Process fabrication III

E. Giulio Villani

UNIVERSITY OF OXFORD

Overview

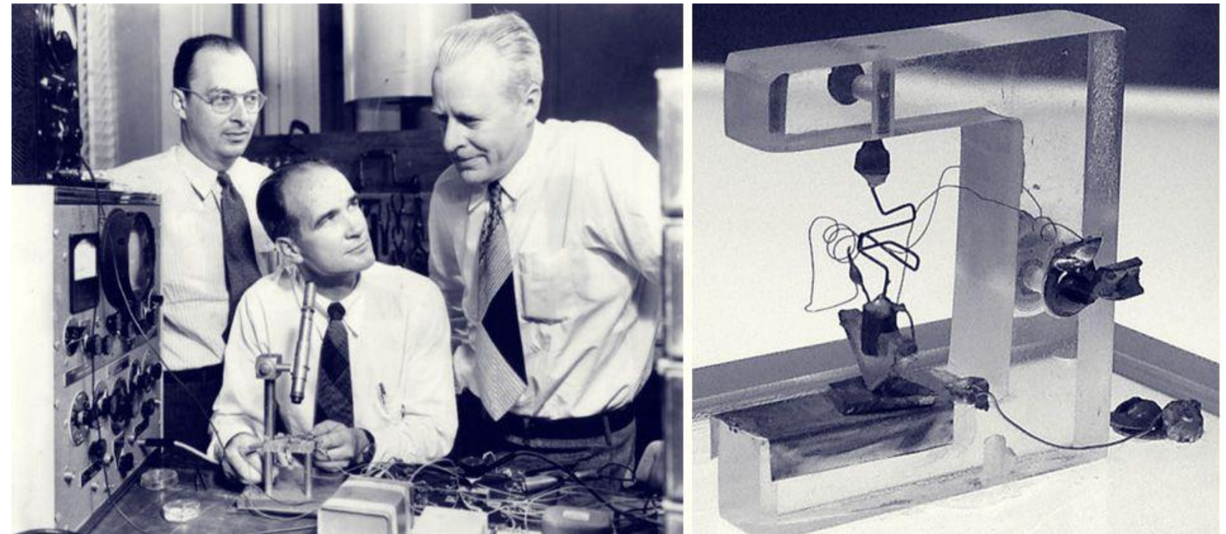
- **Semiconductor technology**
  - *Introduction*
  - *Moore's Law*
  - *Dennard's Law*
- **Semiconductor fabrication process III**
  - *Silicon crystal growth*
  - *Fabrication process flow*
  - *Oxidation*
  - *Resist deposition*
  - *Photolithography*

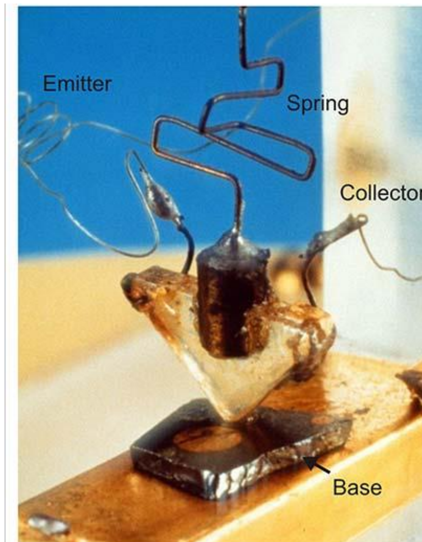
Introduction

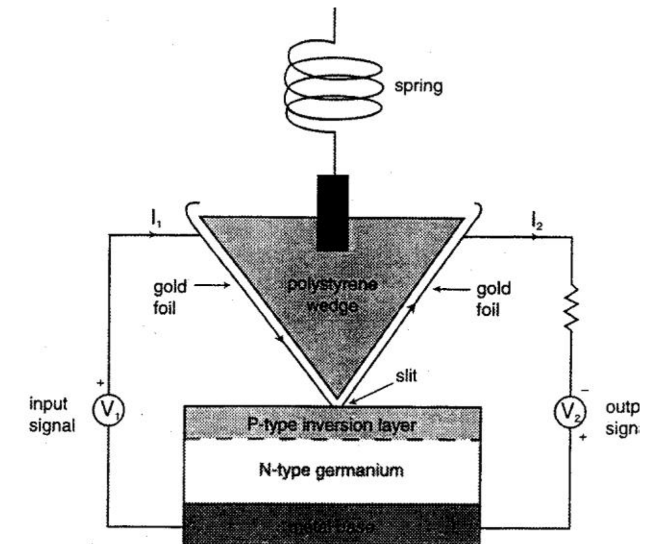
- First bipolar (PNP Ge) transistor invented in 1947 (*US patent 2569347A, 1948*)

- Research on solid state devices to replace vacuum tubes

- Resulted from failed attempts to build a field-effect transistor (FET)





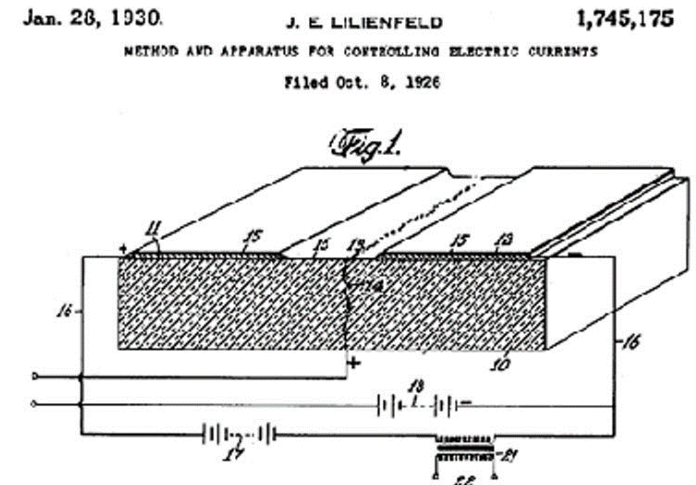


Introduction

- First FET (Field Effect Transistor) patented filed in 1930 (*US patent 1745175A,1930*)

- Serious problems with surface states/ dielectric defects prevented the devices from working

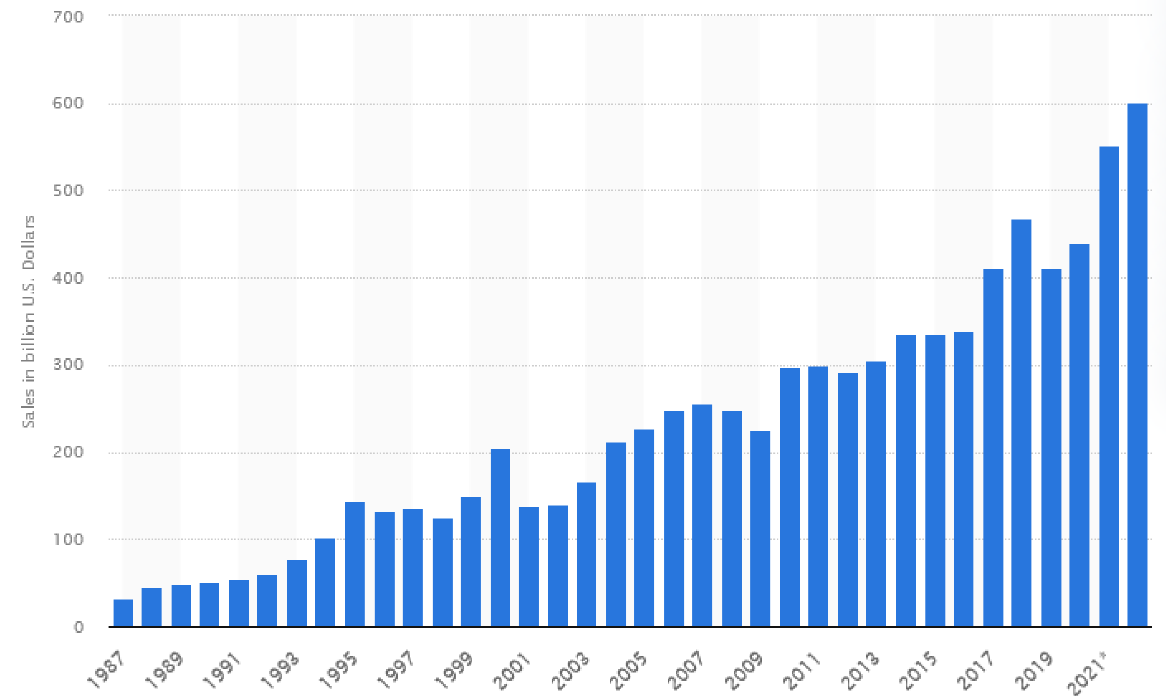
- First practical working device invented in 1959, using SiO_2 as dielectric (MOSFET, Metal Oxide Silicon Field Effect Transistor)



UNIVERSITY OF OXFORD

Introduction

- MOSFETs by far the most widely used (99%) transistor device
- Estimated > 10^{22} devices produced since 1960 (> 10^{9} produced / sec)
- Sales of semiconductors in 2022 $\simeq$ 574 10^{9} USD



Semiconductor market size worldwide from 1987 to 2022

Moore's Law

- Semiconductor industry started around 1960 (First hybrid circuit:, Kilby, Jack S. "Miniaturized Electronic Circuits", U.S. Patent 3,138,743, February 1959)
- In 1965 G. Moore predicted a doubling number of integrated components every year, to promote the idea of integrating devices. From mid '70s the trend is around doubling every ~2 years
- Such trend has been maintained for almost 60 years, leading to $\sim 2^{60/2} \sim 10^9$ transistors count on chip

Moore's Law: The number of transistors on microchips doubles every two years

Our World in Data

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.

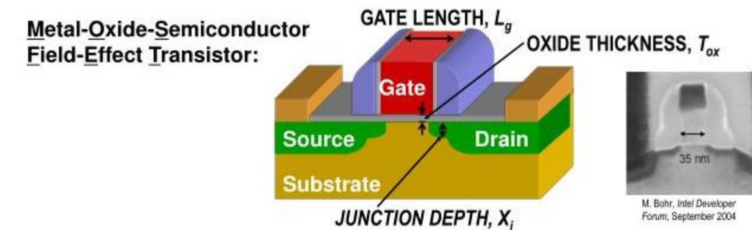
Transistor count

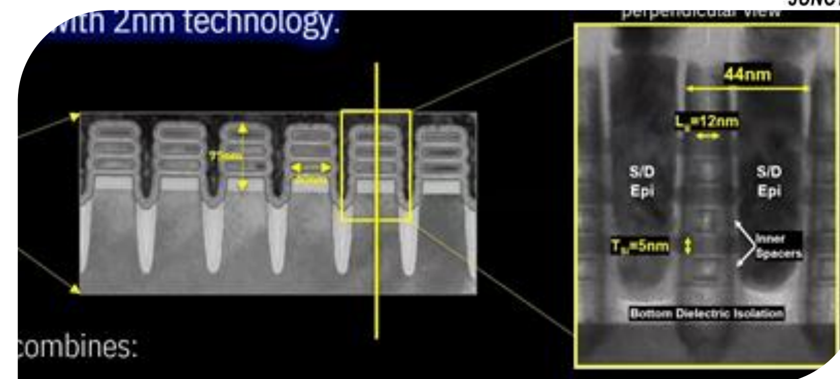
Moore’s Law

- Technology size referred to actual minimum MOS transistor gate length
- Reducing the size of the devices requires considerable technological efforts and investments
- Since the advent of 3D structures the name of nodes means an equivalent gate length size of a planar device

Peak Quoted Transistor Densities (MTr/mm2)				
AnandTech	IBM	TSMC	Intel	Samsung
22nm			16.50	
16nm/14nm		28.88	44.67	33.32
10nm		52.51	100.76	51.82
7nm		91.20	237.18*	95.08
5nm		171.30		
3nm		292.21*		
2nm	333.33			

Data from Wikichip, Different Fabs may have different counting methodologies
* Estimated Logic Density





https://www.anandtech.com/show/16823/intel-accelerated-offensive-process-roadmap-updates-to-10nm-7nm-4nm-3nm-20a-18a-packaging-foundry-emib-foveros

Dennard’s Law of scaling

- From Dennard’s 1974 paper: smaller feature size can only bring benefits
- Smaller -> Faster, Less power dissipation; Power density remains the same!

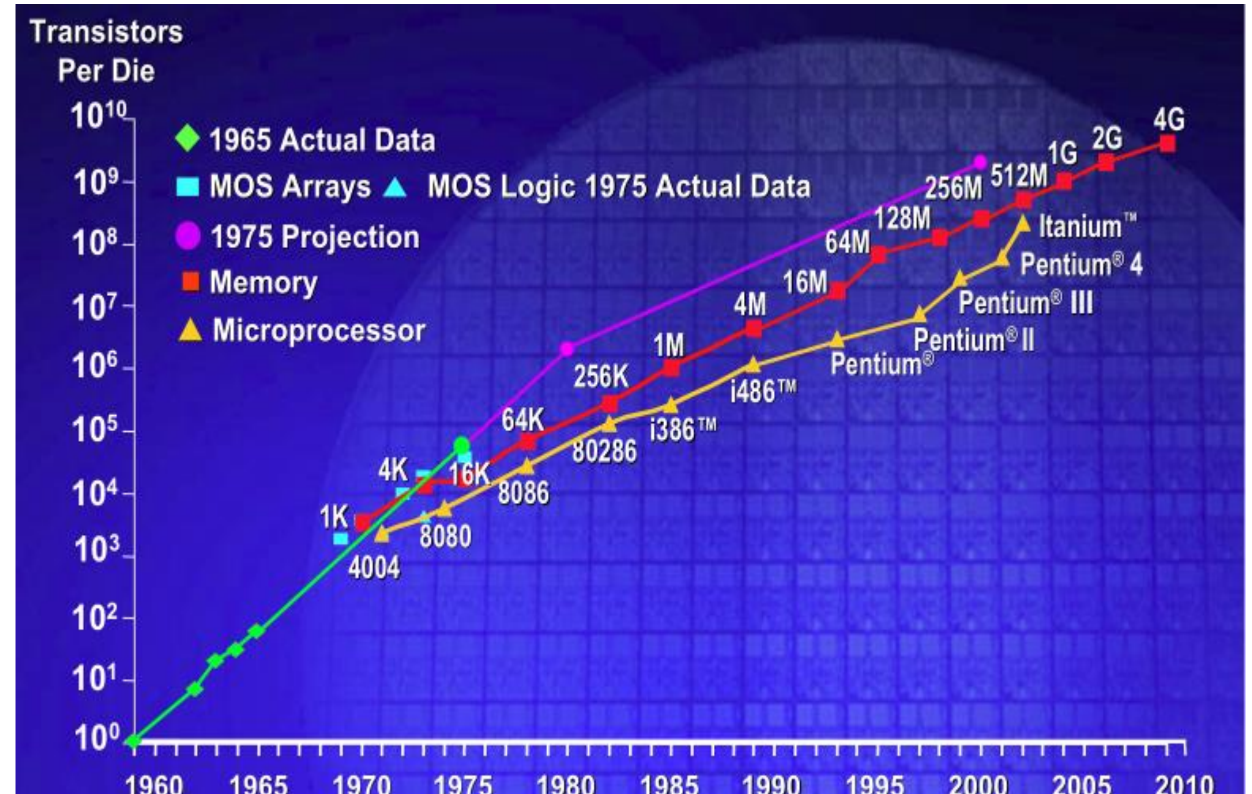
Device or Circuit Parameter	*Scaling Factor*
Device dimension t_{ox}, L, W	$1/k$
Doping concentration N_a	k
Voltage V	$1/k$
Current I	$1/k$
Capacitance eA/t	$1/k$
Delay time per circuit VC/I	$1/k$
Power dissipation per circuit VI	$1/k^2$
Power density VI/A	1

Table I: Scaling Results for Circuit Performance (from Dennard)

R. H. Dennard et al., "Design of ion-implanted MOSFET's with very small physical dimensions," Oct. 1974, doi: 10.1109/JSSC.1974.1050511

Dennard's and Moore's Law

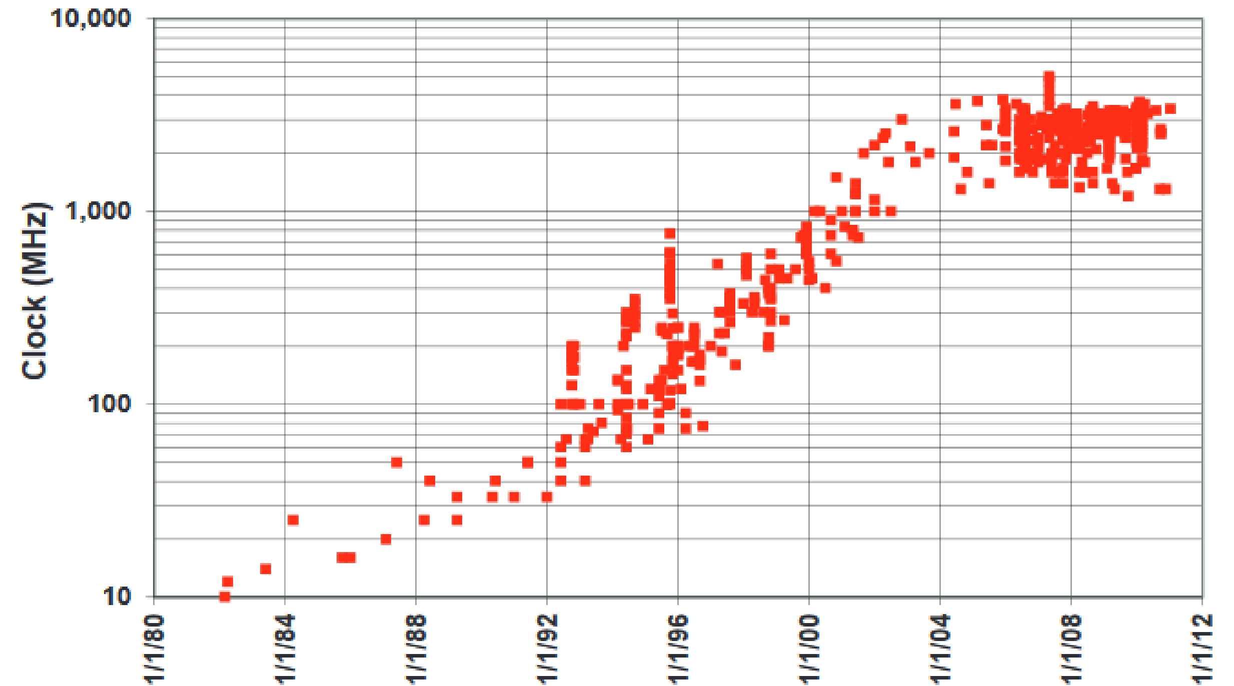
- By keeping the cost/area of chip constant:
  - More powerful chip for the same price
  - Same chip for a lower price: the scaling down of transistor area reduces the cost of a transistors by ~30%/year.
  - This trend has continued for around 20-25 years (1975-2000)

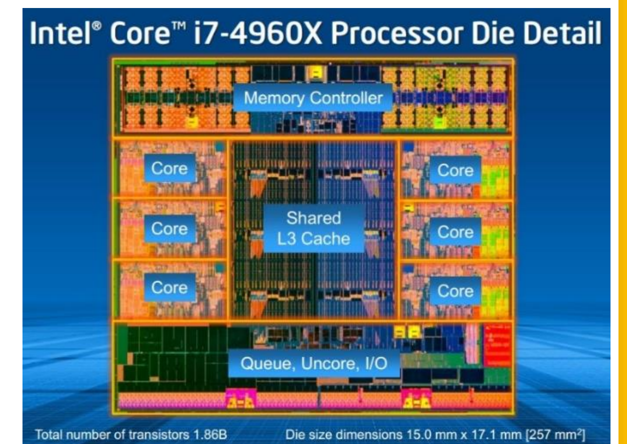


UNIVERSITY OF OXFORD

Dennard’s Law of scaling

- This scaling ignores **quantum tunnelling** and **noise**, i.e. leakage current, which leads to power dissipation, and threshold voltage which do not scale with size (~1/√k)
- Power wall, limiting the frequency of processors to around 4 GHz since 2006



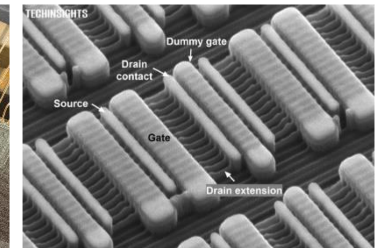


Dennard's and Moore's Law

- Moore's law is not a physics law per se: keeping the cost/area the same (for example increasing the wafer size) despite rising cost of fabrication is the result of human efforts.
- Dennard's law is based on physics, but it does not completely justify anymore the continuous trend in size reduction.
- Nowadays the main benefit in reducing the size of transistor is to reduce cost/function

Semiconductor fabrication process

- The fabrication of semiconductor devices is a rather complex process
- Many intermediate steps and manufacturing precision at atomic level
- It starts with the growth of high purity Si crystals



Silicon crystal growth Czochralski

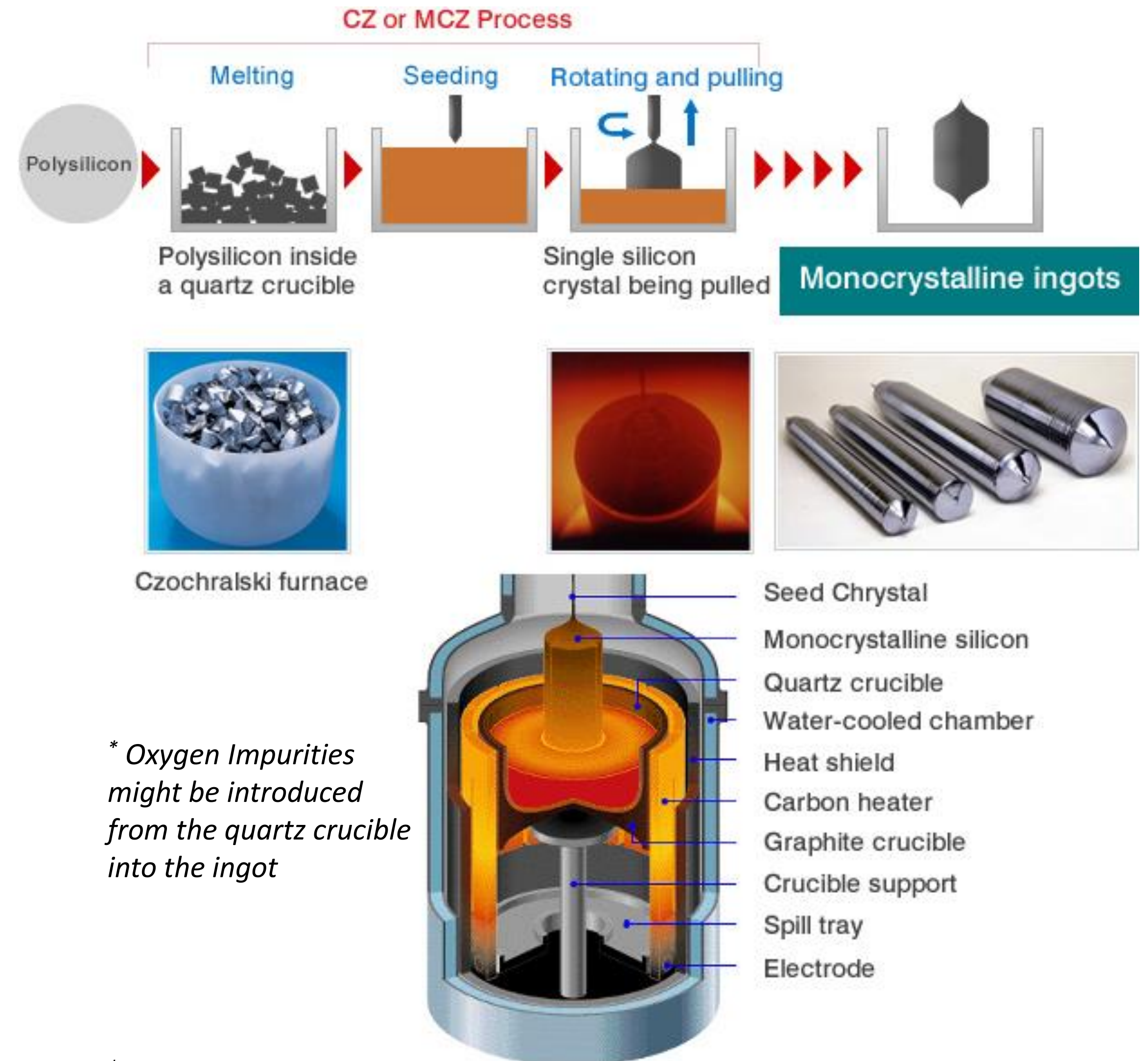
- The majority of ICs are made on single-crystal Si wafers grown in large ingots using **CZ** (Czochralski, 1915 [1]) method

  1. High purity polysilicon (99.9999999% or 9N) crushed into powder, put into a quartz crucible and heated until it melts

  2. A seed crystal is dipped into liquid silicon. As the stick is rotated and slowly pulled up, an ingot of monocrystalline silicon is formed that has the same atomic arrangement as the seed crystal

- For detector grade Si, the smallest concentration of contaminants is required. The Floating- Zone (FZ [2]) process is usually employed for this

[1] Development of Crystal Growth Technique of Silicon by the Czochralski Method, Vol. 124 ACTA PHYSICA POLONICA, 2013
[2] T.F. Ciszek, T.H. Wang, Silicon defect and impurity studies using float-zone crystal growth as a tool, Journal of Crystal Growth, 237, p. 1685-1691, 2002



** Oxygen Impurities might be introduced from the quartz crucible into the ingot*

** Czochralski discovered this by accident: instead of dipping his pen into his inkwell, he dipped it in molten tin, and drew a tin filament, which later proved to be a single crystal*

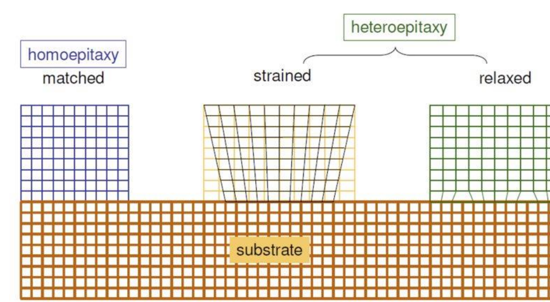
Silicon crystal growth
Epitaxy

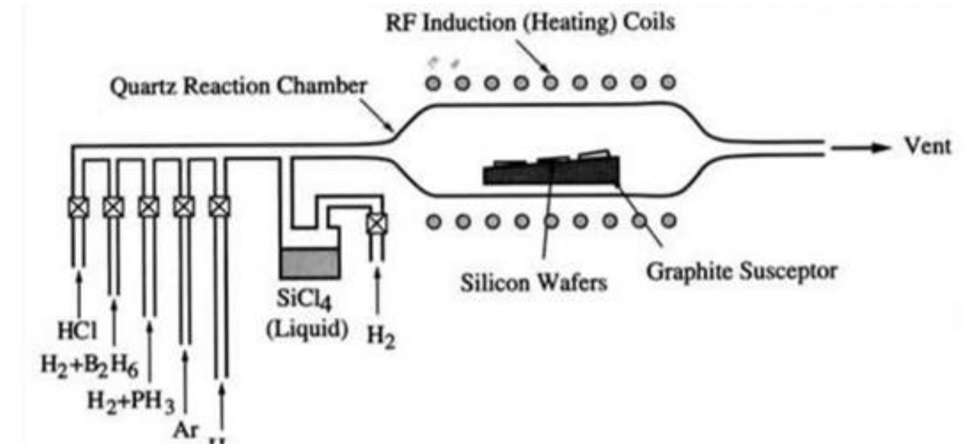
- Arrangement of atoms over existing planes of crystalline substrate, to form an extended crystal

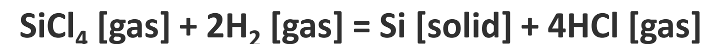
- The deposited layer can have very different doping of the substrate, allowing doping profiles unattainable with implantation.

- Different materials can be grown (heteroepitaxy)

- The epitaxially grown layer can be oxygen free







Chemical Vapour Deposition (CVD) is the formation of solid films on a substrate by exploiting a chemical reaction of reactants in vapour phase.

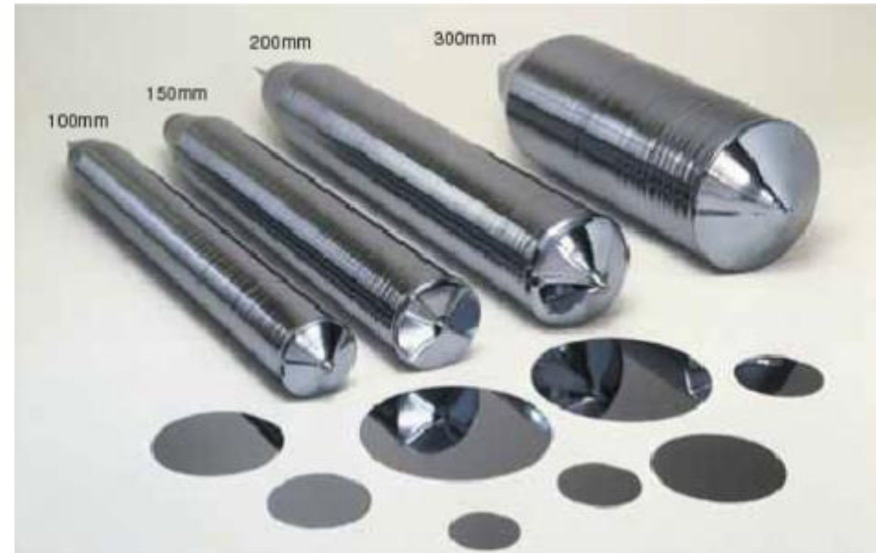
Reactants introduced in reaction chamber decompose and react with the heated surface to form the film

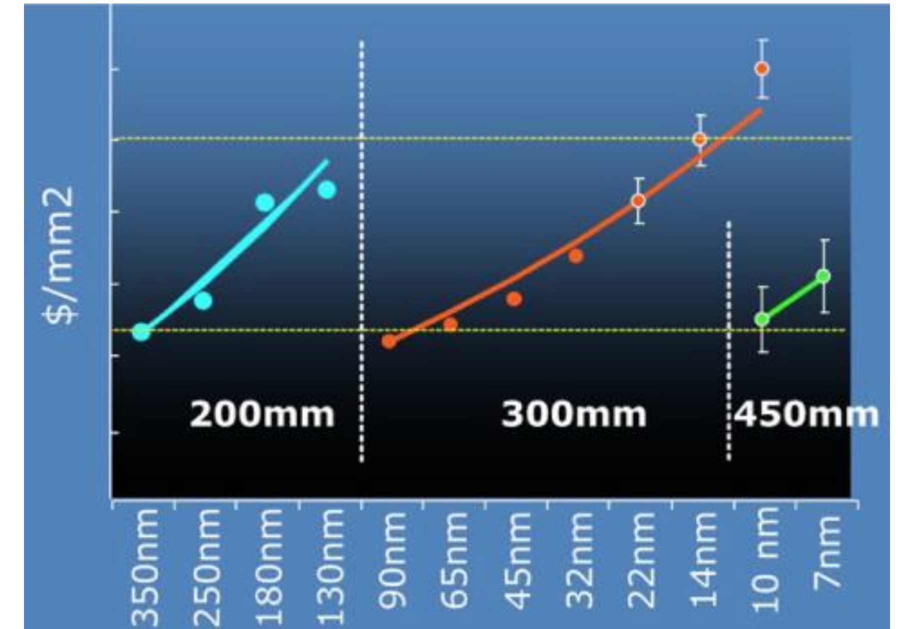
Silicon crystal

- The cylindrical ingot is sliced into thin circular wafers, polished, etched, and cleaned until their surface is almost roughness free (<< 1 nm)

- The diameter of Si ingots grew over the years: currently 300 mm diameter, driven by keeping the cost/area constant

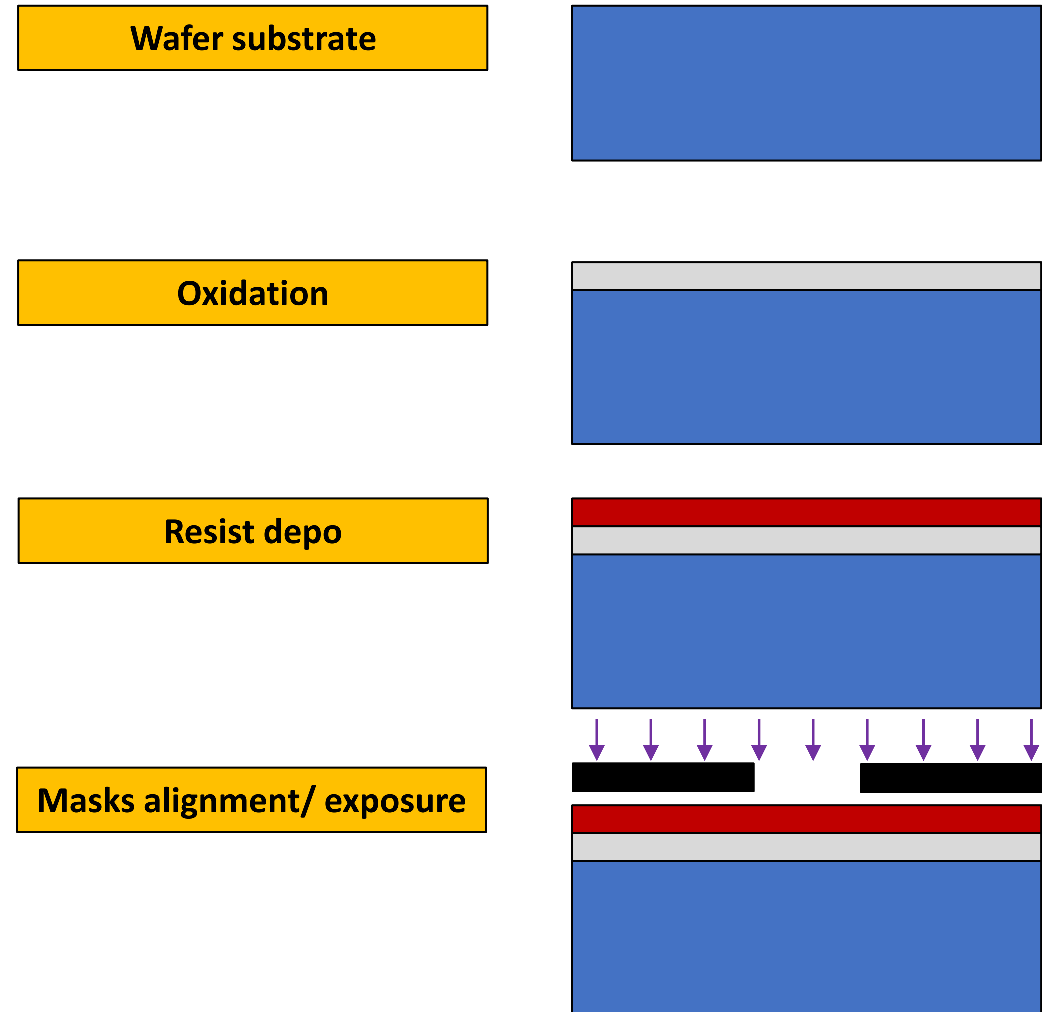
- As per 2022, worldwide Si wafer area around 10e6 m^2 , for an approximate 1.15×10^{12} semiconductor **chips** produced





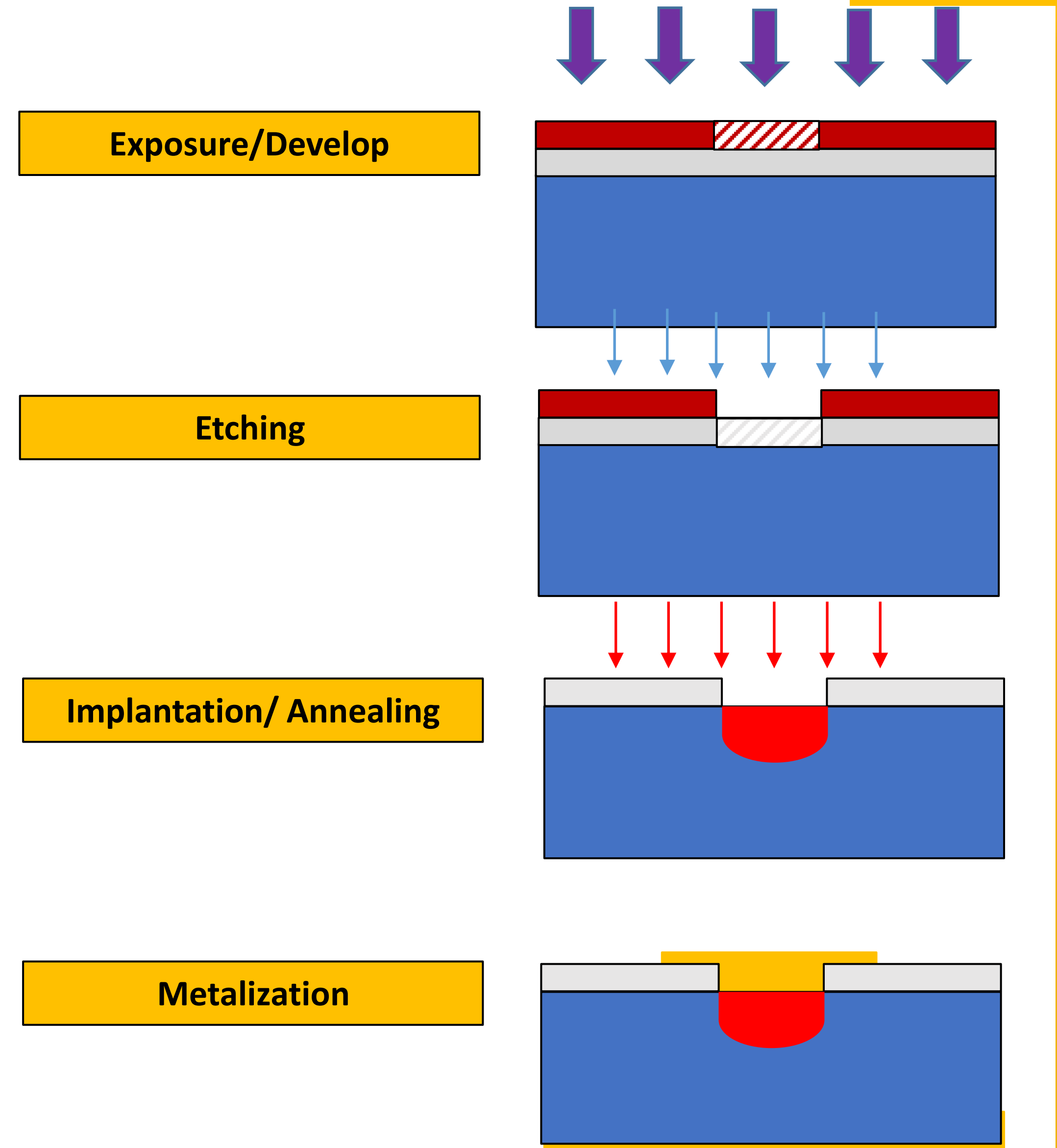
Silicon process fabrication example

- Start with Si wafer cleaned and polished
- The wafer top is covered by insulating layer, SiO_2
- Light-sensitive layer (photoresist) deposited
- A photomask with the desired pattern is aligned with the wafer



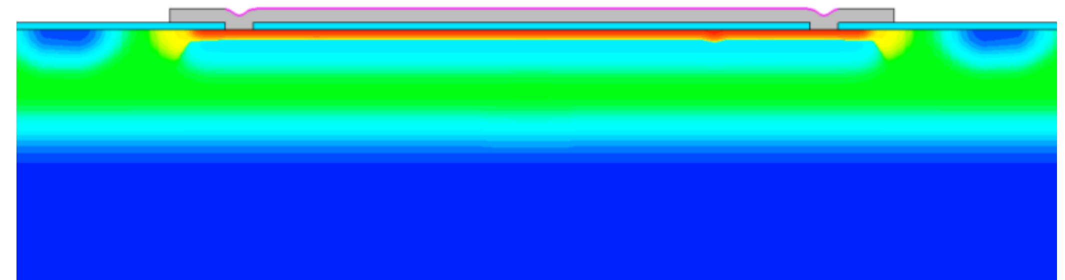
Silicon process fabrication example

- Exposure to UV light makes the photoresist not covered by masks easily removable
- Once the developed resist is removed, the unprotected SiO_2 is etched away, using chemical process. Remaining resist is stripped off
- The exposed areas of Silicon are then doped, e.g. to obtain PN junctions
- Metallization followed by a similar photolithographic process to obtain a final device



Silicon process fabrication example

- Additional layers of conducting or insulating materials can be added, to obtain multilayers structures
- 10s of layers in modern submicron CMOS process



LGAD cross section example – 1 layer

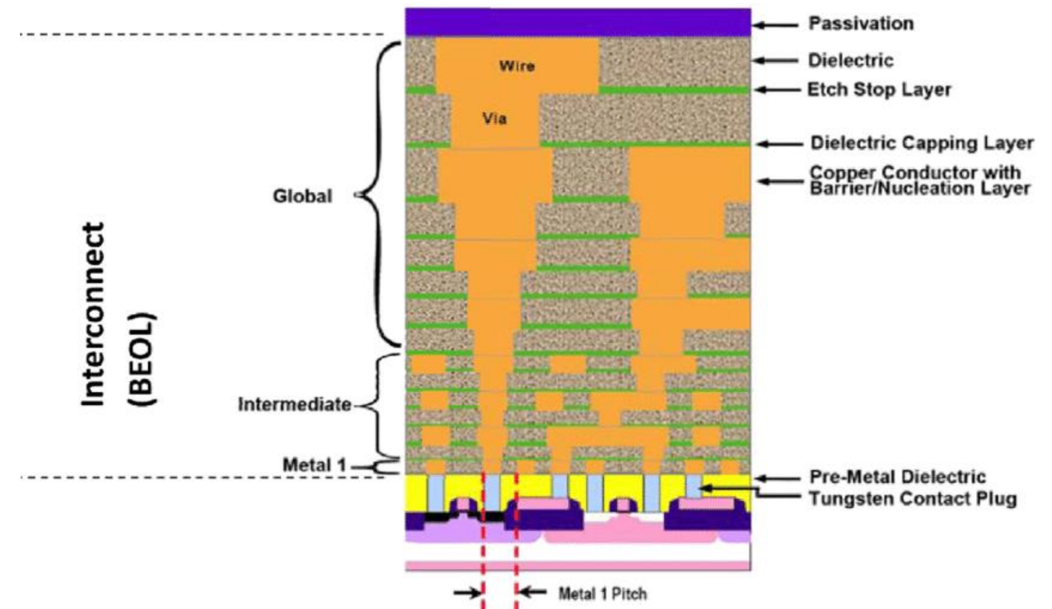


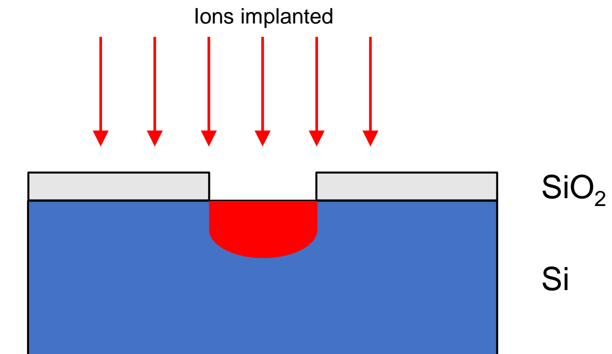
Diagram of a cross-section of a VLSI circuit
from ITRS 2005, http://www.itrs.net/Links/2005itrs/ home2005.htm. 58

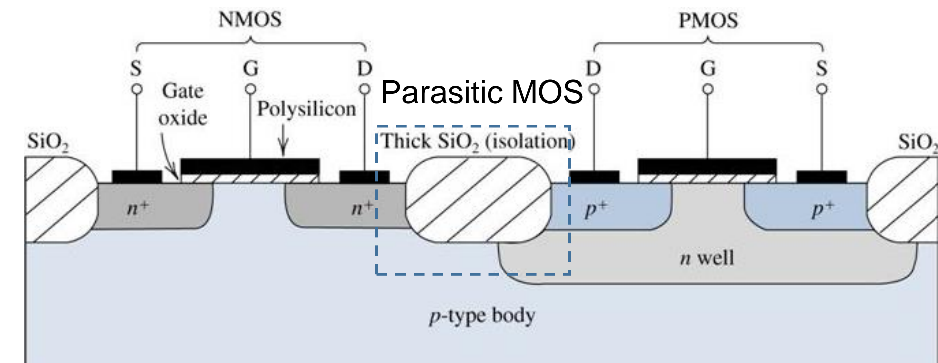
Oxidation process

- The <u>purpose</u> of oxidation is essentially to provide **isolation**:

- Barrier against dopants diffusion/implantation

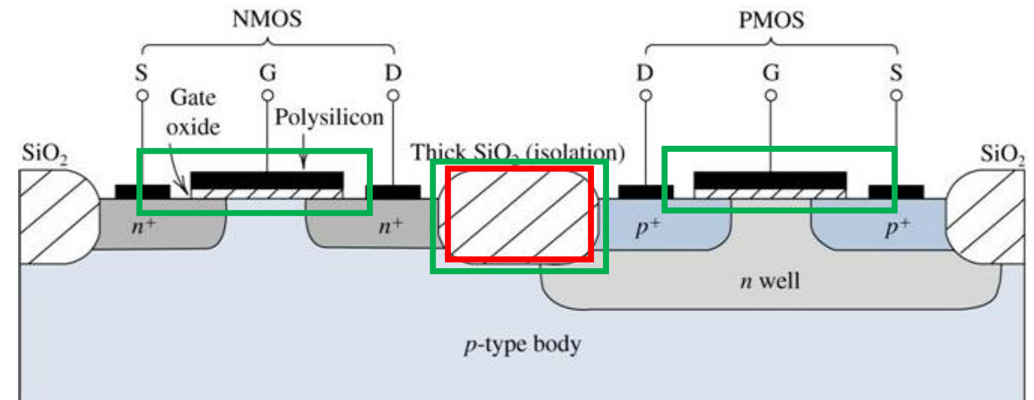
- Electrical insulator between devices (critical field ~ 10^7 V/cm)

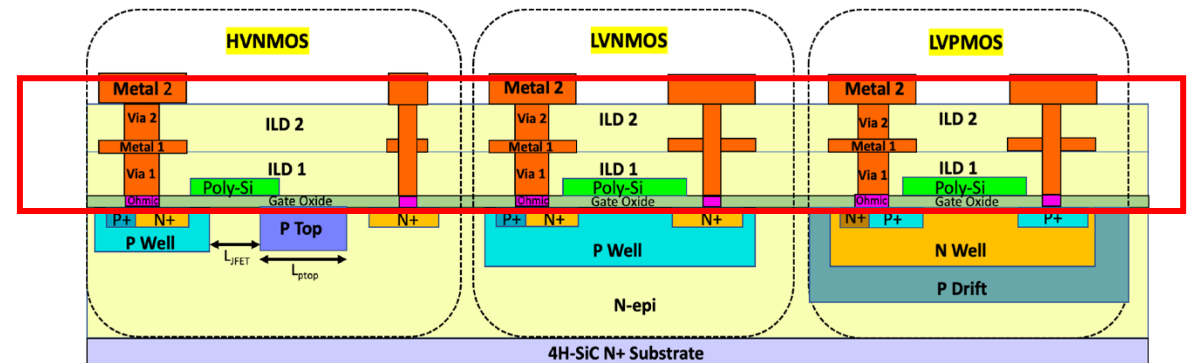




Oxidation process

- Two deposition methods:
  - **Thermal growth**
    - Dry: best quality, slow growth rate: used for gate oxide
    - Wet: lower quality, faster growth rate: used for field oxide (isolation)
  - **Chemical vapor deposition** [3] (lower quality, fast)
    - Used when no Si is available for oxidation (Interlayer Layer Dielectric, ILD)





[3] J. K. Wang, D. R. Denison, Advanced techniques for interlayer dielectric deposition and planarization, Proc. SPIE 2090, https://doi.org/10.1117/12.156535, 1993.

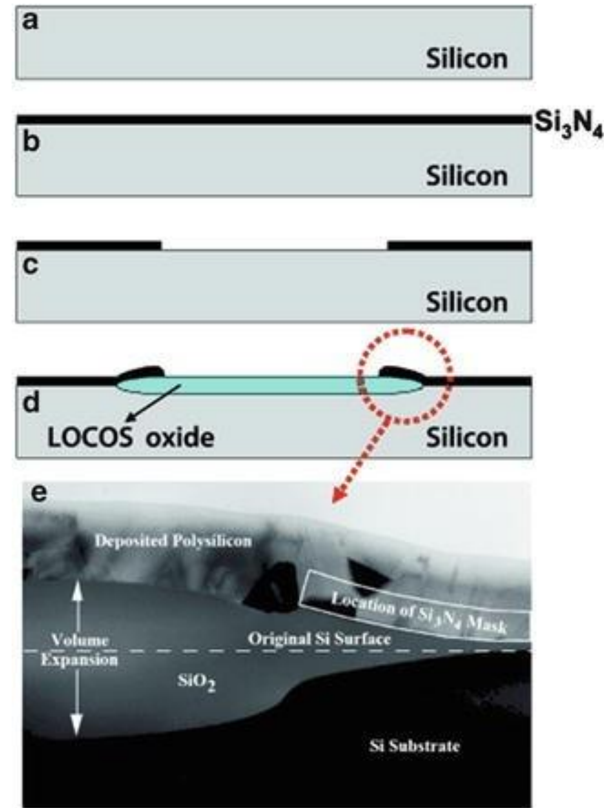
Oxidation process

- **LOCOS** (local Oxidation of Si) process:

  Thermal oxide is grown in etched region: **bird's beak** means Si area loss

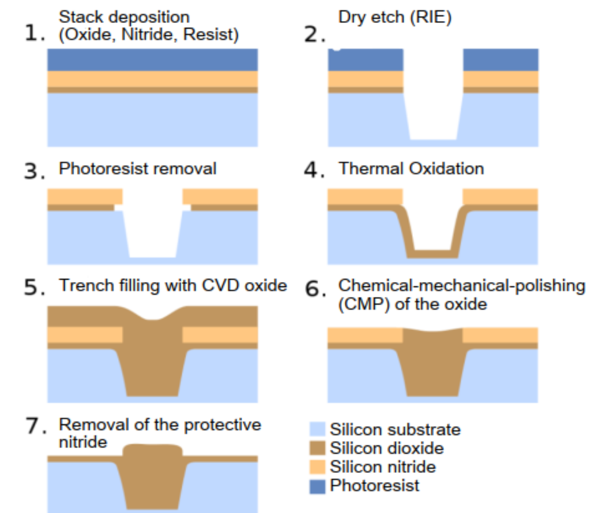
- **STI** (Shallow Trench Isolation) process: (<250 nm)

  CVD oxide is deposited in the etched region: reduced Si area loss



LOCOS process

STI process

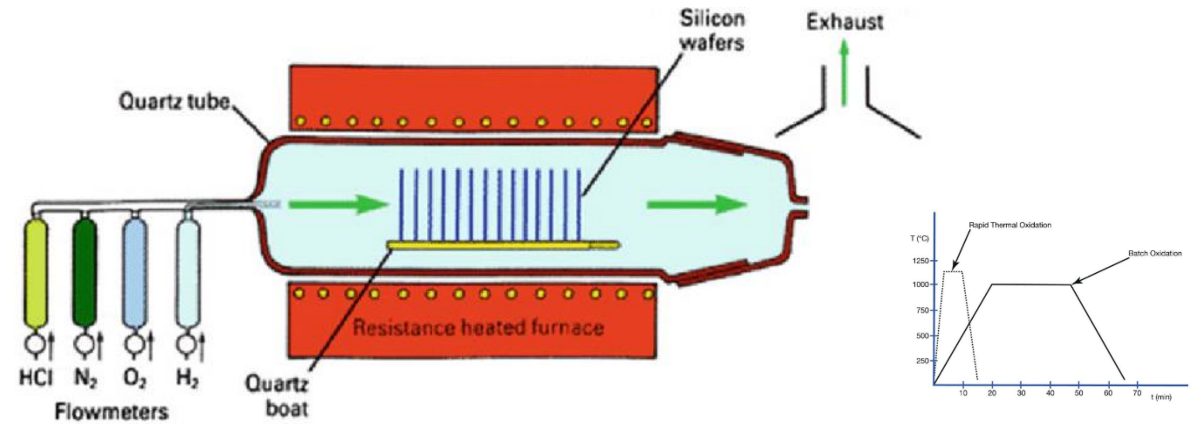


STI fabrication process of modern integrated circuits in cross-sections (from Wikipedia)

Oxidation process tools

- **Thermal oxidation:**
  oxide is grown at high temperature (~ 1000° C) by supplying oxygen that reacts with silicon wafer to form SiO_2 at the surface

- Wafers inserted on a suspended boat into a tubular reactor of quartz, heated by resistance



$$Si + O_2 \rightarrow SiO_2.$$

Dry oxidation: best quality, slow (10 nm/hr)

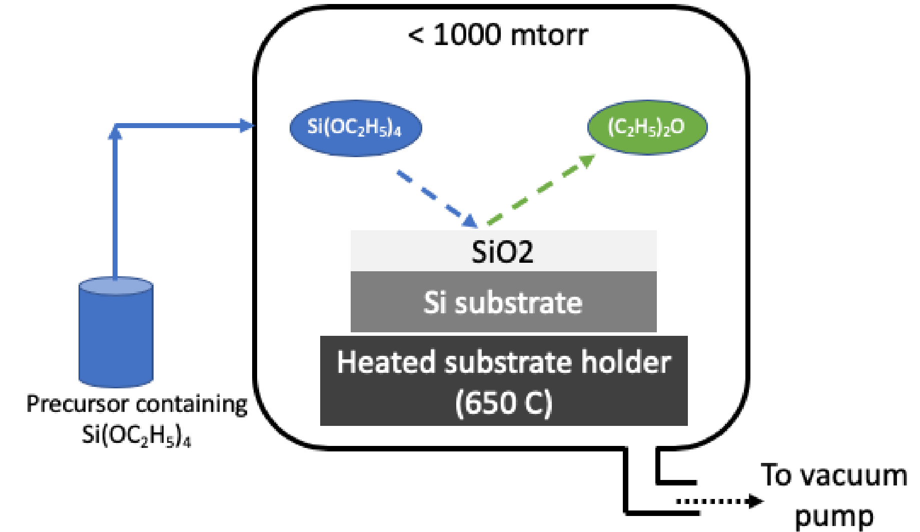
$$Si + 2H_2O \rightarrow SiO_2 + 2H_2.$$

Wet oxidation: lower quality, faster (>x10)

Wet oxidation is faster because H_2O molecules smaller than O_2 leading to faster diffusion through SiO_2

Oxidation process tools

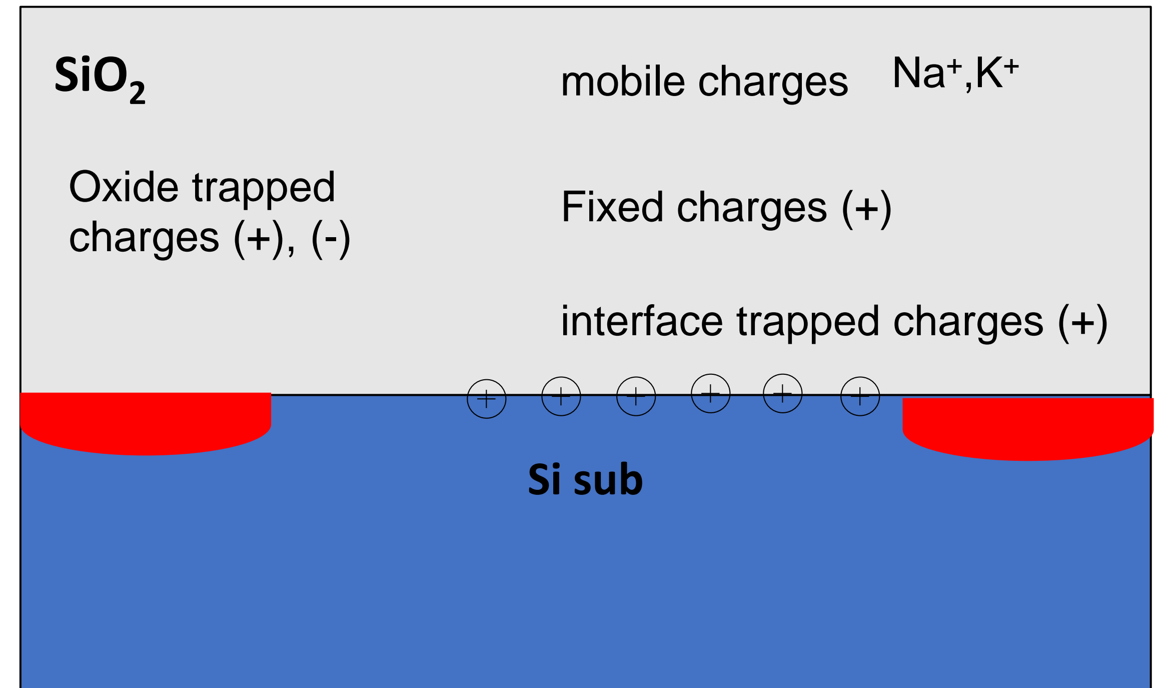
- **Chemical Vapor Deposition (CVD) :** reaction of vapor phase chemicals containing the material to deposit form the solid film on substrate. Reactant gases decompose and form the film.
  - SiO_2 growth rate >x10 compared to thermal oxide, but lower quality
  - CVD is also used for growing dielectric materials of different electrical permittivity (high k/low k material)



Chemical reactions	Techniques
$SiH_4 + O_2 \xrightarrow{430\ °C,\ 1\ bar} SiO_2 + 2H_2$	Silane oxide CVD
$SiH_4 + O_2 \xrightarrow{430\ °C,\ 40\ bar} SiO_2 + 2H_2$	LTO CVD
$SiH_2Cl_2 + 2N_2O \xrightarrow{900\ °C,\ 40\ Pa} SiO_2 + Gas$	HTO CVD
$Si(OC_2H_5)_4 \xrightarrow{700\ °C,\ 40\ bar} SiO_2 + Gas$	TEOS CVD
$Si(OC_2H_5)_4 + O_2 \xrightarrow{400\ °C,\ 0.5\ bar} SiO_2 + Gas$	ACVD
$SiH_4 + 4N_2O \xrightarrow{350\ °C,\ Plasma\ 40\ Pa} SiO_2 + Gas$	PECVD

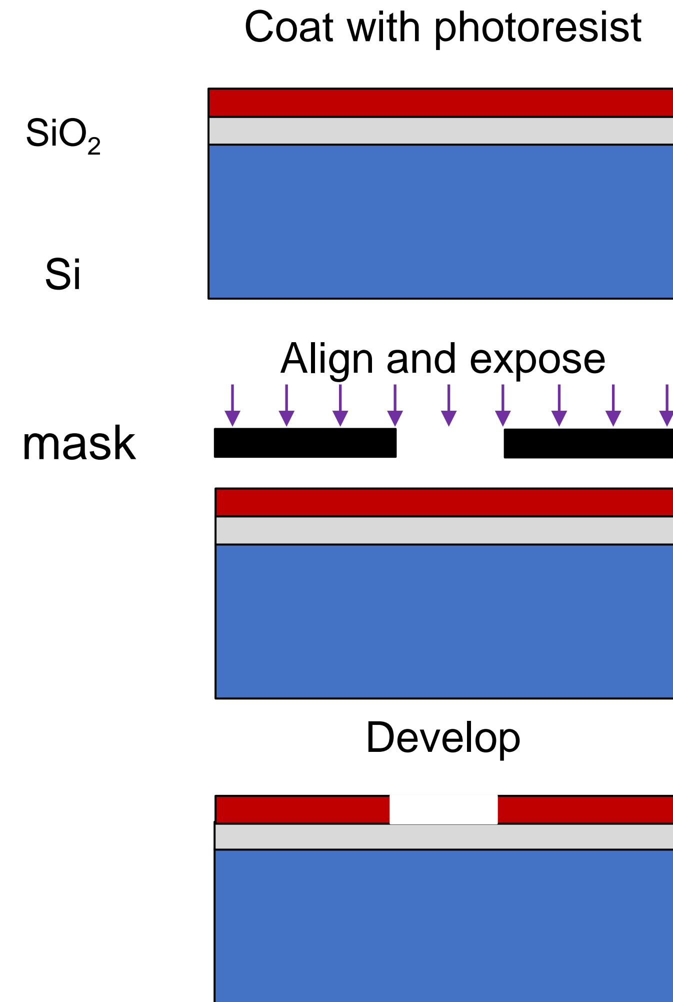
Oxide quality

- Mobile charges (contamination - make insulator conductive)
- Fixed charges (incompletely oxidized Si - create an extra electric field affecting the devices in Si)
- Oxide trapped charge (broken Si-O bonds, due to radiation)
- Interface trapped charges (dangling bonds – affect mobility and increase noise)



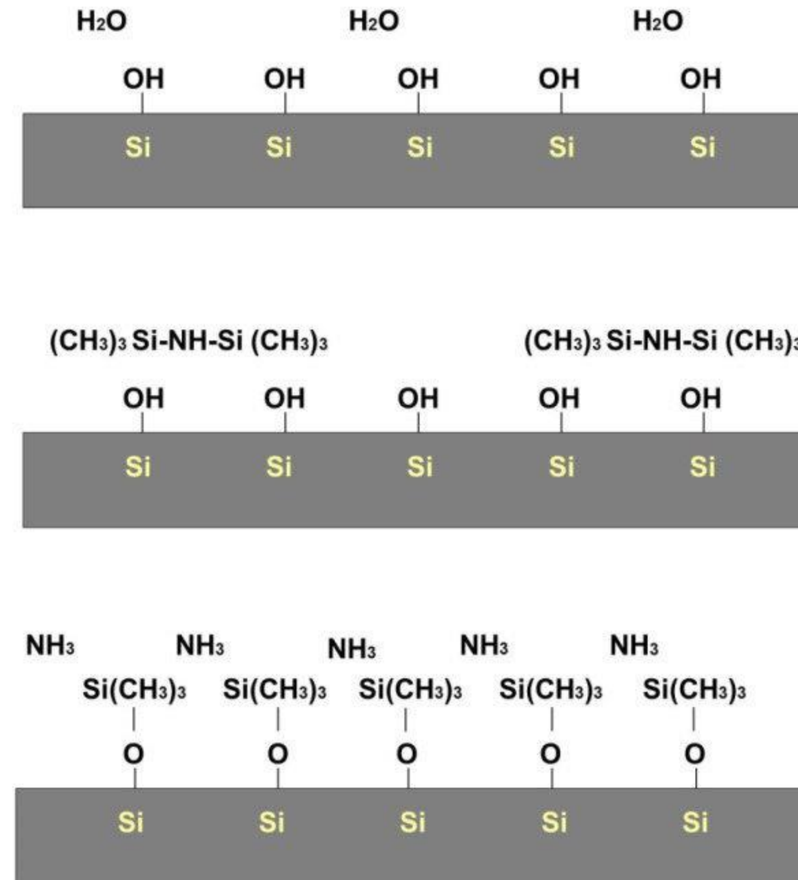
Resist deposition

- The purpose of photoresist deposition is to prepare the wafer for the next step of lithography
- The pattern needs to be reproduced on the photoresist, using a special camera that projects the image of the photomask onto the photoresist
- Photoresist must allow to form a high-res image of the pattern (photo) and be able to stop etching (resist) to transfer the pattern onto the wafer



Resist deposition

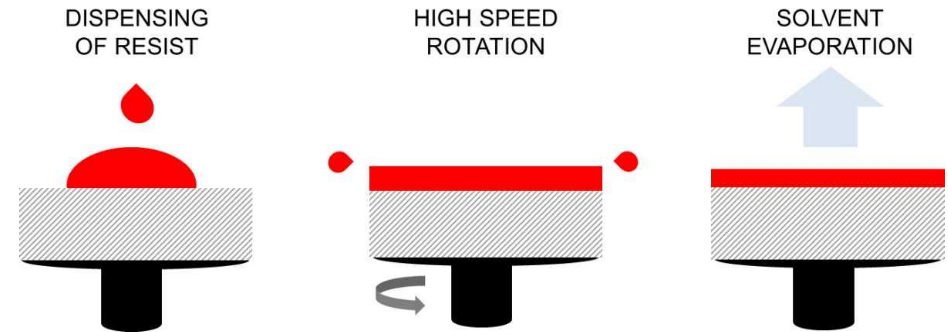
- The wafer must be prepared to improve the adhesion of the photoresist, which is hydrophobic
- Water adheres to the dangling Si bonds at the top of the wafer
- An initial bake to dehydrate the wafer surface is usually performed
- To avoid prompt oxidation, adhesion promoter are often deposited onto the wafer top (**HMDS**, HexaMethylDiSilazane)

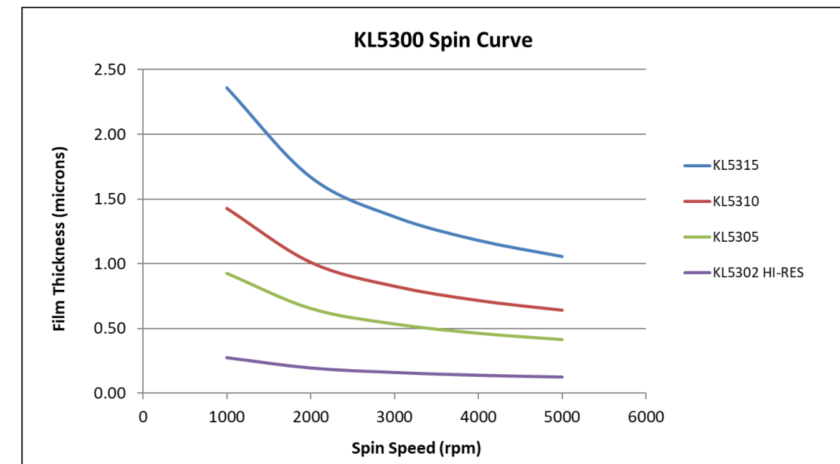


Pics from Dhima, Khalid. (2014). Hybrid lithography –The combination of T-NIL & UV-L.

Resist deposition

- A small amount (~ml) of photoresist, with its solvent, is poured onto the wafer surface
- The wafer is spun at high speeds, the resist spreads out. The air flow increases the viscosity of the photoresists, by evaporating the solvent, which leads to a very uniform photoresist layer spread over the surface
- A post-apply bake is usually performed to further stabilize the film before lithography
- The obtained uniformity is of the order of < 1nm over the entire wafer surface





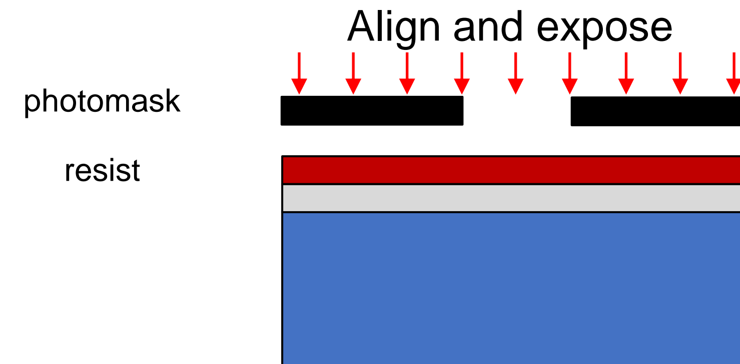
From KemLab website, https://www.kemlab.com/

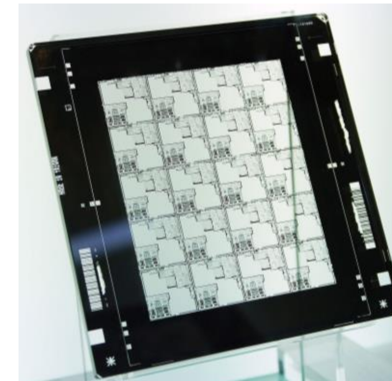
Resist thickness $\sim \frac{1}{\sqrt{\omega}}$

Mask/photolithography

- The photomask, usually made of chrome on quartz substrate, around 5 mm thick, includes the pattern to be reproduced onto the wafer. It can be put in direct contact (contact litho), kept at small distance (proximity litho) or projected onto the wafer (projection litho)

- The photomask is typically bigger (x4-x5) than the wafer pattern



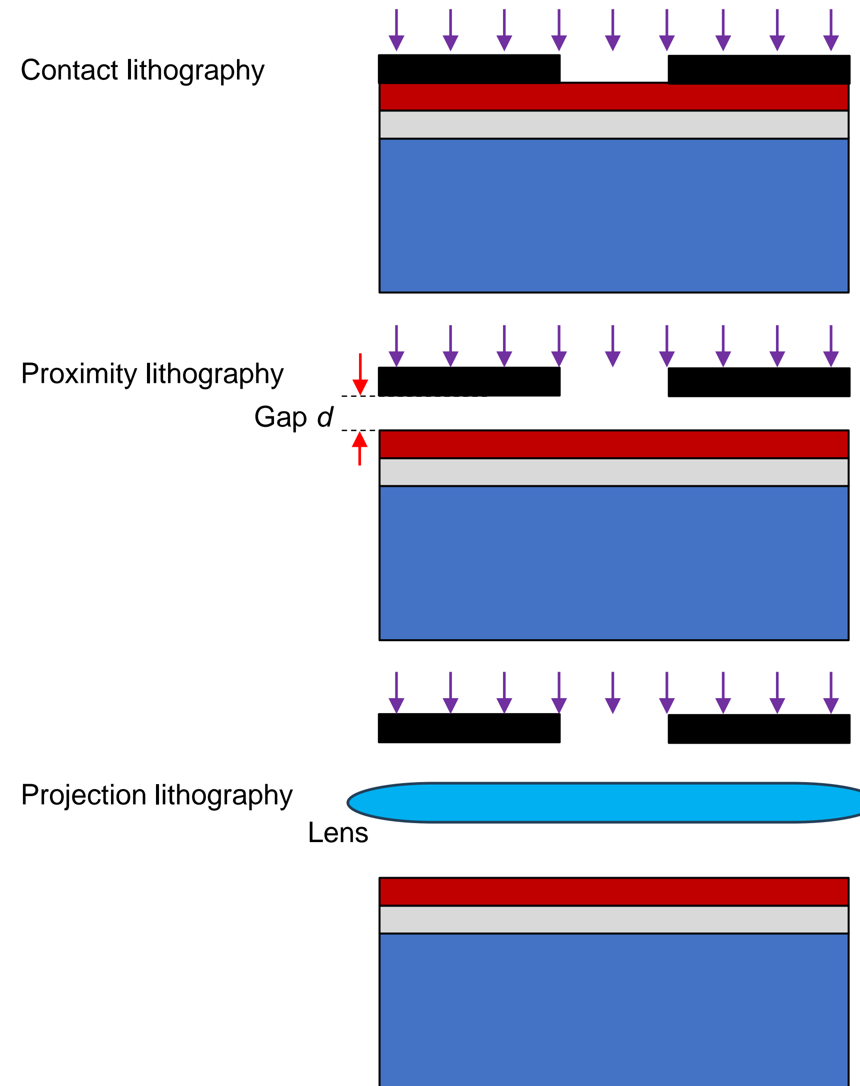


From Wikipedia

The photomask itself is an example of high-resolution photolithography, usually obtained by electron beams to expose a photoresist

Mask/photolithography

- **Contact** printing provides a resolution near the wavelength λ of light used, but every contact damages the mask, that can be used only for a number of times
- **Proximity** printing keeps mask and wafer separated, around 10 um to avoid cumulative defectiveness, but resolution is about $\sqrt{\lambda d}$
- The modern lithography process makes use of **projection** printing, where the mask is illuminated by UV light and the image is projected on the wafer



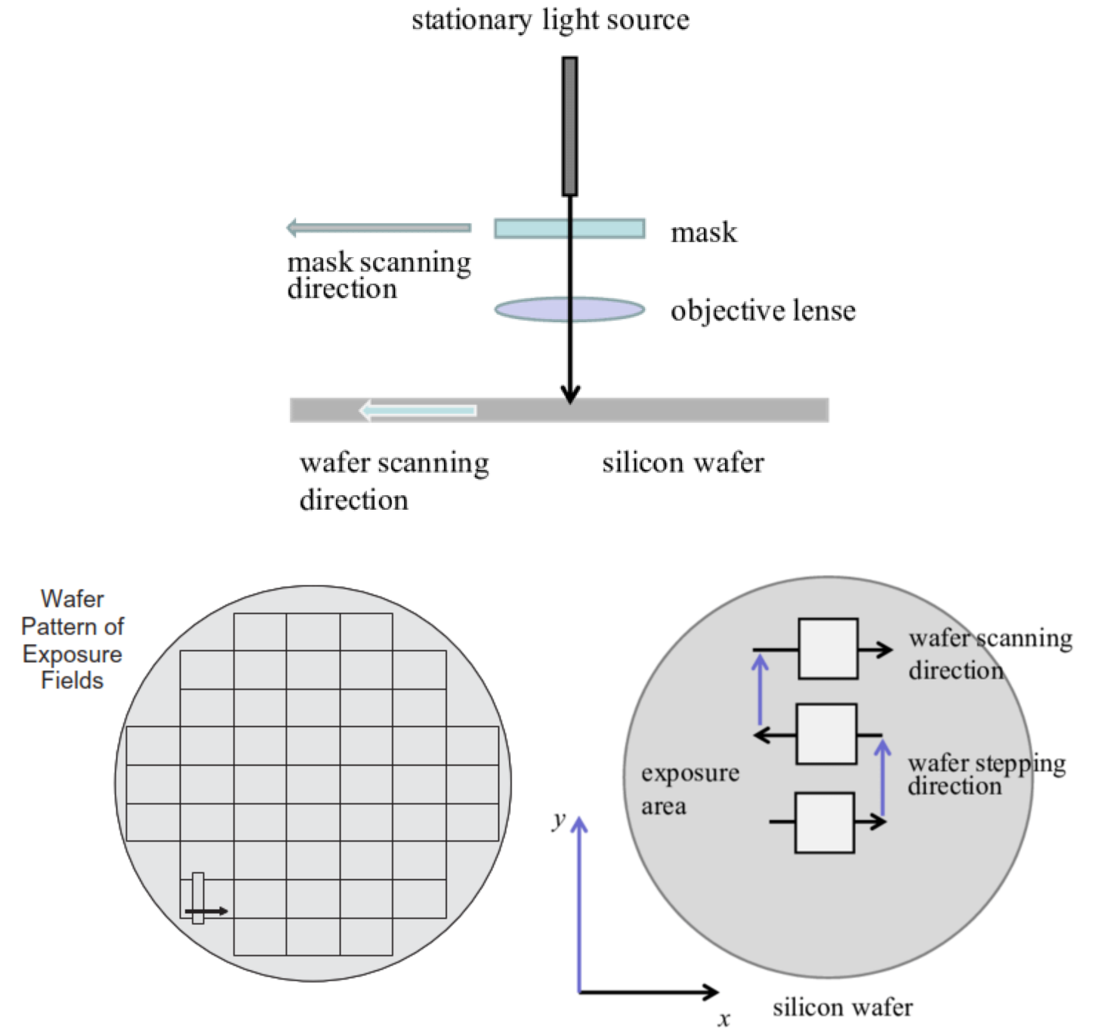
Mask/photolithography

- The cost of lithography process is around 30 % - 50 % of the entire fabrication process and bound to increase for the most modern technology nodes
- Lithography is a gating technology, which leads the advancement in semiconductor industry

From Nikon website

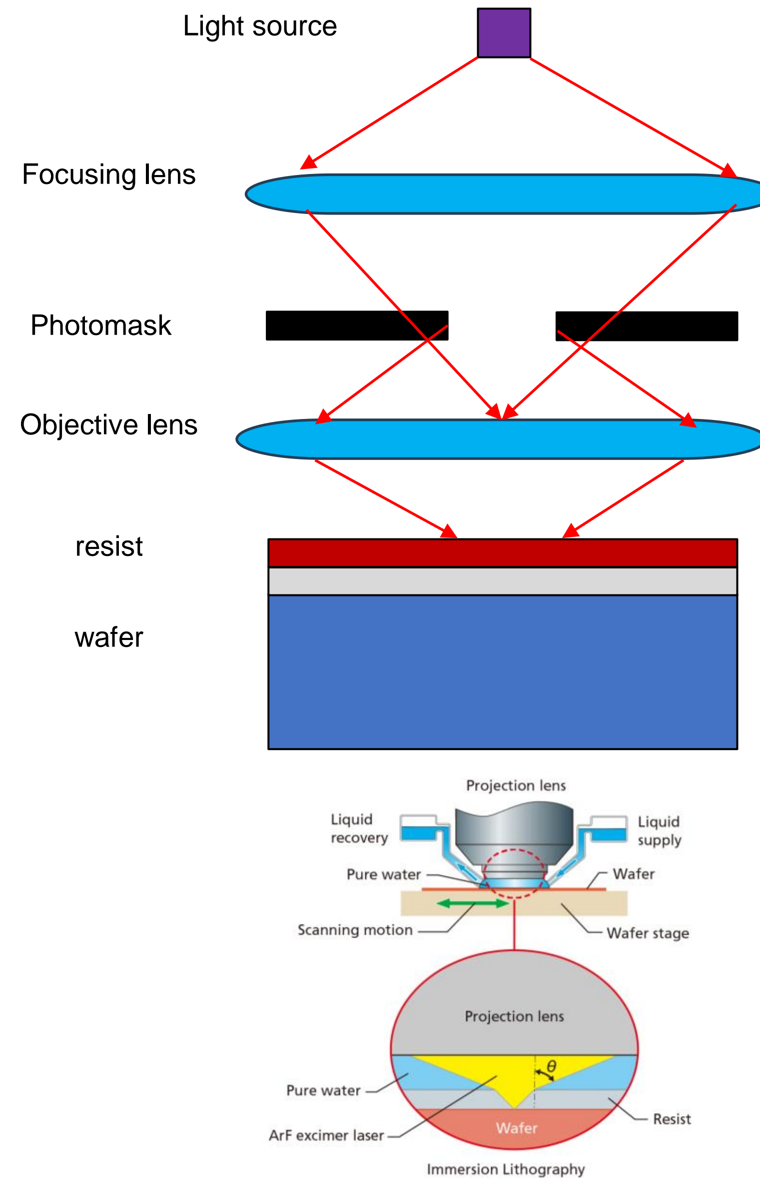
Mask/photolithography

- The patterning of the image is done using a step and scan procedure
- The light goes through a slit and the mask and wafer are scanned across the length of a field (mask pattern)
- Once a field is scanned, the wafer/mask are stepped along the orthogonal direction and the scanning is repeated
- Field size typically 26 x 33 mm^2, slit 25 x 8 mm^2



Mask/photolithography

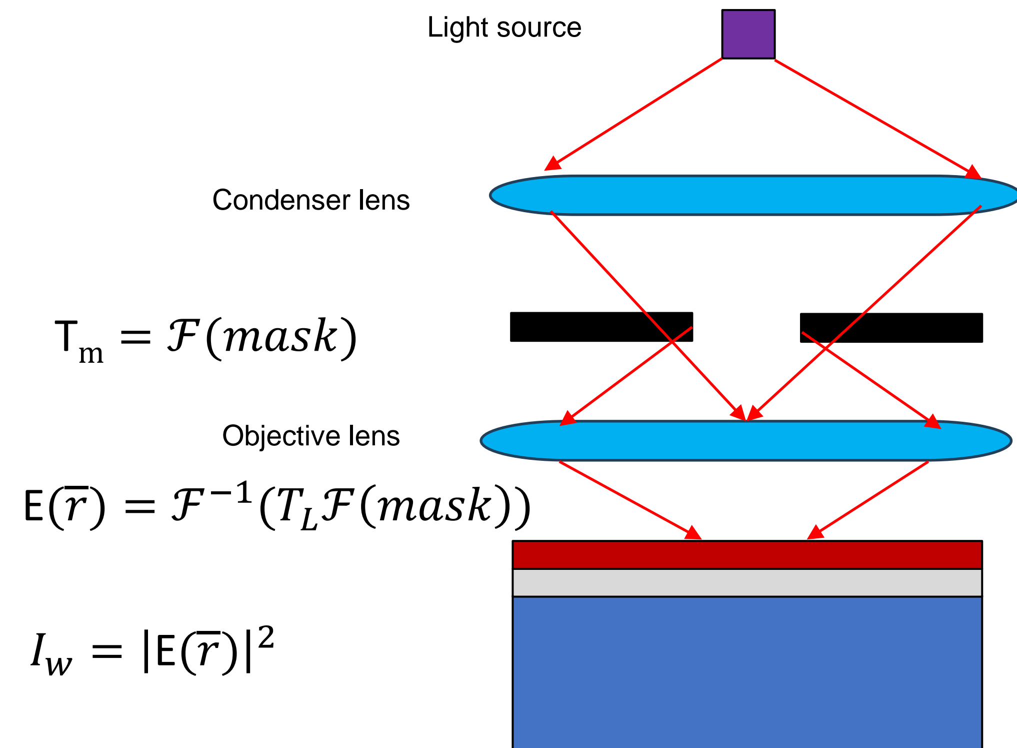
- The projection method used by modern processes consist of a light source and a lens system
- Currently state-of-the-art photolithography [4] uses Deep Ultraviolet (DUV) light λ = 193 nm from an ArF excimer Laser
- The printing of nanometers feature size is possible using **immersion** lithography, where a media of higher refractive index (DI) is used between the lens and the wafer surface



[4] B. Lin, *Optical lithography—present and future challenges*, C. R. Physique 7 (2006) 858–874

Mask/photolithography

- Monochromatic light passing through the mask is diffracted (Fourier transform of the mask function)
- The objective lens produces the Fourier transform of the diffraction pattern entering it
- The Fourier transform of the Fourier transform gives back the 'original' image, i.e. the mask pattern, which is projected onto the wafer
- Due to the practical limitations in the size (hence numerical aperture) of the objective lens, higher frequencies modes of the initial diffraction pattern are cut – off (low pass filter), degrading the final reconstructed pattern on the wafer (resist)
- High-contrast photoresist is used also to enhance the resolution



$$\mathrm{T_m} = \mathcal{F}(mask)$$

$$\mathrm{E}(\overline{r}) = \mathcal{F}^{-1}(T_L \mathcal{F}(mask))$$

$$I_w = |\mathrm{E}(\overline{r})|^2$$

Process fabrication III Summary

Thank you

giulio.villani@stfc.ac.uk

- Introduction to Semiconductor technology
- Differences between Moore's (economic) and Dennard's (physics) law, miniaturization and Power wall

- Process fabrication I, Introduction and typical process flow
- Silicon wafer production
- Oxidation
- Resist deposition
- Photolithography
- [Deal-Groove model]

Appendix :The Deal-Grove oxidation model

- The Deal Grove Model is a simple kinetic model for oxide thermal growth, wet and dry

- Developed by Andy Grove (Intel's CEO) and Bruce Deal in the 60's

JOURNAL OF APPLIED PHYSICS VOLUME 36, NUMBER 12 DECEMBER 1965

General Relationship for the Thermal Oxidation of Silicon

B. E. Deal and A. S. Grove
Fairchild Semiconductor, A Division of Fairchild Camera and Instrument Corporation, Palo Alto, California

(Received 10 May 1965; in final form 9 September 1965)

The thermal-oxidation kinetics of silicon are examined in detail. Based on a simple model of oxidation which takes into account the reactions occurring at the two boundaries of the oxide layer as well as the diffusion process, the general relationship $x_0^2+Ax_0=B(t+\tau)$ is derived. This relationship is shown to be in excellent agreement with oxidation data obtained over a wide range of temperature (700°–1300°C), partial pressure (0.1–1.0 atm) and oxide thickness (300–20 000 Å) for both oxygen and water oxidants. The parameters A, B, and τ are shown to be related to the physico-chemical constants of the oxidation reaction in the predicted manner. Such detailed analysis also leads to further information regarding the nature of the transported species as well as space-charge effects on the initial phase of oxidation.

1. INTRODUCTION

Owing to its great importance in planar silicon-device technology, the formation of silicon dioxide layers by thermal oxidation of single-crystal silicon has been studied very extensively in the past several years.[1–15] Now, with the availability of large amounts of experimental data, it appears that there is much contradiction and many peculiarities in the store of knowledge of silicon oxidation. For instance, reported activation energies of rate constants vary between 27 and 100 kcal/mole for oxidation in dry oxygen; pressure dependence of rate constants has been reported as linear as well as logarithmic. While most of the data on silicon oxidation have been evaluated using the parabolic rate law, certain authors have taken recourse to using empirical power-law dependence,[14] $x_0^n=kt$, where both n and k were complex functions of temperature, pressure, and oxide thickness.

The problems associated with the latter approach can be illustrated by considering Figs 1 and 2. These figures contain a summary of data obtained in these laboratories which are in good general agreement with the corresponding data of Fuller and Strieter[14] and of Evitts, Cooper, and Flaschen.[13] (The experimental methods are dealt with in detail later.) The plots are logarithm of oxide thickness vs the logarithm of oxidation time for dry and wet oxygen (95°C H_2O) at various temperatures. The slope of the lines corresponds to the exponent n in the above power law. These values are indicated at the limiting position of some of the curves. In the case of wet oxygen (Fig. 1), n ranges from 2 for thicker oxides at 1200°C to 1 for the thinner oxide region of the 920°C data. However, for dry oxygen (Fig. 2), the value of n at 1200° approaches 2 as the oxide thickness increases above 1.0 μ; but at lower temperatures and oxide thicknesses the value of n decreases only to about 1.5 and then appears to increase again. Obviously the data cannot be represented by a simple power law.

Most of the previous theoretical treatments of the kinetics of the oxidation of metals emphasize only two limiting types of oxidation mechanisms.[16] In one, the

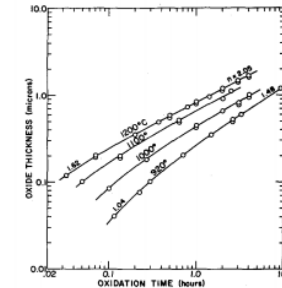
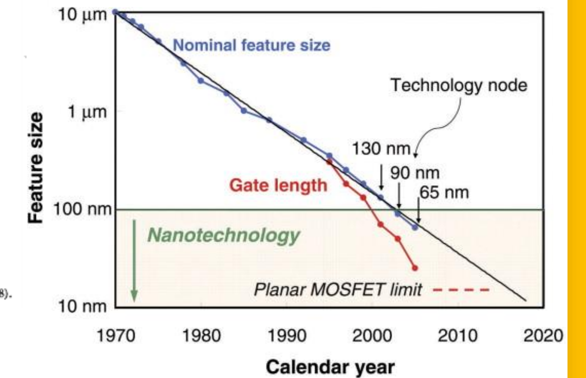


Fig. 1. Oxidation of silicon in wet oxygen (95°C H_2O).

[1] J. T. Law, J. Phys. Chem. **61**, 1200 (1957).
[2] M. M. Atalla, *Properties of Elemental and Compound Semiconductors*, edited by H. Gatos (Interscience Publishers, Inc., New York, 1960), Vol. 5, pp. 163–181.
[3] J. R. Ligenza and W. G. Spitzer, J. Phys. Chem. Solids **14**, 131 (1960).
[4] J. R. Ligenza, J. Phys. Chem. **65**, 2011 (1961).
[5] W. G. Spitzer and J. R. Ligenza, J. Phys. Chem. Solids **17**, 196 (1961).
[6] M. O. Thurston, J. C. C. Tsai, and K. D. Kang, "Diffusion of Impurities into Silicon Through an Oxide Layer," Report 896-Final, Ohio State University, Research Foundation, U. S. Army Signal Supply Agency Contract DA-36-039-SC-83874, March 1961.
[7] P. S. Flint, "The Rates of Oxidation of Silicon," Paper presented at the Spring Meeting of The Electrochemical Society, Abstract No. 94, Los Angeles, 6–10 May 1962.
[8] P. J. Jorgensen, J. Chem. Phys. **37**, 874 (1962).
[9] J. R. Ligenza, J. Electrochem. Soc. **109**, 73 (1962).
[10] B. E. Deal, J. Electrochem. Soc. **110**, 527 (1963).
[11] H. Edagawa, Y. Morita, S. Maekawa, and Y. Inuishi, J. Appl. Phys. (Japan) **2**, 765 (1963).
[12] N. Karube, K. Yamamoto, and M. Kamiyama, J. Appl. Phys. (Japan) **2**, 11 (1963).
[13] H. C. Evitts, H. W. Cooper, and S. S. Flaschen, J. Electrochem. Soc. **111**, 688 (1964).
[14] C. R. Fuller and F. J. Strieter, "Silicon Oxidation," Paper presented at the Spring Meeting of The Electrochemical Society Abstract No. 74, Toronto, 3–7 May 1964.
[15] B. E. Deal and M. Sklar, J. Electrochem. Soc. **112**, 430 (1965).
[16] N. Cabrera and N. F. Mott, Rept. Progr. Phys. **12**, 163 (1948).



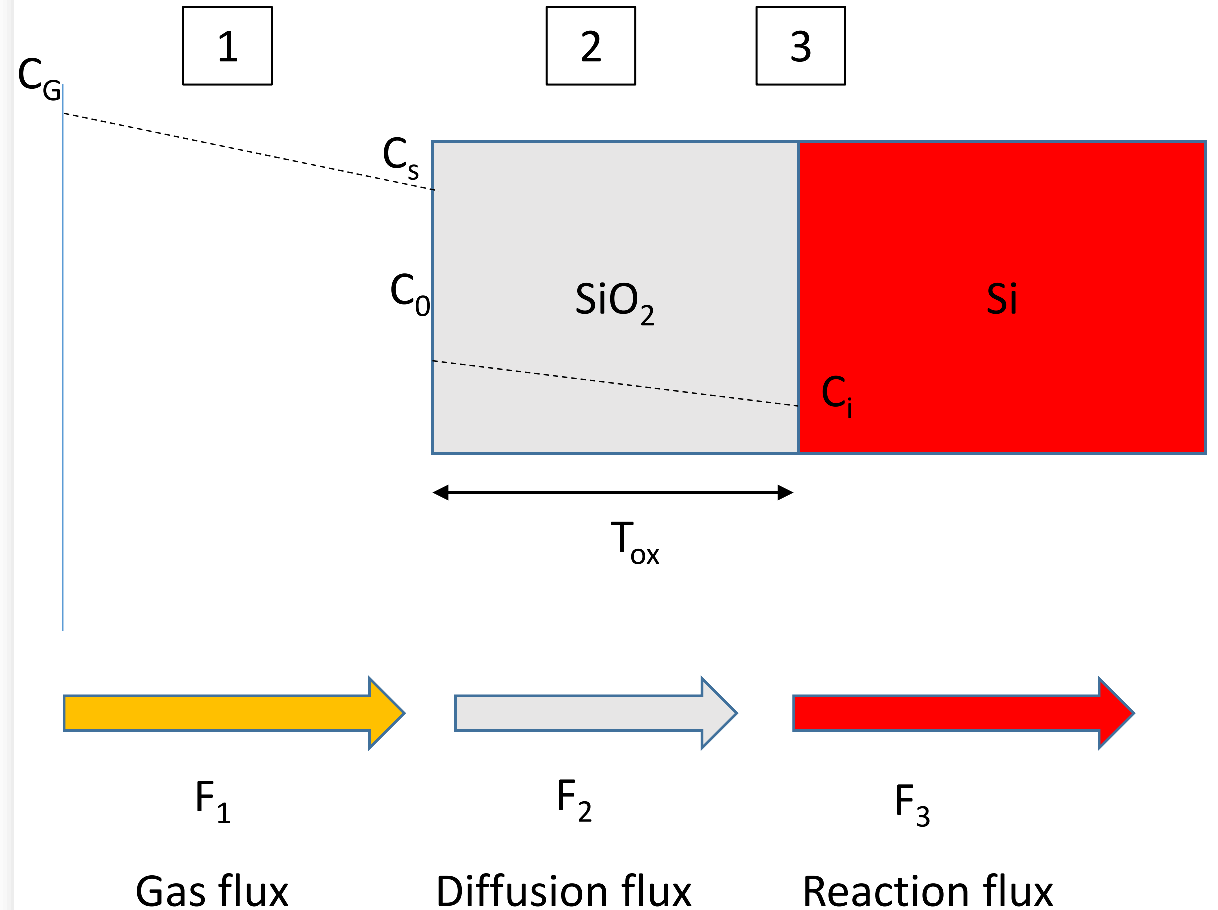
B. E. DEAL AND A. S. GROVE General Relationship for the Thermal Oxidation of Silicon, JOURNAL OF APPLIED PHYSICS VOLUME 36. NUMBER 12 DECEMBER 1965

The Deal-Grove oxidation model

1: Oxygen diffuses from bulk (C_G)to wafer surface (C_S)

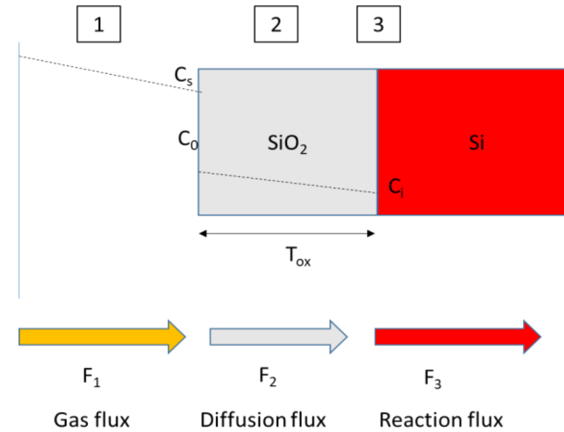
2: Oxygen diffuses from wafer surface (C_O) to Si surface (C_i)

3: Oxygen reacts with Si at interface to form SiO_2



The Deal-Grove oxidation model

- Oxygen diffusion through gas: Fick's 1st Law approximated as linear equation
- Adsorbed concentration C_o on surface ∝ partial pressure (Henry's Law)
- Oxygen diffusion through SiO_2 Fick's 1st Law approximated as linear equation
- 1st order reaction at Si interface



$$\mathbf{1}:\ F_1 = D\frac{dC}{dx} \approx \frac{D_g}{\delta}(C_g - C_s) = h_g(C_g - C_s)$$

$$C_o = HP_s = HC_s kT$$

$$\mathbf{2}:\ F_2 = D\frac{dC}{dx} \approx \frac{D_{ox}}{t_{ox}}(C_o - C_i)$$

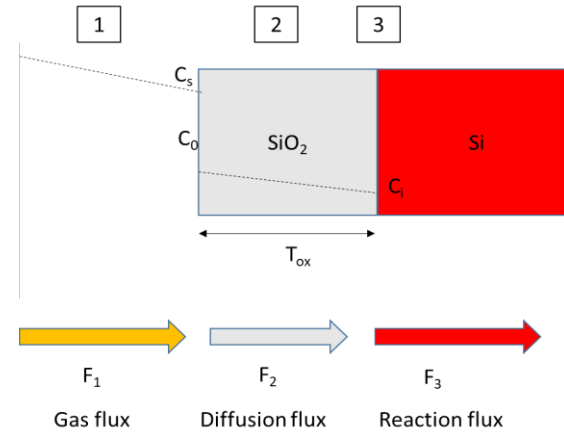
$$\mathbf{3}:\ F_3 = k_s C_i$$

$C_g = \frac{P_g}{kT}$	Reactant concentration
H	Henry's gas law coefficient
t_{ox}	SiO2 thickness
D_{ox}	Oxygen diffusivity in SiO2
h_g	Mass transfer coefficient

The Deal-Grove oxidation model

- Steady state: all fluxes are equal (Si interface reaction is the rate-limiting step)

- Oxygen flux F_{OX} = $v_{ox} N_{ox}$



$$F_1 = F_2 = F_3 = F_{OX}$$

$$F_{OX} = \frac{H k_s P_g}{1 + \frac{k_s}{h} + \frac{k_s t_{ox}}{D_{ox}}} = \frac{d t_{ox}}{dt} N_{ox} \quad ; \; h = \frac{h_g}{HkT}$$

$$\frac{d t_{ox}}{dt} = \frac{H k_s P_g}{1 + \frac{k_s}{h} + \frac{k_s t_{ox}}{D_{ox}}} \frac{1}{N_{ox}}$$

UNIVERSITY OF OXFORD

The Deal-Grove oxidation model

- Integrating over t gives the expression for oxide thickness t_{ox} vs. time t
- The model requires coefficients adjustments for different crystal orientations

$$t_{ox}^2 + At_{ox} = B(t + \tau)$$

$$A = 2D_{ox}\left(\frac{1}{h} + \frac{1}{k_s}\right)$$

$$B = \left(\frac{2D_{ox}HP_g}{N_{ox}}\right)$$

$$\tau = \left(\frac{t_0^2 + At_0}{B}\right)$$

Table 4.1 Oxidation coefficients for silicon

Temperature (°C)	Dry			Wet (640 torr)	
	A (μm)	B (μm²/hr)	τ (hr)	A (μm)	B (μm²/hr)
800	0.370	0.0011	9	—	—
920	0.235	0.0049	1.4	0.50	0.203
1000	0.165	0.0117	0.37	0.226	0.287
1100	0.090	0.027	0.076	0.11	0.510
1200	0.040	0.045	0.027	0.05	0.720

The τ parameter is used to compensate for the rapid growth regime for thin oxides. *(After Deal and Grove.)*

<111> Si

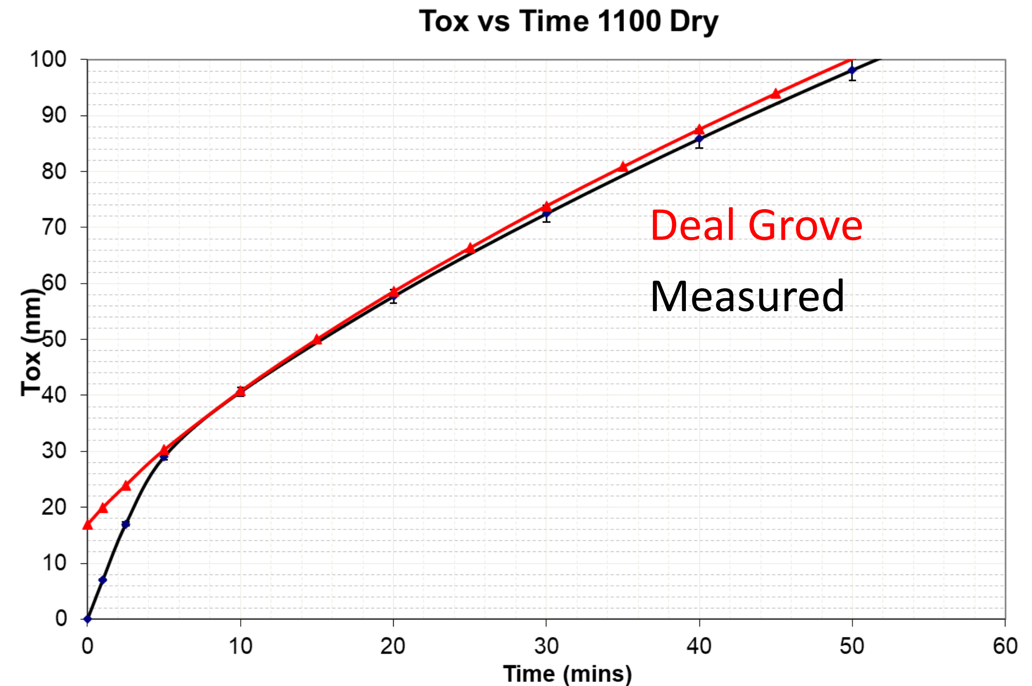
The Deal-Grove oxidation model

- Comparison of Deal-Grove model with <u>measured</u> oxide thickness (Dry oxide)

- With the correct coefficients (P,T, crystal) the Deal-Grove model works nicely for single crystal silicon (~%'s accuracy). Model extensions to 3D exist

- Additional tweaking needed for high Si doping

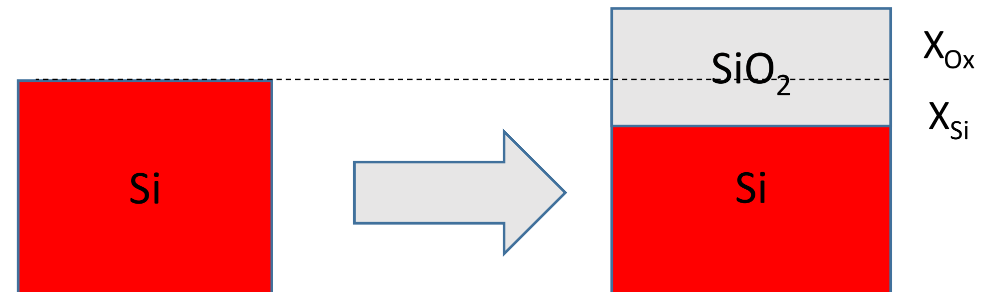
- The model fails for Polysilicon



* Self-limiting growth of native oxide saturates at around 2-3 nm
'**Growth of native oxide on a silicon surface'** Journal of Applied Physics, Volume 68, Issue 3, August 1, 1990, pp.1272-1281

The Deal-Grove oxidation model

- The oxide grows at the expense of Silicon: during oxidation around half of Silicon is consumed



$$X_{Si}\,N_{Si} = X_{Ox}\,N_{Ox} \rightarrow X_{Si} = \frac{X_{Ox}\,N_{Ox}}{N_{Si}} \equiv X_{Ox}\frac{2.3\;10^{22}}{5\;10^{22}} = X_{Ox}\,0.46$$

$N_{Ox} = molecular\ density\ SiO_2$
$N_{Si} = atomic\ density\ Si$