr{E}_M$ at $x = W_P$ to zero at x_P (Figure 2.11). Replacing the left p^+ diffusion of the forward-biased punch-through device junction in Figure 2.8 with an n^+ creates a reach-through device that does not feature carrier injection, but its electric field strength changes more gradually in the lower-doped drift region.

Figure 2.12 depicts that the trapezoidal electric field in the W_R–wide n^- drift region between the n^+ and p^+ end regions in a reach-through device is close to the maximum electric field $\mathscr{E}_M$ existing at the reverse biased junction and drops off sharply in the heavily doped n^+ end region([59], pg. 101).

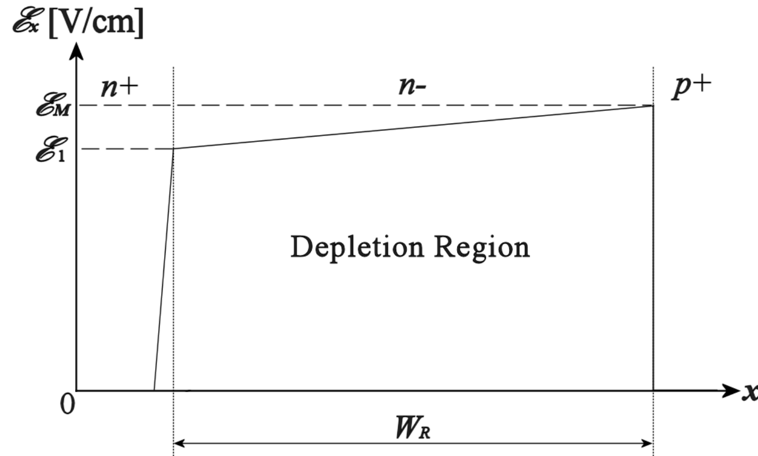


Figure 2.12. Electric field distribution in a reach-through device.

The high electric field throughout the entire drift region causes a larger number of high-energy ("hot") carriers, a higher carrier ionization rates, more carrier multiplication, higher probability of radiative recombination and resultantly a higher Si light source efficiency. The high electric field throughout the depletion region also causes the carriers to travel at their saturated velocity, which gives the device a fast switching speed compared to conventional *pn*-junctions in avalanche.

The electric field E_1 at the interface between the n^- drift and n^+ end region depends on the maximum electric field $\mathscr{E}_M$ and n^- drift region doping concentration N_{DR} through

$$\mathscr{E}_1 = \mathscr{E}_M - \frac{qN_{DR}}{\varepsilon_{Si}} W_R . \tag{2.16}$$

Neglecting the small voltage drop in the n^+ end region allows stating the voltage across the reach-through device as

$$V_{RT} \cong \left(\frac{\mathscr{E}_M + \mathscr{E}_1}{2} \right) W_R . \tag{2.17}$$

Avalanche breakdown will occur when the maximum electric field $\mathscr{E}_M$ is equal to the critical electric field $\mathscr{E}_C$.

Combining (2.16) and (2.17) and substituting $\mathscr{E}_M = \mathscr{E}_C$ renders the reach-through breakdown voltage as

$$V_{BD_RT} \cong \mathscr{E}_C W_R - \frac{qN_{DR} W_P{}^2}{2\varepsilon_{Si}} . \tag{2.18}$$

With $N_{DR} = 10^{17}$ cm^{-3} and $W_R = 400$ nm $V_{BD_RT} \approx 12$ V.

2.2. Optical Properties

This section briefly explores some important optical properties of Si and its associated processing materials and steps that were used in this work. This includes defining integrated light source efficiency (subsection 2.2.1.2), presenting the average power spectrum of typical CMOS light sources (subsection 2.2.1.3), quantifying refractive indices and extinction coefficients of Si, silicon dioxide (SiO_2) and silicon nitride (Si_xN_y or Si_3N_4) in subsection 2.2.2, photon absorption within Si (subsection 2.2.3.1) and optical transmission through the SiO_2 and Si_xN_y layers covering ICs (subsection 2.2.3.2),.

2.2.1. Electroluminescence

2.2.1.1. Radiative Recombination

Avalanche electroluminescence in semiconductors is mostly caused by radiative recombination of mobile electrons and holes. Two types of radiative recombination can occur: direct (band-to-band) radiative recombination and radiative band-to-impurity recombination ([60], pg. 161).

Figure 2.3 on pg. 8 showed that silicon, in contrast to GaAs, has its minimum conduction band energy not at the same waver vector as its maximum valence energy at Γ, which makes it an indirect band-gap semiconductor. Since electron-hole radiative recombination requires that momentum be conserved, this means that a phonon with equal but opposite momentum to the electron's initial state in the conduction. Such a two-step process has a lower occurrence probability compared to direct recombination, which makes Si an inefficient light emitter ([56], pg. 3).

Since the photon energy $\hbar\omega$ emitted during radiative recombination is equal or slightly larger than the recombination energy gap E_G, the recombination path(s) can be estimated from the wavelength (spectrum) of the generated light through

$$E_G \cong \hbar\omega. \tag{2.19}$$

Since both electrons and holes are necessary for band-to-band recombination, the radiative recombination rate R_{BTB} is proportional to the np-product

$$R_{BTB} \propto np - n_i^2. \tag{2.20}$$

Similarly, the band-to-impurity radiative recombination rate R_{BTI} is proportional to the product of the electron-hole pair concentration N_P and the concentration of impurities N_I involved in the radiative recombination process through

$$R_{BTI} \propto N_P N_I . \quad (2.21)$$

2.2.1.2. Integrated Light Source Efficiency

Two types of integrated light source efficiencies can be considered: the quantum conversion efficiency and the external power efficiency.

2.2.1.2.1 Quantum Conversion Efficiency

The quantum-conversion efficiency η of a light source is defined by the ratio

$$\eta = \frac{\text{average number of photons emitted per unit time}}{\text{average number of electrons passing through source per unit time}} . \quad (2.22)$$

Generally, two types of quantum efficiency definitions are used. While the internal quantum-conversion efficiency η_i relates the emitted photons to electrons right at the light generation site, the external quantum efficiency η_e uses the externally available photon count that incorporates optical losses (for example surface reflection and media absorption) associated with the environment between source and region where the light is used.

Although the quantum efficiency of a source is a valid performance criterion parameter, it is not very practical in physical implementations.

2.2.1.2.2 External Power Efficiency

The external power efficiency *EPE* of an integrated light source is defined as the ratio of the externally available optical power $P_{Optical_Out}$ to the electric power $P_{Electrical_In}$ invested to generate the light

$$EPE = \frac{P_{Optical_Out}}{P_{Electrical_In}} . \quad (2.23)$$

Since the external power efficiency relates how many Watt of electrical power are necessary to generate a specific Wattage of optical power it is much more useful comparison figure in practical applications, where a certain optical power is desired for a minimum of electrical power needed to generate the light.

2.2.1.3. CMOS Light Source

Figure 2.13 illustrates that the shallow n^+ in p-well and p^+ in n-well junctions in the standard CMOS process mainly constitute horizontal diodes; i.e. most of the pn-interface is oriented in the horizontal plane, with most of the depletion region (indicated by dotted lines) extending into the underlying lighter doped material under applied reverse bias.

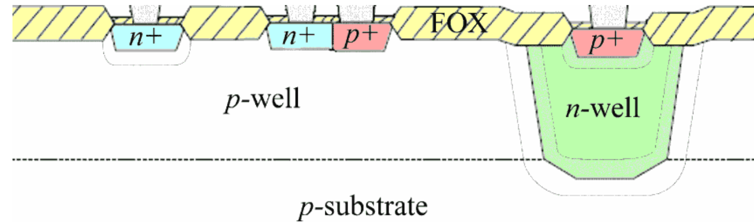


Figure 2.13. Cross-sectional view of the three typical CMOS pn-junction diode types.

The typical CMOS light source employs either the n^+ in p-well or p^+ in n-well junctions in avalanche. Depending on the electric field distribution, the light generation site is usually located along one sidewall of either junction.

Figure 2.14 depicts average measured optical power spectra of 0.35 µm CMOS light sources operating at a current of 650 µA when the light source is placed directly in front of a spectrometer's fibre-optical receptor and the spectrometer's response (see Addendum B) is taken into account.

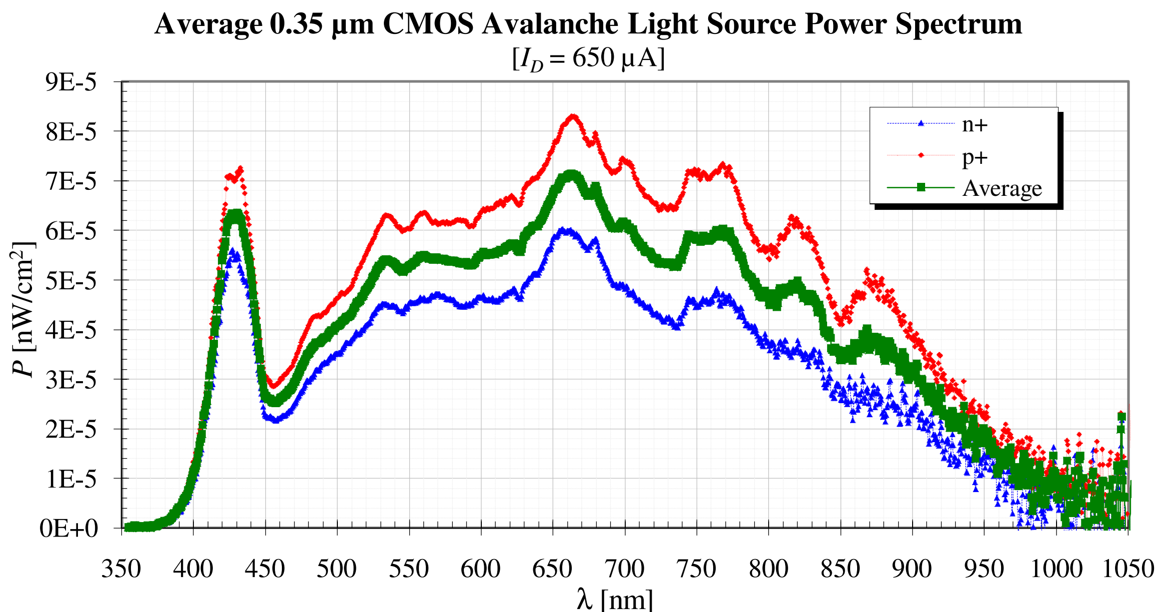


Figure 2.14. Average 0.35 µm CMOS avalanche light source power spectrum (I_D = 650 µA).

The silicon light source spectrum in above figure indicates that the silicon light source has its maximum optical power emission around 660 nm, which is in the visible region.

It should be kept in mind that above power spectral density plot still incorporates absorption from the bulk silicon depth from which the light originates and transmission losses from the silicon dioxide and nitride layers covering the chip, both of which introduce attenuation of the shorter wavelengths. Furthermore, the optical acceptance angle of the spectrometer only captures a small solid angle of the complete optical radiation pattern.

The light emission power spectrum of a thin Si wire junction is expected to not only be much stronger, but should also be much more pronounced for shorter wavelengths.

2.2.1.4. Carrier Injection

du Plessis *et al* ([38] and [39]) found that the quantum conversion efficiency of Si electroluminescence can be improved by injecting “cool” (low energy) carriers from a forward-biased junction into the high electric field existing in the depletion region of a reverse-biased junction so that they can recombine with “hot” (high energy) carriers. In contrast to a simple *pn*-junction in which the avalanche-multiplied minority carriers are due to the leakage current thermally generated in the drift region, the carrier injection technique improves the optical power efficiency of an integrated light source by making more carriers available to avalanche multiplication at the same reverse bias.

Such a Si light source is achievable by placing a forward-biased injecting junction in close proximity of an avalanching reverse-biased junction. The resultant Si light source can have two or three terminals. In the three-terminal configuration, the third connection to the lower-doped drift region that contains the reverse-bias depletion region allows separate manipulation of the forward- and reverse-biased junction currents. In the two-terminal configuration, the intermediate drift region remains floating and the forward and reverse biased junctions are forced to have the same current. Bulk CMOS injection light sources have the disadvantage that it is quite difficult to direct all of the injected carriers into the avalanching depletion region. The forward-biased junction injection can be illustrated with an analogy to an overflowing dam with equal dam wall height all around it: the water would spill equally in all directions. This unidirectional carrier injection into the bulk of the forward-biased junction has the effect that only a relatively small fraction of the “cool” carriers reaches the depletion region – the rest is lost in the bulk, but because it still contributes to the forward current, it decreases the power efficiency of the light source.

One way of increasing the electroluminescent efficiency is to ensure that all injected carriers enter the depletion region by implementing the two junctions next to each other in a Si wire surrounded by an insulator. The SOI manufacturing technology is a perfect candidate for realizing such devices as it provides a thin Si layer on top of a thick SiO_2 layer. Etching the superfluous Si away on adjacent sides of the junctions would leave a Si channel that constricts the carrier movement in it. The overflowing water analogy would entail having the dam surrounded by mountains except for one spill-way channel that the water is forced to flow through.

Another way of ensuring that most of the injected carriers reach the depletion region is to employ the punch-through effect introduced in section 2.1.9 (pg. 14). As the reverse-bias depletion region reaches the forward biased junction, it only decreases the barrier energy at that location so that most of the cool carriers enter directly into the encroaching depletion region. Depending on the degree of barrier lowering, some carriers can still be thermally injected over the higher barrier at other locations. The water analogy would employ a dam with a depression in its otherwise equal-height dam wall surrounding it. Most of the water would flow out at this lower barrier, although, depending on the depth of the localized dam wall depression, the wind could still spill some water over the higher dam wall at other places. By reverse biasing the injector-junction slightly with respect to the bulk, the thermally excited component can be suppressed so that only the localized punch-through region is forward biased and injecting cool carriers into the touching reverse-bias depletion region.

An example of an injection-enhanced Si light source would be the punch-through device. Manufacturing this device in an SOI technology with a drift region doping level selected in such a way that the reverse-biased depletion region touches the forward biased junction (enabling carrier injection) while at the same time exhibiting avalanche breakdown should render a Si light source with improved quantum conversion efficiency.

2.2.2. Refractive Indices and Extinction Coefficients

2.2.2.1. Silicon

Figure 2.15 plots silicon's refractive index and extinction coefficient against wavelength.

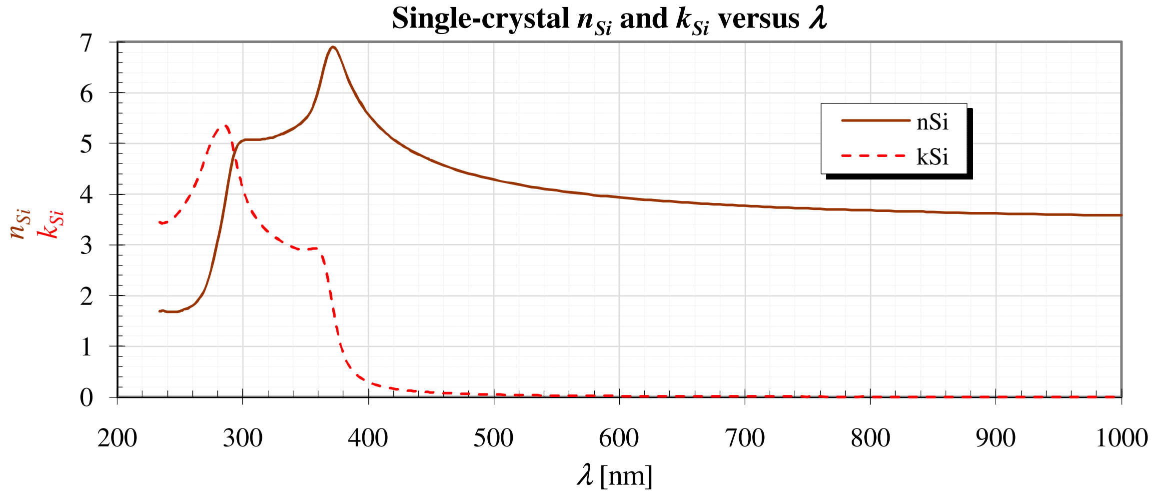


Figure 2.15. Single-crystal silicon refractive index n_{Si} and extinction coefficient k_{Si} ([61] - [63]).

A material's extinction coefficient k is relatable to the material's optical absorption coefficient α (see subsection 2.2.3.1) through

$$k = \frac{\alpha\lambda}{4\pi}. \tag{2.24}$$

Due to the strong gradient variation of silicon's refractive index and extinction coefficient between 286 nm and 372 nm, curve-fitting is easier when both are plotted against the photon energy E_{Ph} ($\approx$ 1.24 [eV/nm]/λ) as shown in Figure 2.16.

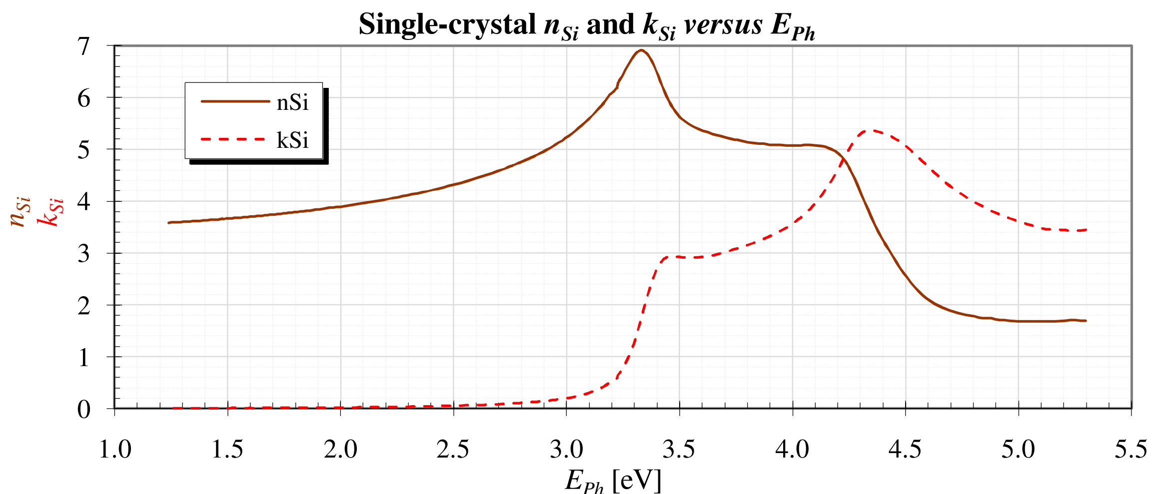


Figure 2.16. Silicon refractive index n_{Si} and extinction coefficient k_{Si} against photon energy E_{Ph}.

Jellison *et al* [62] propose that the silicon index of refraction n_{Si} below the direct band edge, $E_{Ph} < 3.33$ eV (372 nm < λ), can be represented by

$$n_{Si}(E_{Ph},T) = \sqrt{4.386 - 0.00343T + \frac{99.14 + 0.062T}{E_G{}^2 - E_{Ph}{}^2}}, \tag{2.25}$$

where T is the temperature in ºC and the fitted energy gap $E_G = 3.652$ eV.

For the region above the direct band edge, 3.33 eV < E_{Ph} < 4.33 eV (286 nm ≤ λ < 372 nm), Figure 2.17 shows a polynomial fit for n_{Si}.

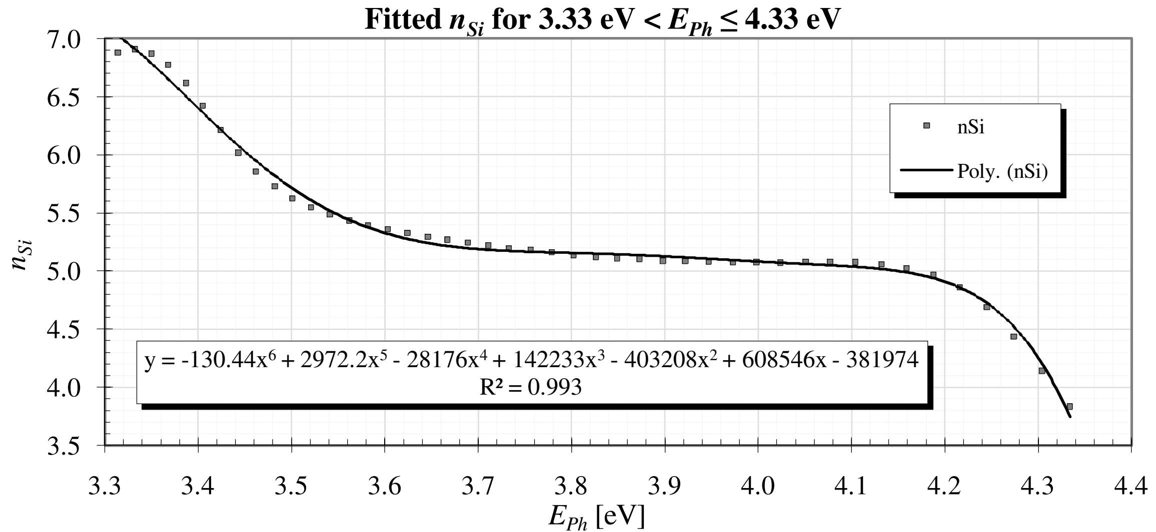


Figure 2.17. Fitted silicon refractive index for 286 nm ≤ λ ≤ 372 nm.

Due to large gradient variations, the best silicon extinction coefficient curve-fit is achieved in three overlapping parts over the range 1.24 eV ≤ E_{Ph} < 4.33 eV (286 nm ≤ λ ≤ 1 µm)

Figure 2.18 shows a k_{Si} polynomial fit for 1.24 eV ≤ E_{Ph} < 3.11 eV (398 nm ≤ λ ≤ 1 µm).

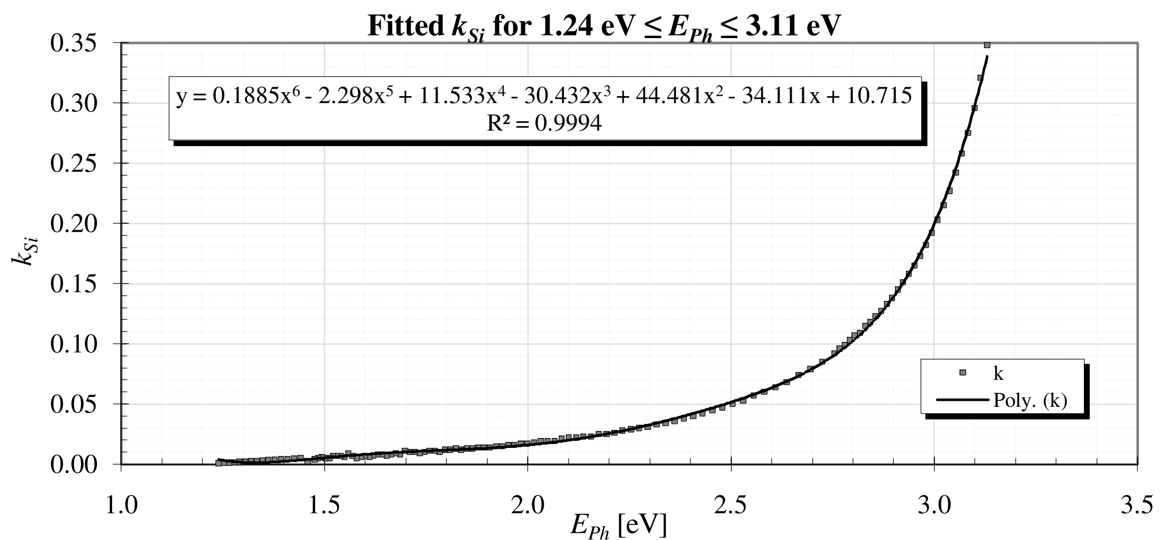


Figure 2.18. Silicon extinction coefficient k_{Si} polynomial fit for 1.24 eV ≤ E_{Ph} < 3.11 eV.

Figure 2.19 shows a k_{Si} polynomial curve fitting for 3.1 eV ≤ E_{Ph} < 3.42 eV (362 nm ≤ λ ≤ 400 nm)

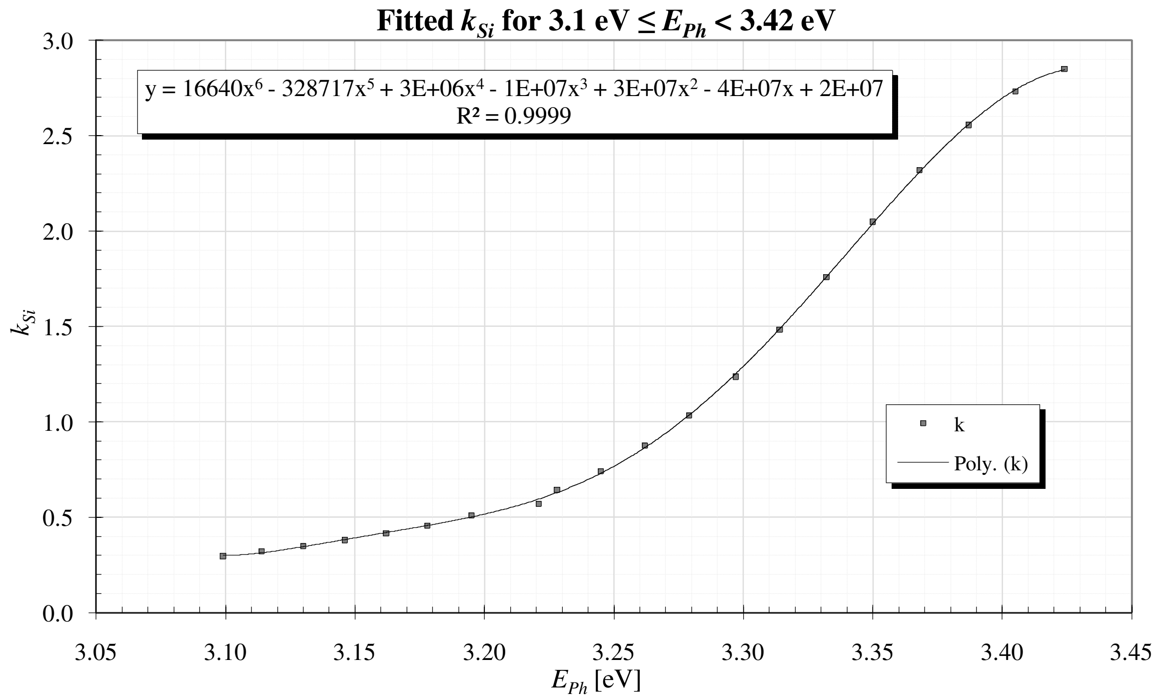


Figure 2.19. Silicon extinction coefficient k_{Si} polynomial fit for 3.1 eV ≤ E_{Ph} < 3.42 eV.

Figure 2.20 shows a k_{Si} polynomial curve fit for 3.39 eV ≤ E_{Ph} ≤ 4.33 eV (286 nm ≤ λ ≤ 366 nm).

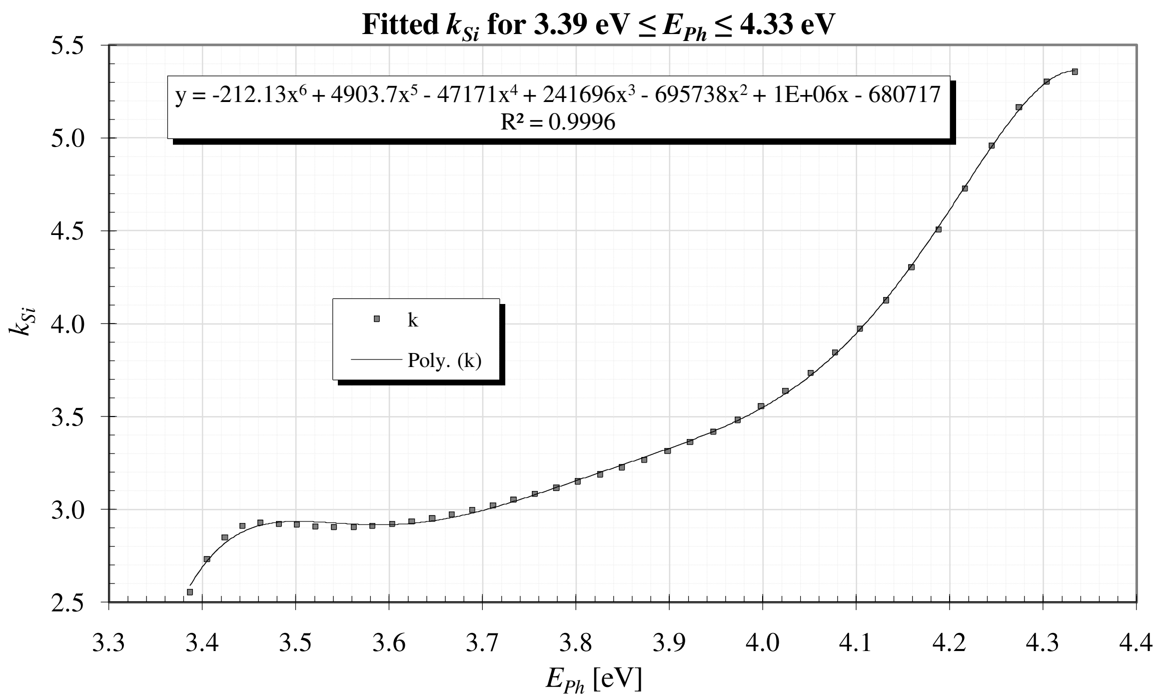


Figure 2.20. Silicon extinction coefficient k_{Si} polynomial fit for 3.39 eV ≤ E_{Ph} ≤ 4.33 eV.

2.2.2.2. Silicon Dioxide and Silicon Nitride

The refractive indices of chemical vapour deposited (CVD) SiO_2 and Si_xN_y are usually assumed fixed as $n_{SiO2} \approx 1.46$ and $n_{SixNy} \approx 1.99$ respectively although material properties like the polarizability of impurity ions [64] and the manufacturing process can, for example, vary n_{SiO2} between 1.43 and 1.51 and n_{SixNy} between 1.98 and 2.

The absorption coefficients of SiO_2 and Si_xN_y are very close to zero and usually only become significant in the deep-UV spectrum.

2.2.3. Optical Loss Mechanisms

2.2.3.1. Silicon Photon Absorption

Light travelling in silicon is absorbed as photons annihilate themselves by generating hole-electron pairs that diffuse from their generation site.

The wavelength-dependant average photon travelling distance $x_{Absotption}$ in which a material absorbs 63 % of the light is related to the material's absorption coefficient α through

$$\overline{x_{Absorption}} = \frac{1}{\alpha}. \tag{2.26}$$

Figure 2.21 illustrates the wavelength dependence of the optical absorption coefficient α and the average photon travelling distance $x_{Absorption}$ in silicon ([53], pg. 53, [65] and [66]).

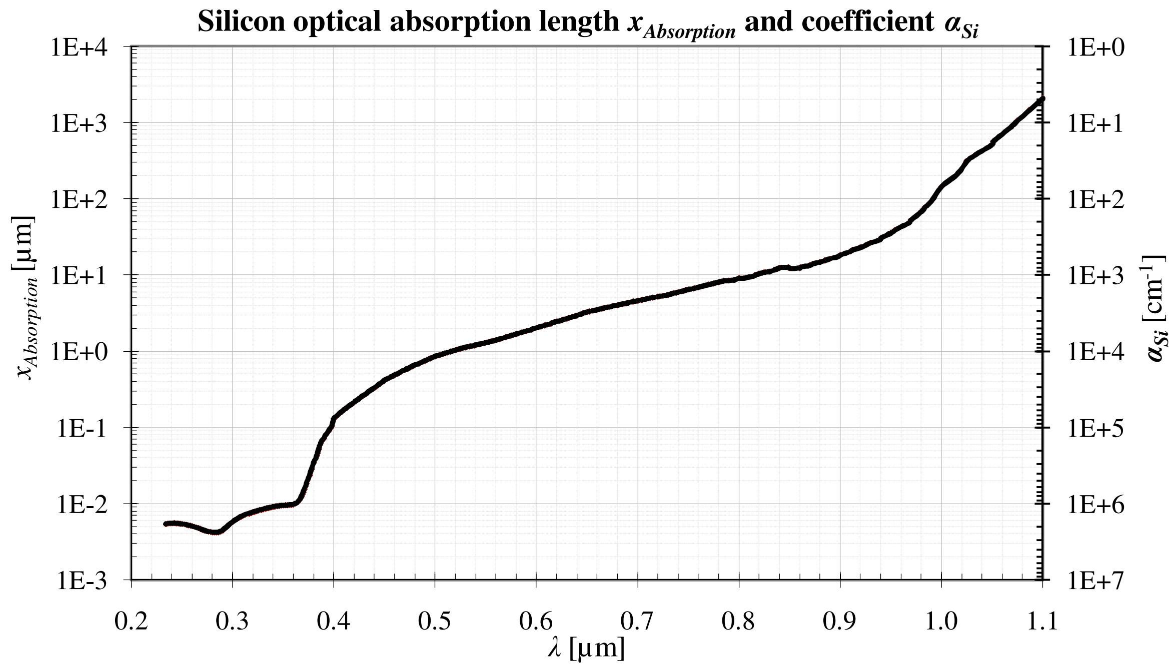


Figure 2.21. Silicon optical absorption distance $x_{Absorption}$ and coefficient α_{Si} against wavelength λ.

Figure 2.21 illustrates how 63 % of light with λ = 350 µm is absorbed within 10 nm of silicon, but light with λ = 1 µm can travel more than 100 µm through silicon.

According to [65] for 400 nm ≤ λ ≤ 850 nm silicon's optical absorption coefficient α_{Si} can be numerically approximated in terms of wavelength λ (in µm) by

$$\alpha_{Si} \cong 10^{13.2131-36.7985\lambda+48.1893\lambda^2-22.5562\lambda^3} \text{ cm}^{-1}. \tag{2.27}$$

From (2.26) the photon flux at a distance x from a light source in a material is expressible as ([65], pg. 22)

$$\Phi_x = \Phi_0 \alpha e^{\alpha x}, \tag{2.28}$$

where Φ_0 is the photon flux at the source ($x = 0$).

The optical power P_{Out} leaving the material's surface with total surface reflection coefficient R can be related the surface photon-flux Φ_S by

$$P_{Out} = \frac{\Phi_S E_{ph}}{(1-R)}, \tag{2.29}$$

where the photon energy E_{ph} depends on Planck's constant h, speed of light c (Addendum A) and wavelength λ by

$$\begin{aligned} E_{ph} &= h\frac{c}{\lambda} \\ &\approx \frac{1.24\,[\text{eV}\cdot\mu\text{m}]}{\lambda}. \end{aligned} \tag{2.30}$$

Since the absorption coefficient α, photon-flux Φ_0 and reflection coefficient R in equations (2.28) and (2.29) usually vary with wavelength, different light wavelengths will have different absorption lengths within a material.

The discussion in this subsection makes it evident that one way of increasing the externally available optical power from a silicon light source is to minimize the distance that the light has to travel through silicon. The possibility of decreasing the photon absorption within the silicon material between light generation site and surface is another incentive to manufacture nanometre-scale silicon light sources.

2.2.3.2. Optical Transmission through Si, SiO_2 and Si_xN_y Stacks

In addition to the light absorbed in the silicon around the light generation site, the different refractive indices of the different materials in the optical path will cause reflections and refractions that introduce a wavelength-dependant light attenuation.

Figure 2.22 illustrates how incident light *I* generated in the Si is reflected and transmitted by the SiO_2 and Si_xN_y layers covering the silicon surface [67].

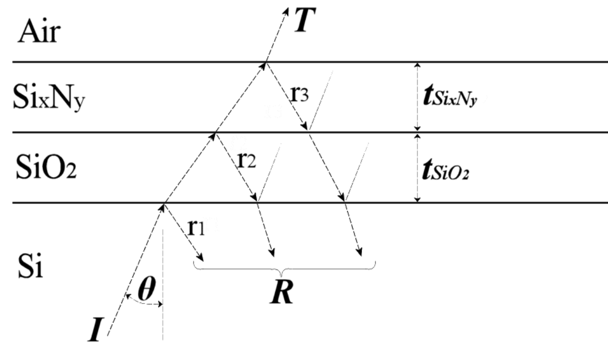


Figure 2.22. Light transmission through the Si, SiO_2, Si_xN_y and air layer stack.

In above figure the incident light (*I*) enters with angle θ to the normal of the Si-SiO_2 surface, *T* is the total light transmitted through to the air above the chip and *R* is the total reflection consisting of the reflection components r_1, r_2 and r_3 at the Si-SiO_2, SiO_2-Si_xN_y and Si_xN_y-air interfaces with thicknesses t_{SiO2} and t_{SixNy} respectively. As indicated, the light experiences multiple forward and backward reflections between the layer interfaces, which cause constructive and destructive interference and standing waves in a layer when the wavelength of the light is a multiple of the layer thickness.

Considering only perpendicular incidence (θ = 0) of light polarized in the plane of the Si surface, simplifies the electromagnetic wave analysis to the bidirectional field components illustrated in Figure 2.23.

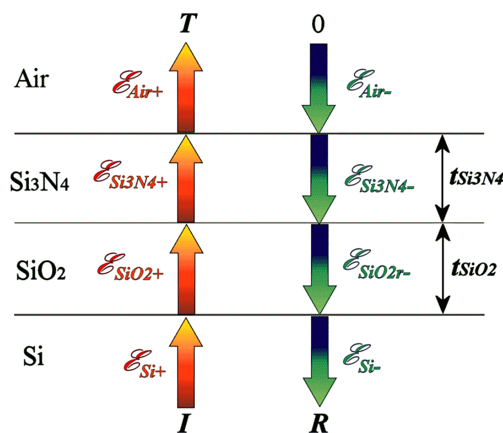


Figure 2.23. Electromagnetic fields in the Si-SiO_2-Si_xN_y-air interfaces.

In previous figure $\mathscr{E}_{x+}$ and $\mathscr{E}_{x-}$ denote the forward and reverse electromagnetic field components in material *x* respectively. The usage of these two distinct wave direction components is only to ease the analysis of a single complex electromagnetic field that is actually present.

Since the tangential electromagnetic field components at each dielectric interface must be continuous [68], the field strengths either side of an interface between materials *x* and *y* must be interrelated by

$$\mathscr{E}_{x+} - \mathscr{E}_{x-} = \mathscr{E}_{y+} - \mathscr{E}_{y-}. \tag{2.31}$$

Furthermore the refractive indices n_x and n_y of the materials either side of the interface interrelate the forward transmission and reverse reflection components by [69]

$$\begin{bmatrix} 1 & 1 \\ n_x & -n_x \end{bmatrix} \begin{bmatrix} \mathscr{E}_{x+} \\ \mathscr{E}_{x-} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ n_y & -n_y \end{bmatrix} \begin{bmatrix} \mathscr{E}_{y+} \\ \mathscr{E}_{y-} \end{bmatrix}. \tag{2.32}$$

Rewriting previous relation to express the field components in material *x* in terms of the field components existing in material *y* renders

$$\begin{aligned} \begin{bmatrix} \mathscr{E}_{x+} \\ \mathscr{E}_{x-} \end{bmatrix} &= \begin{bmatrix} 1 & 1 \\ n_x & -n_x \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 \\ n_y & -n_y \end{bmatrix} \begin{bmatrix} \mathscr{E}_{y+} \\ \mathscr{E}_{y-} \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} 1+\dfrac{n_y}{n_x} & 1-\dfrac{n_y}{n_x} \\ 1-\dfrac{n_y}{n_x} & 1+\dfrac{n_y}{n_x} \end{bmatrix} \begin{bmatrix} \mathscr{E}_{y+} \\ \mathscr{E}_{y-} \end{bmatrix}, \end{aligned} \tag{2.33}$$

which shows that $\mathscr{E}_{x+} = \mathscr{E}_{y+}$ is completely independent of $\mathscr{E}_{x-} = \mathscr{E}_{y-}$ when $n_y/n_x = 1$ and that the forward and reverse components "feed" into each other for $n_x \neq n_y$.

The relationship between the electromagnetic field strengths at opposite ends of material *x* depends on the layer's thickness t_x, the material's refractive index n_x and wavelength λ in the diagonal phase matrix

$$\Phi_x = \begin{bmatrix} \cos\phi_x + i\sin\phi_x & 0 \\ 0 & \cos\phi_x - i\sin\phi_x \end{bmatrix}, \tag{2.34}$$

where $\phi_x = 2\pi n_x t_x/\lambda$ and $i = \sqrt{-1}$. Above matrix models the constructive and destructive interference when the light's wavelength λ is a multiple of the layer thickness t_x.

Using the matrices in (2.32) and (2.34) the electromagnetic fields $\mathscr{E}_{Air}$ and $\mathscr{E}_{Si}$ of Figure 2.23 are relatable by

$$\begin{bmatrix} \mathscr{E}_{Air+} \\ \mathscr{E}_{Air-} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ n_{Air} & -n_{Air} \end{bmatrix}^{-1} \begin{bmatrix} \cos\phi_{Si_xN_y} & i\dfrac{\sin\phi_{Si_xN_y}}{n_{Si_xN_y}} \\ n_{Si_xN_y}\, i\sin\phi_{Si_xN_y} & \cos\phi_{Si_xN_y} \end{bmatrix} \begin{bmatrix} \cos\phi_{SiO_2} & i\dfrac{\sin\phi_{SiO_2}}{n_{SiO_2}} \\ n_{SiO_2}\, i\sin\phi_{SiO_2} & \cos\phi_{SiO_2} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ n_{Si} & -n_{Si} \end{bmatrix} \begin{bmatrix} \mathscr{E}_{Si+} \\ \mathscr{E}_{Si-} \end{bmatrix}. \tag{2.35}$$

The total power reflection R_P and transmission T_P of the stacked layers are calculated by

$$R_P = \left| \frac{\mathscr{E}_{Si-}}{\mathscr{E}_{Si+}} \right|^2 \text{ and} \tag{2.36}$$

$$T_P = \left| \frac{\mathscr{E}_{Air+}}{\mathscr{E}_{Si+}} \right|^2 . \tag{2.37}$$

Rewriting equation (2.35) in a simpler form as

$$\begin{bmatrix} \mathscr{E}_{Air+} \\ \mathscr{E}_{Air-} \end{bmatrix} = \begin{bmatrix} U & V \\ W & X \end{bmatrix} \begin{bmatrix} \mathscr{E}_{Si+} \\ \mathscr{E}_{Si-} \end{bmatrix}$$
$$\begin{bmatrix} T \\ 0 \end{bmatrix} = \begin{bmatrix} U & V \\ W & X \end{bmatrix} \begin{bmatrix} I \\ R \end{bmatrix}, \tag{2.38}$$

where the fact that no light is incident from the air ($E_{Air-} = 0$) can be used to express the total reflected power as

$$R_P = \left| \frac{-W}{X} \right|^2 . \tag{2.39}$$

Using (2.35) to obtain W and X, the power transmission T_P through all layers becomes

$$\begin{aligned} T_P &= 1 - R_P \\ &= 1 - \left| \frac{W}{X} \right|^2 \\ &= 1 - \frac{(n_{Si}a - n_{Air}d)^2 + (n_{Air}n_{Si}b - n_{Air}c)^2}{(n_{Si}a + n_{Air}d)^2 + (n_{Air}n_{Si}b + n_{Air}c)^2}, \end{aligned} \tag{2.40}$$

where

$$a = \cos\phi_{SiO_2} \cos\phi_{Si_xN_y} - \frac{n_{SiO_2}}{n_{Si_xN_y}} \sin\phi_{SiO_2} \sin\phi_{Si_xN_y} , \tag{2.41}$$

$$b = \frac{\sin\phi_{SiO_2} \cos\phi_{Si_xN_y}}{n_{SiO_2}} + \frac{\cos\phi_{SiO_2} \sin\phi_{Si_xN_y}}{n_{Si_xN_y}}, \tag{2.42}$$

$$c = n_{Si_xN_y} \cos\phi_{SiO_2} \sin\phi_{Si_xN_y} + n_{SiO_2} \sin\phi_{SiO_2} \cos\phi_{Si_xN_y} \text{ and} \tag{2.43}$$

$$d = \cos\phi_{SiO_2} \cos\phi_{Si_xN_y} - \frac{n_{Si_xN_y}}{n_{SiO_2}} \sin\phi_{SiO_2} \sin\phi_{Si_xN_y} . \tag{2.44}$$

As before, $\phi_x = 2\pi n_x t_x / \lambda$ (where "x" is a placeholder for either Si_xN_y or SiO_2).

With the availability of the thin-film interference expression in (2.40), knowledge of the SiO_2 and Si_xN_y layer thicknesses t_{SiO2} and t_{SixNy} will allow the estimation of the optical power passing through the SiO_2 and Si_xN_y layers covering the wafer.

To briefly investigate the optical transmission effects of SiO_2 and Si_xN_y layers covering an IC the typical CMOS layer thicknesses in Table 2.1 are used in an example.

Table 2.1. Typical AMS 0.35 µm SiO_2 and Si_xN_y layer thicknesses.

Layer	**Thickness** [nm]
t_{FOx}	55.2
t_{GOx}	7.6
t_{P1}	282
t_{P1M1Ox}	918
t_{M1}	665
t_{IMOX12}	1000
t_{M2}	640
t_{IMOX23}	1000
t_{M3}	640
t_{IMOX34}	1000
$\boldsymbol{t_{SiO2}}$	**6208**
t_{Prot1}	1030
t_{Prot2}	1000
$\boldsymbol{t_{Si3N4}}$	**2030**

Figure 2.24 compares the nominal Si-air, Si-SiO_2-air and Si-SiO_2-Si_xN_y-air interface transmissions of above typical CMOS layer thicknesses.

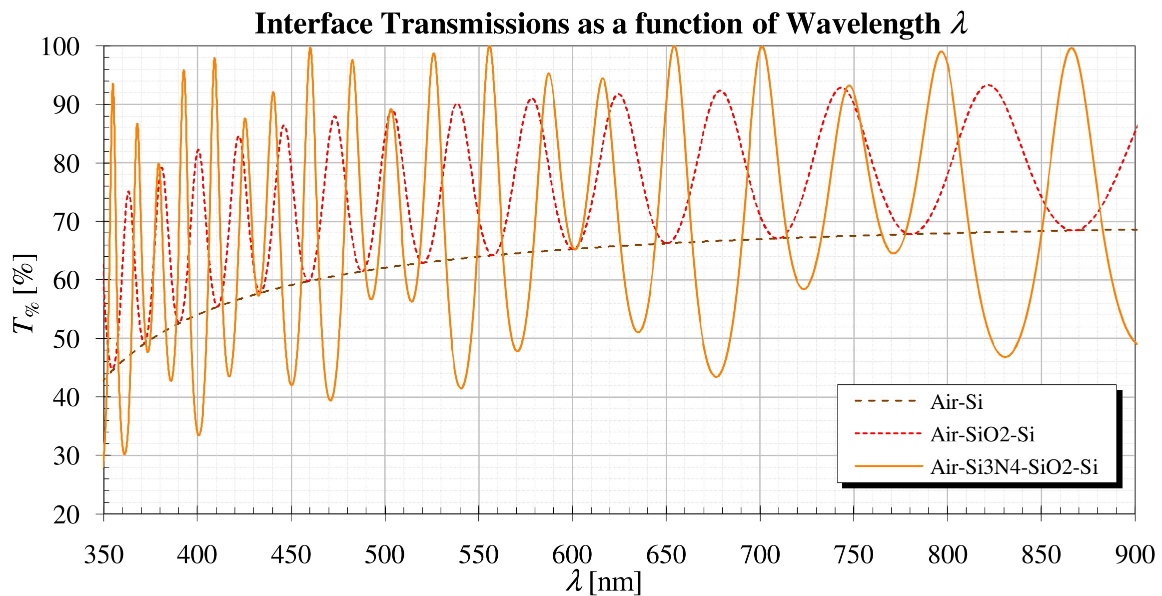


Figure 2.24. Comparison of Air-Si_xN_y-SiO_2-Si, Air-SiO_2-Si and Air-Si transmissions.

For all light transmission profiles above it is evident that the average transmission initially decreases slowly with decreasing wavelength, but then drops off significantly for $\lambda < 400$ nm. This characteristic attenuation of shorter wavelengths is due to the wavelength-dependant refractive index of the bulk Si (section 2.2.2.1) and independent of the SiO_2 and Si_xN_y layer properties.

Since it might be expected that removing the SiO_2 and/or Si_xN_y layer(s) could remove the wavelength-dependant transmission and lead to a higher optical coupling into the air, these scenarios are discussed next.

It is evident in Figure 2.24 that the addition of the SiO_2 layer on top of the silicon “builds on” the Si-air transmission by periodically improving the transmission at certain wavelengths, but never reducing the transmission below the Si-air transmission. This is due to constructive interference of standing waves in the SiO_2 layer.

Adding the Si_xN_y layer on top of the SiO_2 also improves the transmission at certain wavelengths, but reduces the transmission for other wavelengths.

Figure 2.25 compares the average light transmissions through the Si-air, Si-SiO_2-air and Si-SiO_2-Si_xN_y-air interface systems.

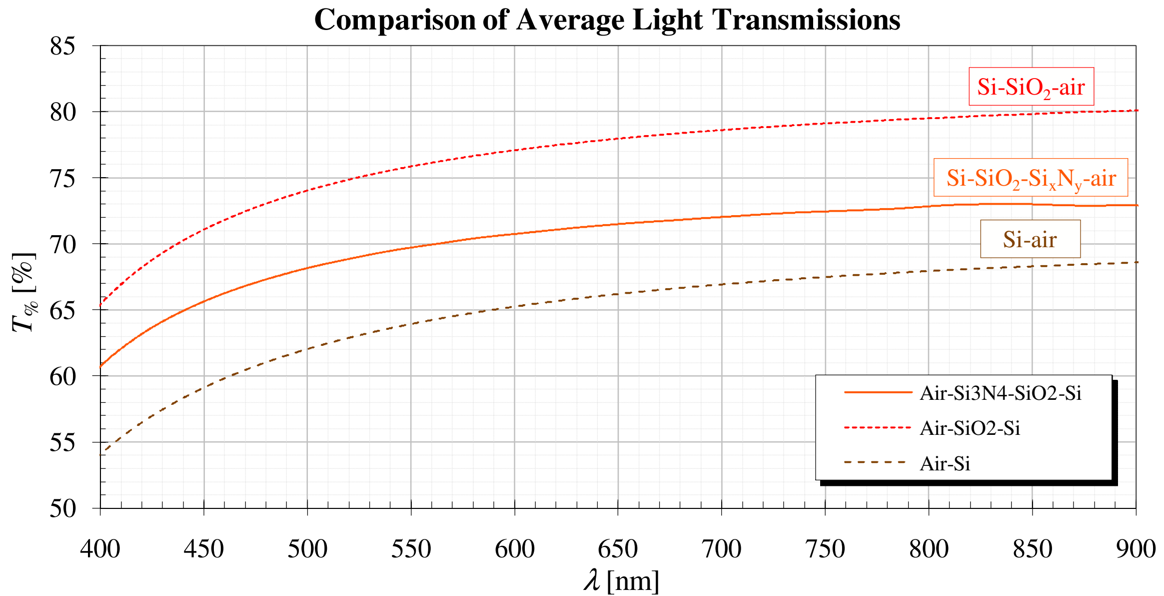


Figure 2.25. Average light transmissions through Si-air, Si-SiO_2-air and Si-SiO_2-Si_xN_y-air stacks.

As shown in above plot, the simple Si-air interface (without SiO_2 and Si_xN_y layers), while eliminating the interference patterns in Figure 2.24, also results in the lowest average light power exiting into the air above the chip.

Figure 2.25 shows that the highest average light transmission occurs for the Si-SiO_2-air interface system that transmits on average about 6 % more than the Si-SiO_2-Si_xN_y-air interfaces and about 12 % more than the Si-air system. The cause of this phenomenon is that while constructive and destructive light interference causes the wavelength-dependant transmission variations (when the light wavelength is a multiple of a layer thickness), the relation between the refractive indices of the materials determines the average light transmission and reflection through them. While the SiO_2 refractive index of about 1.46 is between the refractive indices of air (n_{Air} = 1) and silicon ($3.5 \leq n_{Si} \leq 5$ for $0.4\ \mu\text{m} \leq \lambda \leq 0.9\ \mu\text{m}$), the refractive index of Si_xN_y (≈ 2) is not between the refractive indices of air and SiO_2. The criteria for increasing the average transmission therefore requires that the refractive indices of intermediate layers satisfy the relationship $n_{Si} \geq n_1 \geq \ldots \geq n_x \geq n_{Air}$, where materials 1 to x are arbitrary transparent layers covering the chip. For this reason, the SiO_2 layer acts as antireflection coating that reduces the total light reflection by introducing an intermediate refractive index value between the refractive indices of Si and air.

The previous transmission calculations ignore the optical absorption within the SiO_2 and Si_3N_4 layers as their extinction coefficients are usually considered negligibly small in the wavelength range of interest (350 to 950 nm), but their processing-dependant absorption coefficients could become influential when sufficient impurities are present within them.

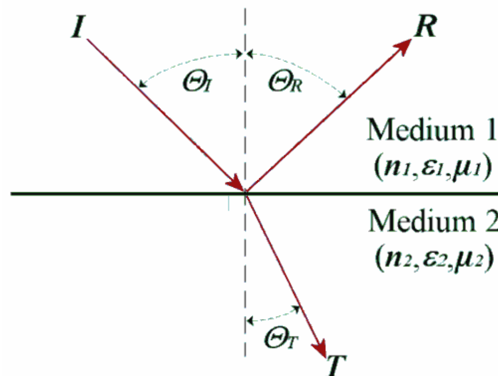
2.2.3.3. Oblique Light Incidence

The thin-film interference analysis in the previous section assumed orthogonal light incidence. Since light from the source will not always pass all interfaces in the optical path perpendicularly, it is worthwhile to calculate the effect of varying incidence angle on material interface transmission and reflection.

With reference to Figure 2.26, Snell's laws of reflection and refraction state that the incidence, transmission and reflection angles of an electromagnetic wave impinging on an interface between two materials with refractive indices n_1 and n_2 and equal permeability ($\mu_1 = \mu_2$) are related by [68]

$$\Theta_R = \Theta_I \text{ and} \tag{2.45}$$

$$\frac{\sin\Theta_T}{\sin\Theta_I} = \frac{n_1}{n_2} = \sqrt{\frac{\varepsilon_1}{\varepsilon_2}} = \sqrt{\frac{\varepsilon_{r1}}{\varepsilon_{r2}}}, \tag{2.46}$$



where equation **Error! Reference source not found.** has been used to relate material refractive index to permeability.

Figure 2.26. Medium and angle definitions for Snell's laws of reflection and refraction.

Figure 2.27 plots the transmission refraction angle Θ_T against incident angle Θ_I when each medium interface is considered separately.

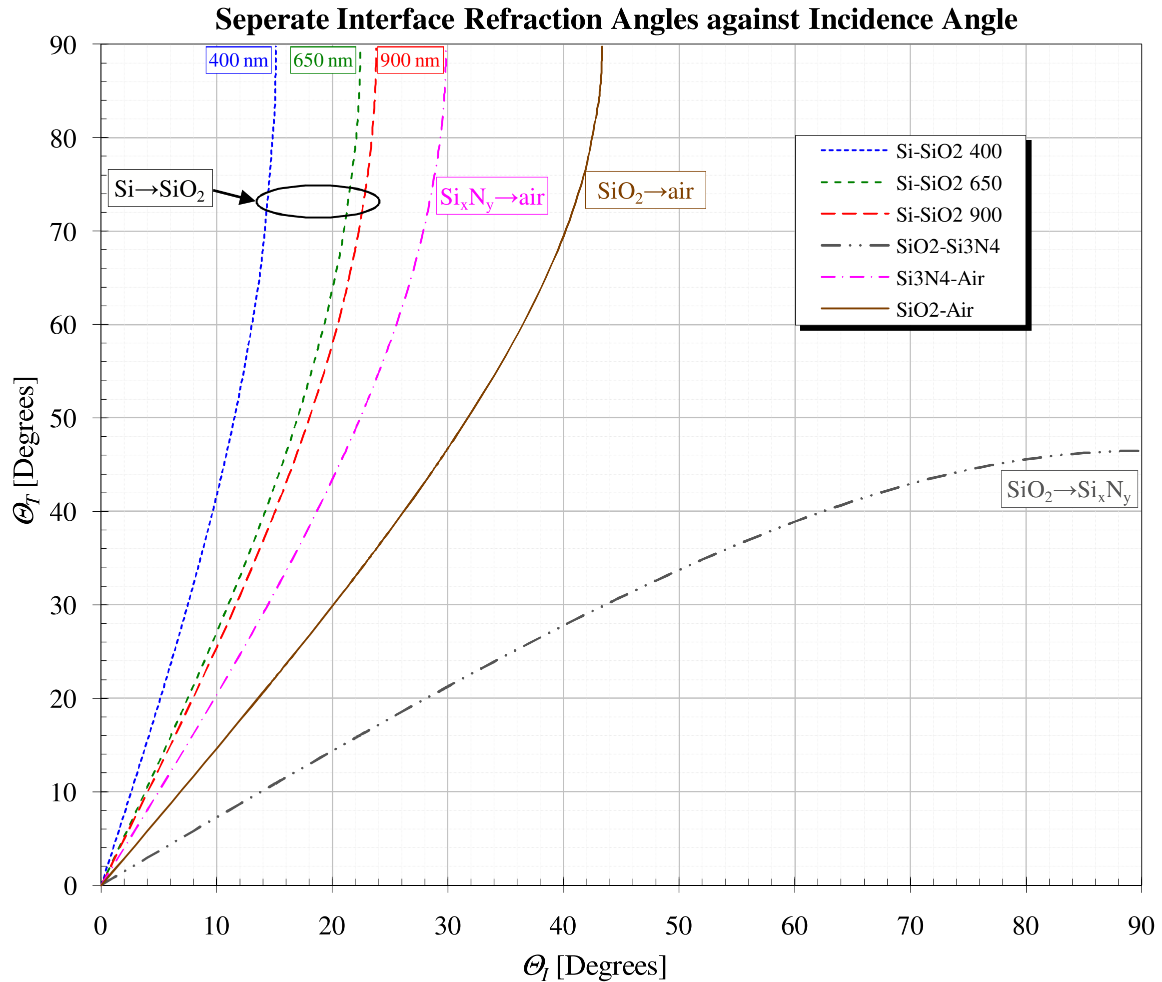


Figure 2.27. Single interface transmission angles against incidence angle.

Except for the $SiO_2 \rightarrow Si_xN_y$ interface, all other interfaces have $\Theta_T > \Theta_I$, i.e. they bend the refracted light away from the normal to the interface.

The maximum exit angle of about 46 ° for the $SiO_2 \rightarrow Si_xN_y$ interface is due the interface's critical angle Θ_C, which is equal to

$$\Theta_C = \sin^{-1} \frac{n_{SiO_2}}{n_{Si_xN_y}} \tag{2.47}$$

and only occurs when light passes from a less dense into a denser medium.

To establish which acceptance angles on the Si side of the Si-SiO_2 boundary will still allow light to exit the top-most SiO_2 or Si_xN_y surface into the air, the refracted angle from one interface can be used as incidence angle for the next interface above it.

This is done in Figure 2.28, which shows the final SiO_2-air exit angle Θ_{Air} against the Si-SiO_2 entry angel Θ_{Si}.

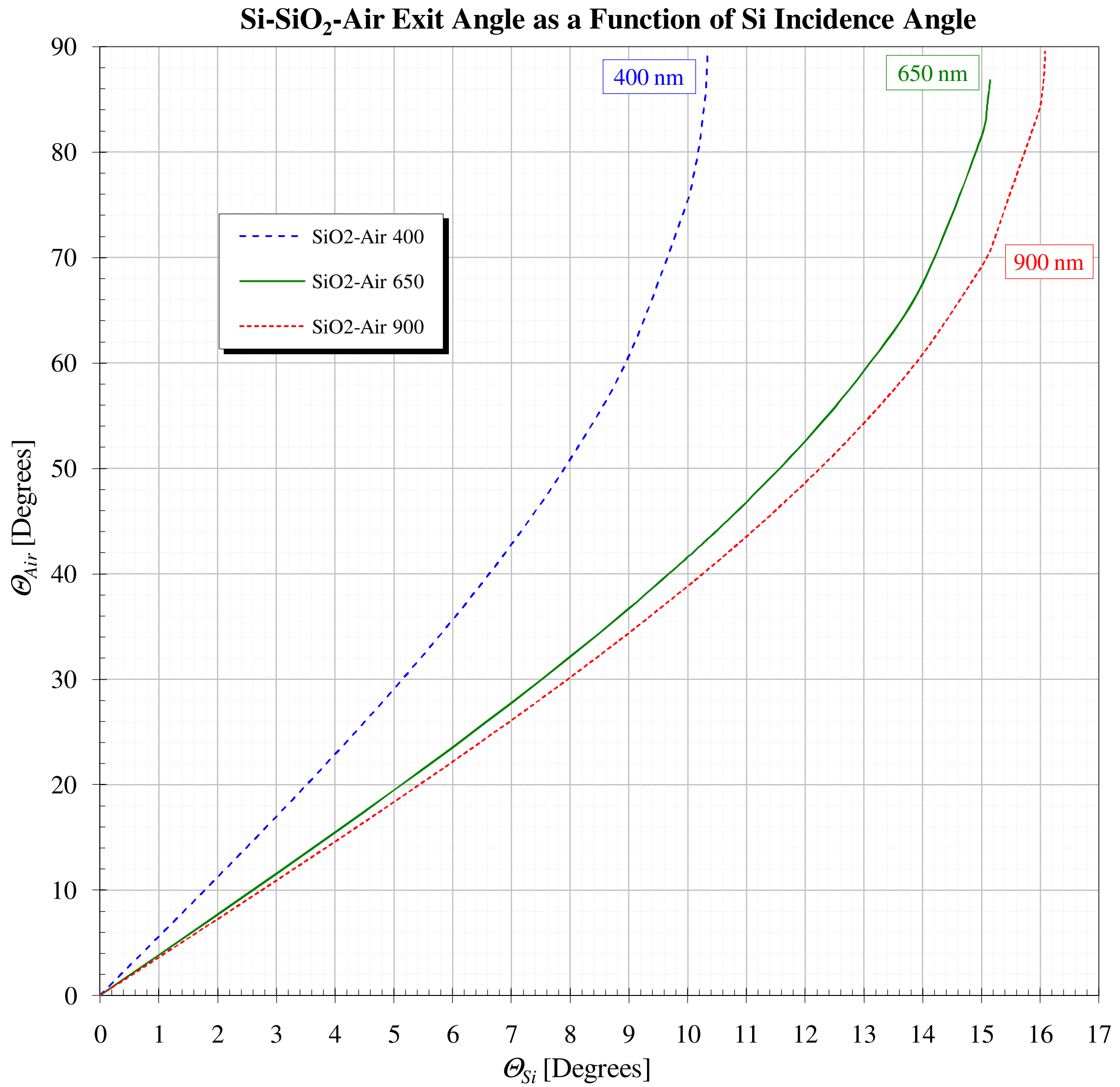


Figure 2.28. Single interface transmission angles against incidence angle.

Figure 2.28 shows that for an air exit-angle Θ_{Air} = 90° the acceptance angle on the silicon side of the Si-SiO_2–air boundary Θ_{Si} varies with wavelength between 10.4° and 16.1° for 400 nm ≤ λ ≤ 900 nm. Exactly the same bulk Si acceptance angle limit is present whether a Si_xN_y layer is introduced between the SiO_2 and air or even when the SiO_2 is removed from the chip's surface.

Symbolic substitutions into equation (2.46) confirm that the exit angle from a stack of transparent media depends only on the refractive indices of the initial and final media if the refraction angles do not exceed the critical angles of intermediate interfaces.

$$\begin{aligned}
\Theta_{Si} &= \sin^{-1}\left(\frac{n_{SiO2}}{n_{Si}}\sin\Theta_{SiO2}\right) \\
&= \sin^{-1}\left[\frac{n_{SiO2}}{n_{Si}}\sin\left\{\sin^{-1}\left(\frac{n_{Si3N4}}{n_{SiO2}}\sin\Theta_{Si3N4}\right)\right\}\right] \\
&= \sin^{-1}\left(\frac{n_{Si3N4}}{n_{Si}}\sin\Theta_{Si3N4}\right) \\
&= \sin^{-1}\left[\frac{n_{Si3N4}}{n_{Si}}\sin\left\{\sin^{-1}\left(\frac{n_{Air}}{n_{Si3N4}}\sin\Theta_{Air}\right)\right\}\right] \\
&= \sin^{-1}\left(\frac{n_{Air}}{n_{Si}}\sin\Theta_{Air}\right).
\end{aligned} \tag{2.48}$$

Above finding makes it clear that, depending on distance from the Si-SiO_2 interface, a large percentage of the light generated at the isotropic Si light source generation site is lost and does not exit into the air above the IC.

One way of making more light externally available would therefore be to manufacture circular or hemispherical light sources in which light reaching the Si surface will impinge with an angle smaller than about 10 ° (for λ = 400 nm).

2.3. Processing Properties

2.3.1. Self-limiting Thermal Oxidation

Some metals (Fe, Al, Si, …) react with oxygen very easily. Over time, oxygen chemically combines with the atoms on the materials surface to form an oxide layer. Two requirements are necessary for this oxidation: oxygen in the environment (O_2 or H_2O) and thermal energy (i.e. heat). Since both of these are readily available in most uncontrolled environments, Si is usually covered with a native oxide layer a few nanometres thick.

Figure 2.29 shows how thermal oxidation reduces the thickness of Si by an amount ΔSi by growing a SiO_2 layer with thickness t_{SiO2} on the Si.

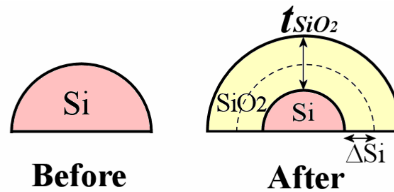


Figure 2.29. Si finger thinning oxidation dimensions.

The ratio of ΔSi removed to SiO_2 thickness t_{SiO2} grown during oxidation$\approx$ 0.44 so that for every micron of SiO_2 grown about 440 nm of Si are consumed.

The thermal oxidation of Si in an oxygen ambient involves two kinetic processes: the diffusion of O_2 through the SiO_2 and the chemical reaction [70]

$$\mathrm{Si} + \mathrm{O}_2 \Leftrightarrow \mathrm{SiO}_2. \tag{2.49}$$

As oxidation progresses in time, the thicker growing SiO_2 layer covering the Si makes it more and more difficult for oxygen to diffuse through it to chemically react with the Si underneath and form SiO_2. This means that the oxidation rate slows down more and more with time until it seems to have stopped. This effect depends on the diffusivity of oxygen in SiO_2, which depends on the oxidation temperature T_{Ox}. At room temperature T_{Ox} is rather low ($\approx$ 300 K), which explains why the native Si oxide thickness is only a few nanometres thick. The self-limiting oxidation effect described by the Deal-Grove model [71] can be clearly seen in the measured thermal oxidation data plotted in Figure D.3.

The SiO_2 thickness t_{SiO2} depends on oxidation temperature and oxidation time t_{Ox} through

$$t_{SiO_2}{}^2 + A t_{SiO_2} = B\left(t_{Ox} + \tau\right), \tag{2.50}$$

where A [µm] is the linear growth factor, B [µm^2/h] is the parabolic growth factor and τ [h] is the oxidation time offset that compensates for an initial oxide thickness at $t_{Ox} = 0$.

A and B depend on oxidation temperature and whether the oxidation is wet or dry. Since the initial (native) oxide thickness for wet oxidations is very small, $\tau \approx 0$ for wet oxidation.

For short oxidation times the oxidation is reaction rate limited and the oxide growth in this "linear regime" is calculated as

$$t_{SiO2} = \frac{B}{A}(t_{Ox} + \tau). \tag{2.51}$$

Longer oxidation times are diffusion limited where the oxide growth in the "parabolic regime" can be expressed as

$$t_{SiO2}{}^2 = B(t_{Ox} + \tau). \tag{2.52}$$

For dry oxidation the parabolic SiO_2 growth rate constant B at an oxidation temperature T_{Ox} can be expressed as

$$B = 772\,\mathrm{e}^{\frac{-1.23\,\mathrm{eV}}{k_B T_{Ox}}}\ [\mu\mathrm{m}^2/\mathrm{h}], \tag{2.53}$$

and for wet oxidation with water temperature T_{H2O} as

$$B = \left(0.217\,\mathrm{e}^{\frac{T_{H_2O}}{62.1°C}}\right)^2 386\,\mathrm{e}^{\frac{-0.78\,\mathrm{eV}}{k_B T_{Ox}}}\ [\mu\mathrm{m}^2/\mathrm{h}]. \tag{2.54}$$

The parabolic SiO_2 growth rate constant B is about 25 times higher at T_{Ox} = 1050 ºC in a wet ambient compared to dry oxidation. This is primarily due to the much higher solid solubility of H_2O in SiO_2 than of O_2 in SiO_2, thus providing a much higher supply of oxidizing O_2 to the underlying Si for oxidation.

Etching the SiO_2 away and subjecting the Si again to an thermal oxidizing environment starts the oxidation sequence again in the linear region.

Oxidation of Si with a non-planar surface (i.e. wires, dots or other structures with curved surfaces) introduces another limiting factor into the oxidation process. As SiO_2 grows on the curved Si surface, the accumulating mechanical stress in the oxide slows down the chemical oxidation reaction rate and the diffusion of O_2 through the stressed SiO_2 ([72] and [73]). Similar stresses in grown SiO_2 are also responsible for the well-known "birds beak" shape of the LOCOS oxide [74].

This effect is especially pronounced in Si nano-structures since with decreasing structure dimensions and consequently decreasing surface curvature radius the oxidation rate decreases much more dramatically compared to planar bulk Si [75]. These self-limiting oxidation effects can therefore be used to create nanometre-scale Si structures like wires and nano-dots with predictable final dimensions that can be estimated from the initial Si dimensions, crystal orientation and the oxidation parameters (oxidation time and temperature) [76].

According to Kedzierski *et al* [77] the oxidation progress of SOI Si nano-wires can be expected to progress as shown in Figure 2.30.

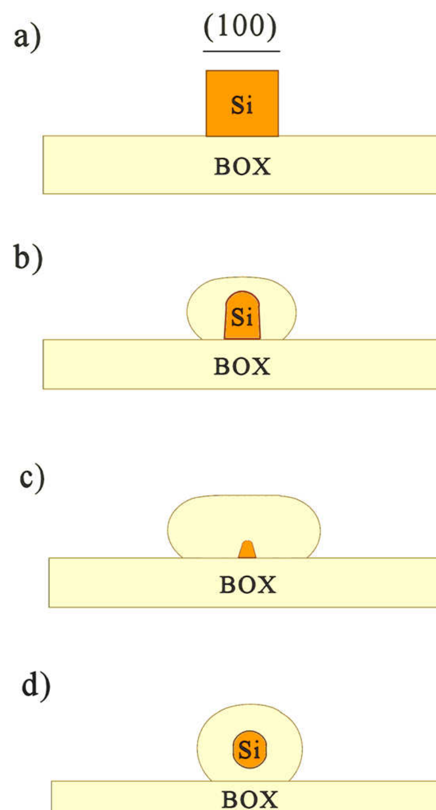


Figure 2.30. Si shape a) before oxidation, during b), c) continuous and d) undercutting oxidation.

Figure 2.30 a) shows the cross-sectional view of the initial Si bar. Figure 2.30 b) then shows that the Si wire will first be predominantly oxidized from the sides parallel to the (100) plane and from the top. Since less O_2 diffuses to the bottom of the structure, it will oxidize slower than the top. Figure 2.30 c) shows that if oxidation is continued, the wire will first get thinner from the sides until the top side of the Si is consumed and a smaller triangular shape remains close to the BOX. Figure 2.30 d) shows one way of achieving round Si finger wires. By etching the grown SiO_2 and the BOX, the Si finger is also oxidized from the bottom, which gives it a rounder shape.

2.3.2. Dopant-dependant Oxidation

The oxidation of extrinsic silicon can be modelled by modifying the Deal-Grove linear-parabolic growth law in (2.51) in subsection 2.3.1. The dependence of silicon oxidation kinetics on doping concentration manifests as part of the linear rate constant, where the physical effects of higher doping levels has been explained primarily as an electrical effect. The modified linear rate constant that includes the doping dependence becomes [78]

$$\frac{B}{A} = \left(\frac{B}{A}\right)_i \left(\frac{B}{A}\right)_{Doping} \tag{2.55}$$

where $(B/A)_i$ is the linear rate constant on intrinsic Si and [79]

$$\left(\frac{B}{A}\right)_{Doping} = \left\{1 + B_{K0} e^{-\frac{B_{FE}}{k_B T}} \left(\frac{V^*}{V_i^*} - 1\right)\right\} \tag{2.56}$$

where V^* is the equilibrium vacancy concentration in the doped Si at the Si/SiO_2 interface, V_i^* is the equilibrium vacancy concentration in intrinsic Si, B_{K0} and B_{FE} are experimentally-determined Arrhenius coefficients relating to the doping dependence of the oxidation rate.

In general, the doping type and concentration in the Si cause a Fermi level shift that alters the Si equilibrium vacancy concentration, composed of vacancy defects in different charged states [80]. Figure 2.31 shows the functional dependence of the equilibrium vacancy concentration at 950 ºC versus doping concentration for commonly used Si dopants.

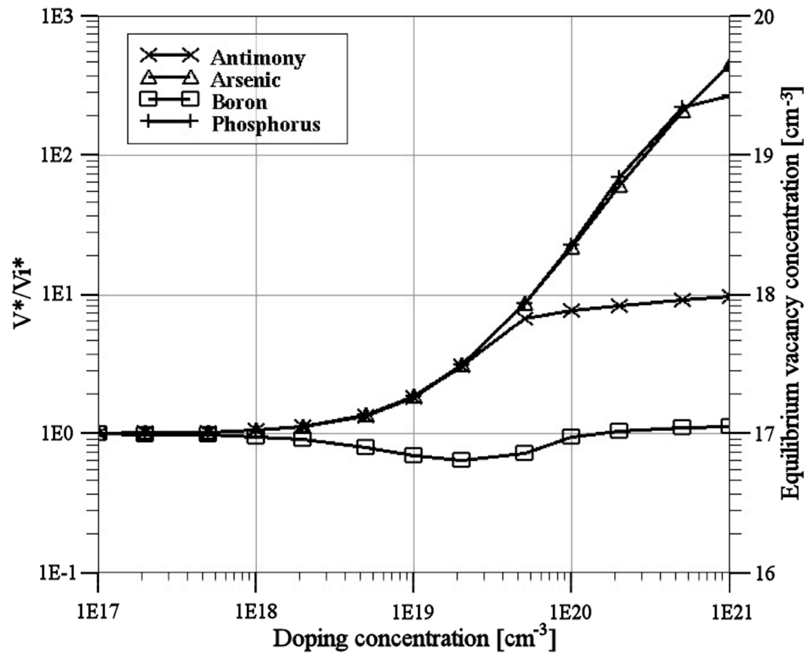


Figure 2.31. Si equilibrium vacancy concentration versus doping density and species at 950 ºC.

The physical significance of an increase in the vacancy concentration is a higher availability of unoccupied Si lattice reaction sites for the incoming oxidant molecules, which in turn enhances the oxidation rate.

An increased equilibrium vacancy concentration and consequently larger Si oxidation (with respect to the lower doping concentrations) is observed with increasing *n*-type doping, but remains essentially constant for the *p*-type dopant.

2.4. Summary

Section 2.1 revised important semiconductor properties of Si. Aspects like resistivity, intrinsic carrier concentration, built-in potential and depletion region width were used to design the layout dimensions of the SOI light sources while the influence of silicon's energy band diagram, impact ionization, avalanche breakdown, punch-through and reach-through properties on electroluminescence were introduced.

Subsection 2.2.1.2 investigated the radiative recombination mechanism in Si.

Subsection 2.2.1 determined that the optical power spectrum expected from a Si nanometre-scale finger junction would be similar to the typical CMOS light source power spectrum in Figure 2.14, but that it would have a higher amplitude with higher intensities for shorter wavelengths.

Subsection 2.2.1.4 discussed how carrier injection can be used to increase the quantum efficiency and resultant power efficiency of integrated Si light sources.

Comparing the average light transmissions through the Si-air, Si-SiO_2-air and Si-SiO_2-Si_xN_y-air interface systems in Figure 2.25 (subsection 2.2.3.2) revealed that the Si-SiO_2-air system transmits the most optical power from the Si light source into the air above.

Section 2.3 showed that geometrical factors, doping species and concentration affect the oxidation rate of Si.

3. PHYSICAL AND OPTICAL SIMULATION

Simulations were useful during the design stage to predict physical and optical properties achievable at the end of the manufacturing process.

3.1. Impurity Redistribution during Oxidation

The low diffusivity and solubility of arsenic (As) in SiO_2 results in the "snow-shovel" effect that causes As to pile up against the moving SiO_2 boundary during thermal oxidation, ([81] - [83]).

Boron (B) with its higher diffusivity and solubility in SiO_2 is absorbed into the SiO_2 during thermal oxidation, therefore decreasing its concentration in the Si, ([84] - [89]). Since the lower B background doping concentration of n^+p junctions plays a larger role in determining junction characteristics like depletion region width w_d than the higher As concentration, only the spatial B concentration variation with thermal oxidation was further considered.

The SOI B concentration $C(y,t)$ as a function of distance y from the BOX-Si interface and oxidation time t during thermal oxidation can be expressed as [84]

$$\frac{C(y,t)}{N_B} = 1 - \frac{\frac{k-m}{2}\sqrt{\frac{B\pi}{D_B}}\left\{\operatorname{erfc}\left(\frac{l-m\sqrt{Bt}-y}{2\sqrt{D_B t}}\right)+\operatorname{erfc}\left(\frac{l-m\sqrt{Bt}+y}{2\sqrt{D_B t}}\right)\right\}}{1-\mathrm{e}^{-\left(\frac{l-m\sqrt{Bt}}{\sqrt{D_B t}}\right)^2}+\frac{k-m}{2}\sqrt{\frac{B\pi}{D_B}}\left\{1+\operatorname{erfc}\left(\frac{l-m\sqrt{Bt}}{2\sqrt{D_B t}}\right)\right\}}, \tag{3.1}$$

where $m \approx 0.44$ is the ratio of ΔSi removed to SiO_2 thickness t_{SiO2} grown during oxidation (section 2.3.1), B is the parabolic SiO_2 growth-rate constant, D_B the B diffusion constant and l is the initial SOI active layer thickness. $k = C_{SiO2}/C_{Si}$ is the segregation coefficient defined by C_{SiO2}, the B concentration on the SiO_2 side of the Si/SiO_2 boundary and C_{Si} the B concentration on the Si side of the same interface.

The B diffusion constant D_B can be expressed as

$$D_B = D_{B\infty}\,\mathrm{e}^{\frac{-E_{As}}{k_B T_{Ox}}}\,[\mathrm{cm}^2/\mathrm{s}], \tag{3.2}$$

where $D_{B\infty}$ is a B diffusion constant and E_{As} is the activation energy for B diffusing through Si.

Figure 3.1 shows the simulated B concentration decrease from the initial implanted concentration during successive oxidations.

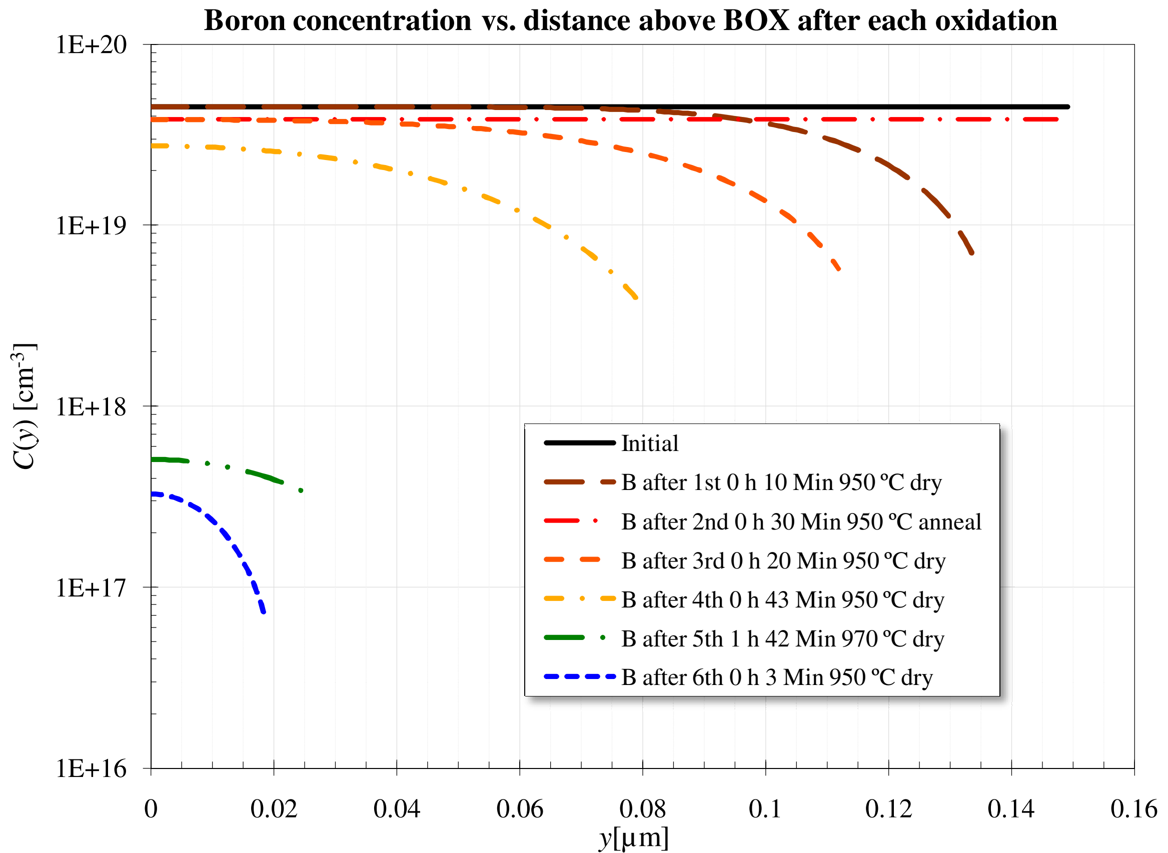


Figure 3.1. Boron redistribution during thermal oxidation and anneal processing steps.

Prior knowledge of oxidation steps therefore allowed specifying the initial B implantation dose so that the desired final average finger background B concentration of ≈ 10^{17} cm^{-3} would be achievable.

In reality, some B segregation and re-diffusion into the Si also occurs at the BOX interface [90], but was ignored in the simulation shown in above Figure.

3.2. Optical Radiation Simulation

Employing the geometrical optics RAYTRACE[1] software enabled the simulation of spatial light radiation characteristics of the Si fingers with varying finger geometries and light source locations.

Figure 3.2 for example shows how light generated at the centre of a finger with rounded corners focuses into four lobes (recognizable by higher ray density) emanating from the rounded top surface of the Si finger.

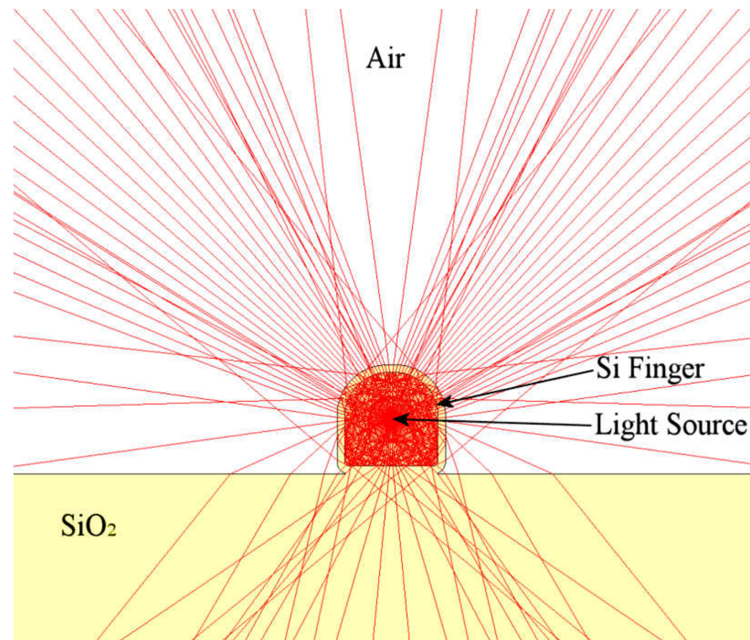


Figure 3.2. SOI finger spatial light radiation pattern simulation.

It was *inter alia* determined that the maximum useful light emission directed away from the chip surface is achieved with a hemispherical round top finger surface to minimize internal reflection beyond the critical interface incidence angle and a flat bottom surface to maximize reflection back to the top surface. Shaping the Si fingers in this way can increase the usefully radiated light by up to a factor five.

[1]Version 2.22, IME software, Ian Moore 1994 – 2006, http://www.ozemail.co.au/~imesoft.