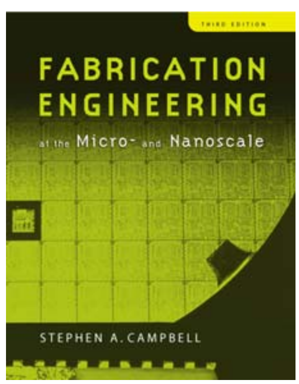
ECE 416/516
IC Technologies

Lecture 16:
Back-End Processing
& Manufacturing

Professor James E. Morris
Spring 2012

Chapter 15

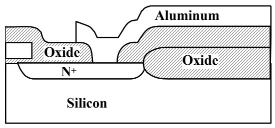
Device Isolation, Contacts, and Metallization



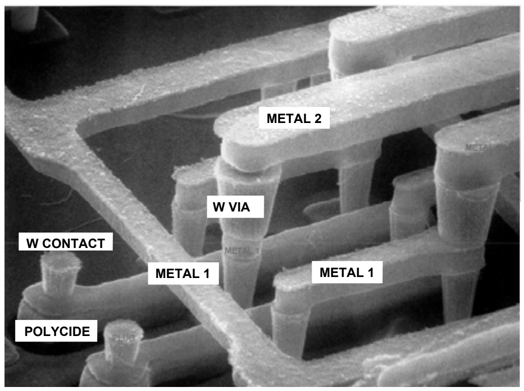
Fabrication Engineering at the Micro- and Nanoscale   Campbell

BACKEND TECHNOLOGY

Introduction



- Backend technology: fabrication of interconnects and the dielectrics that electrically isolate them.
- Early structures were simple by today's standards.

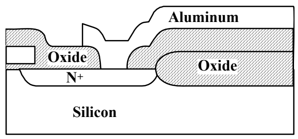


- More metal interconnect levels increases circuit functionality and speed.
- Interconnects are separated into local interconnects (polysilicon, silicides, TiN) and intermediate/ global interconnects (Cu or Al).
- Backend processing is becoming more important.
- Larger fraction of total structure and processing.
- Starting to dominate total speed of circuit.

Year of Production	1998	2000	2002	2004	2007	2010	2013	2016	2018
Technology Node (half pitch)	250 nm	180 nm	130 nm	90 nm	65 nm	45 nm	32 nm	22 nm	18 nm
MPU Printed Gate Length		100 nm	70 nm	53 nm	35 nm	25 nm	18 nm	13 nm	10 nm
Min Metal 1 Pitch (nm)				214	152	108	76	54	42
Wiring Levels - Logic				10	11	12	12	14	14
Metal 1 Aspect Ratio (Cu)				1.7	1.7	1.8	1.9	2.0	2.0
Contact Aspect Ratio (DRAM)				15	16	>20	>20	>20	>20
STI Trench Aspect Ratio				4.8	5.9	7.9	10.3	14	16.4
Metal Resistivity (μohm-cm)	3.3, 2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2
Interlevel Dielectric Constant	3.9	3.7	3.7	<2.7	<2.4	<2.1	<1.9	<1.7	<1.7

- More sophisticated analysis from the 2003 ITRS interconnect roadmap.
- Global interconnects dominate the RC delays.
- “In the long term, new design or technology solutions (such as co-planar waveguides, free space RF, optical interconnect) will be needed to overcome the performance limitations of traditional interconnect.” (ITRS)

A. Contacts



- **Early structures were simple Al/Si contacts.**
- **Highly doped silicon regions are necessary to insure ohmic, low resistance contacts.**

$$\rho_c = \rho_{co} \exp\left(\frac{2\phi_B\sqrt{m^*\varepsilon_s}}{\hbar\sqrt{N_D}}\right)$$

- **Tunneling current through a Schottky barrier depends on the width of the barrier and hence N_D.**
- **In practice, N_D, N_A > 10^{20} are required.**

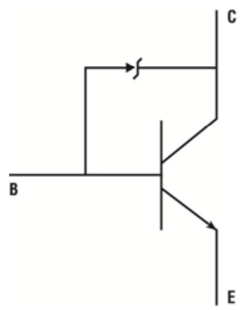
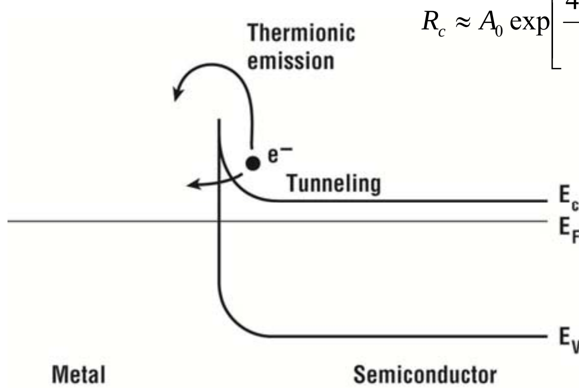


Figure 15.17 Schottky shunted bipolar transistor used for nonsaturating bipolar logic.

5

$$R_c \approx A_0 \exp\left[\frac{4\pi\phi_b\sqrt{m_n^* k_{Si}\varepsilon_0}}{h\sqrt{N_D}}\right]$$



Al

Si

Figure 15.22 Two carrier transport mechanisms typically found in metal semiconductor contacts.

Fabrication Engineering at the Micro- and Nanoscale Campbell

$$I = I_0\left(\exp\frac{eV}{kT} - 1\right) \text{ where } I_0 = RT^2 A \exp - \frac{\phi_b}{kT}$$

$$\text{so } J_S = A^* T^2 \exp - \frac{\phi_b}{kT}\left(\exp\frac{eV}{kT} - 1\right) \text{ and}$$

$$\text{specific contact resistance } R_c = \left(\frac{\partial J}{\partial V}\right)^{-1}_{V=0}$$

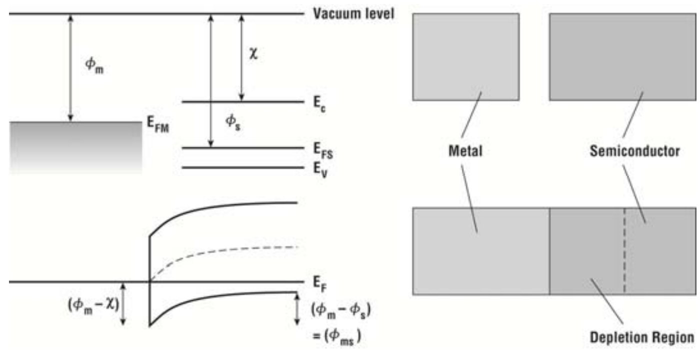


Figure 15.18 Band diagram for an ideal Schottky contact: before contact (right) and after contact (left).

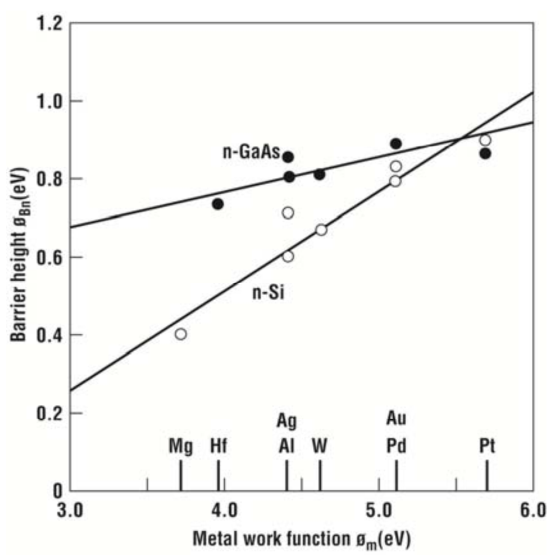
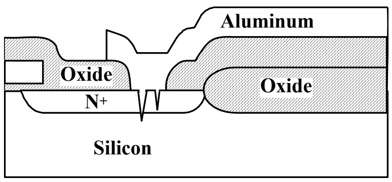
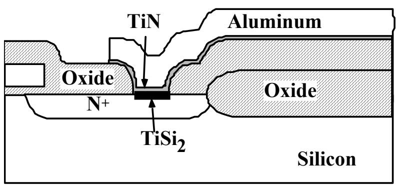


Figure 15.19 Experimentally measured Schottky barrier heights for silicon and GaAs *(from Sze, reprinted by permission, Wiley).*

• Another practical issue is that Si is soluble in Al (≈ 0.5% at 450˚C). This can lead to "spiking" problems.



• Si diffuses into Al, voids form, Al fills voids ⇒ shorts!
• 1st solution - add 1-2% Si in Al to satisfy solubility. Widely used, but Si can precipitate when cooling down and increase ρ_c.



• Better solution: use barrier layer(s). Ti or $TiSi_2$ for good contact and adhesion, TiN for barrier.

Si dissolves in Al at high T

0.5% @ 450°C

1.0% @ 525°C

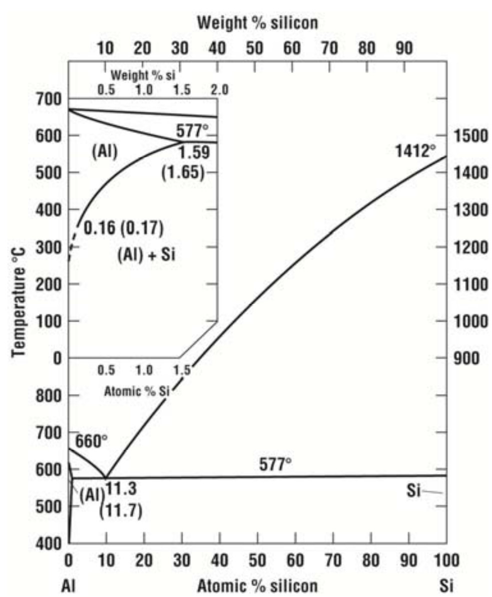


Figure 15.24 Phase diagram of Al/Si. Inset shows the low concentration region.

High T sinter:

Si dissolves in Al to saturation

Si diffuses away along Al conductor

More Si can dissolve

Voids

Spiking

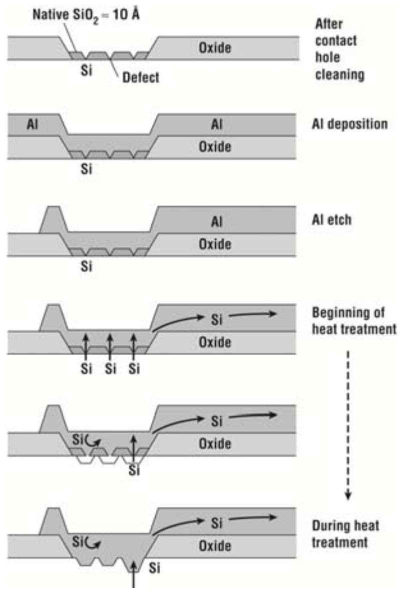


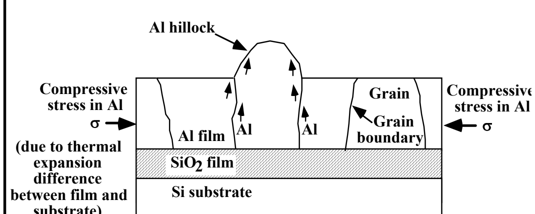
Figure 15.25 Cross-sectional diagrams of Al on Si contact formation process *(after Wolf [1], reprinted by permission, Lattice Press).*

Fabrication Engineering at the Micro- and Nanoscale   Campbell

B. Interconnects And Vias

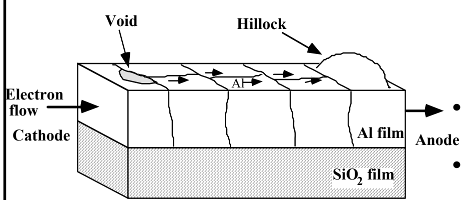
- **Al has historically been the dominant material for interconnects.**
  - **- low resistivity**
  - **- adheres well to Si and SiO_2**
  - **- can reduce other oxides**
  - **- can be etched and deposited easily**

- **Problems:**
  - **-relatively low melting point and soft.**
  - **-need a higher melting point material for gate electrode and local interconnect ⇒ polysilicon.**
  - **- hillocks and voids easily formed in Al.**



- **Hillocks and voids form because of stress and diffusion in Al films. Heating places Al under compression causing hillocks. Cooling back down can place Al under tension ⇒ voids.**
- **Adding a few % Cu stabilizes grain boundaries and minimizes hillock formation.**

12



- A related problem with Al interconnects is "electromigration." High current density (0.1-0.5 MA/cm^2) causes movement of Al atoms in direction of electron flow.
- Can cause hillocks and voids, leading to shorts or opens in the circuit.
- Adding Cu (0.5-4 weight %) can also inhibit electromigration.
- Thus Al is commonly deposited with 1-2 wt % Si and 0.5-4 wt % Cu.

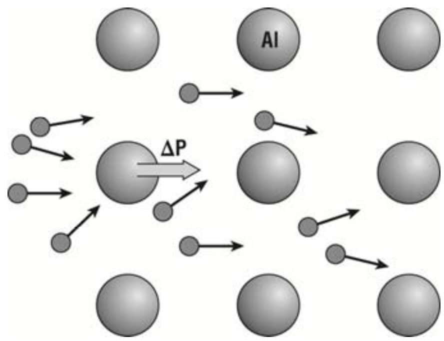


Figure 15.30 Momentum transfer between electrons and matrix atoms in an interconnect material.

13

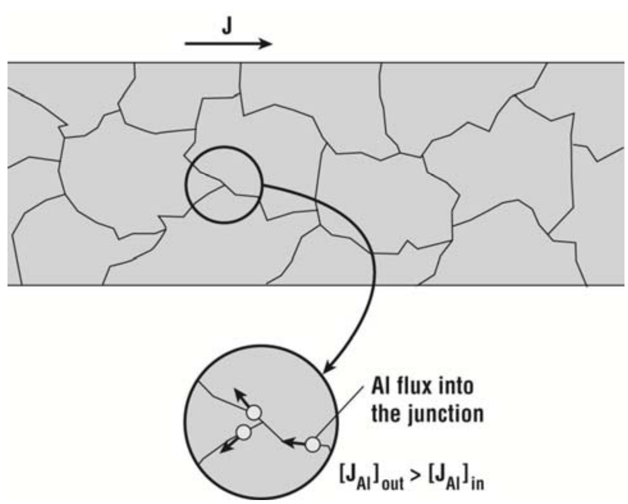
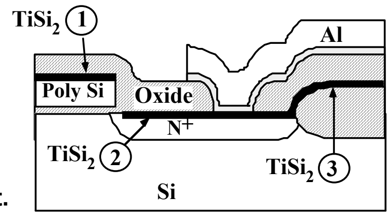
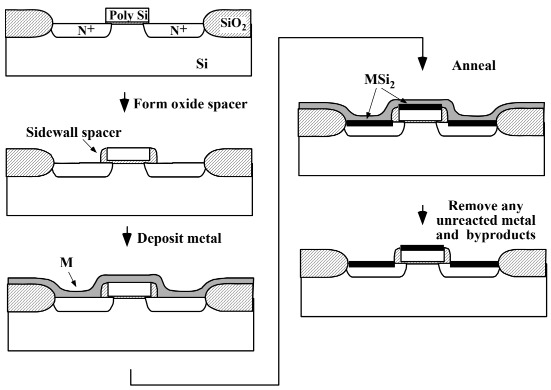


Figure 15.31 Typical point of flux divergence in an interconnect material.

- Next development was use of other materials with lower resistivity as local interconnects, like TiN and silicides.
- Silicides used to:
  1. strap polysilicon,
  2. strap junctions,
  3. as a local interconnect.



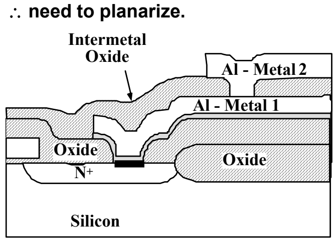


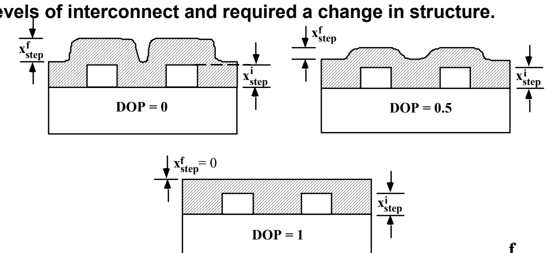
- Self-aligned silicide ("salicide") process.
- Also, simultaneous TiN, $TiSi_2$ formation in CMOS process.

15

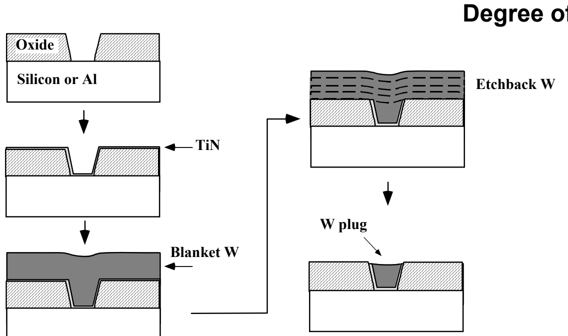
- Early two-level metal structure (early 1980's). Non-planar topography leads to lithography, deposition, filling issues.
- These issues get worse with additional levels of interconnect and required a change in structure.

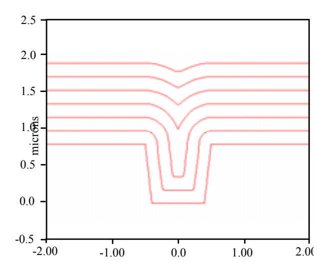
∴ need to planarize.





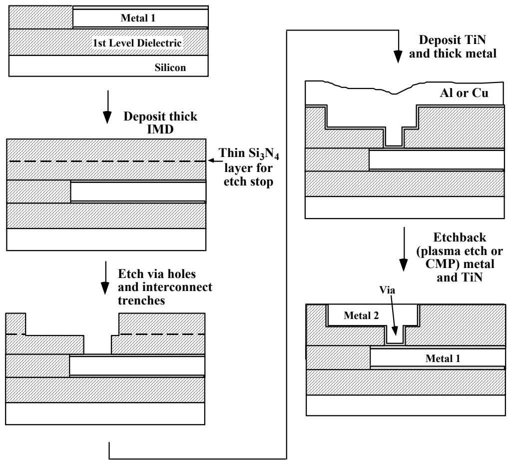
Degree of planarization is $\text{DOP} = 1 - \dfrac{x_{step}^{f}}{x_{step}^{i}}$



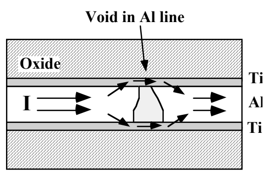


- One early approach to planarization incorporated W plugs and a simple etchback process. (Damascene process.) SPEEDIE simulation.

16



- More advanced version of the damascene process provides both the via/contact and interconnect levels simultaneously.
- In this "dual damascene" process, both the openings in the IMD for the metal interconnect and for the contact or vias underneath are opened, one after the other.
- Metal is then deposited into both layers at once followed by a CMP etchback.



- Interconnects have also become multilayer structures.
- Shunting the Al helps mitigate electromigration and can provide mechanical strength, better adhesion and barriers in multi-level structures. TiN on top also acts as antireflection coating for lithography.

IMD: Inter-metal dielectric

17

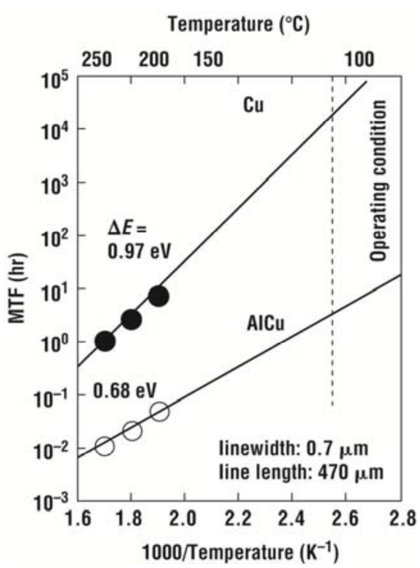
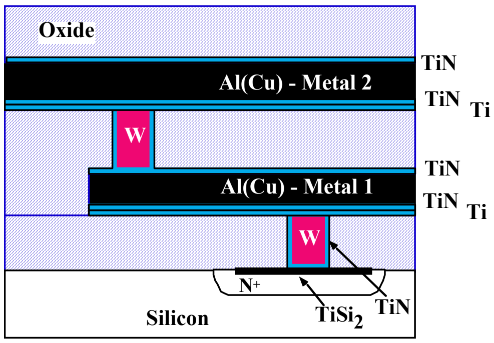


Figure 15.36 Median time to failure of Cu and AlCu lines (*from Sun, reprinted with permission IEEE*).

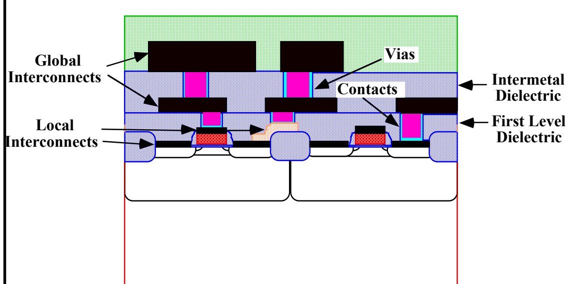
- The biggest change that has occurred in the past 5 years is the widespread introduction of Cu, replacing aluminum.
- Cu cannot be easily etched since the byproducts, copper halides are not volatile room temperature.
- Electroplating plus a damascene process (single or dual) is the obvious solution and is widely used today.
- Cu is the dominant material in logic chips today (μp, ASICs), but not in most memory chips.



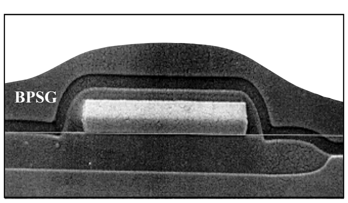
- Typical modern interconnect structure incorporating all these new features.

19

C. Dielectrics



- Dielectrics electrically and physically separate interconnects from each other and from active regions.
- Two types:
  - First level dielectric
  - Intermetal dielectric (IMD)



- First level dielectric is usually SiO_2 “doped” with P or B or both (2-8 wt. %) to enhance reflow properties.
- PSG: phosphosilicate glass, reflows at 950-1100˚C
- BPSG: borophosphosilicate glass, reflows at 800˚C.
- SEM shows BPSG oxide layer after 800˚C reflow step, showing smooth topography over step.
- Undoped SiO_2 often used above and below PSG or BPSG to prevent corrosion of Al .

20

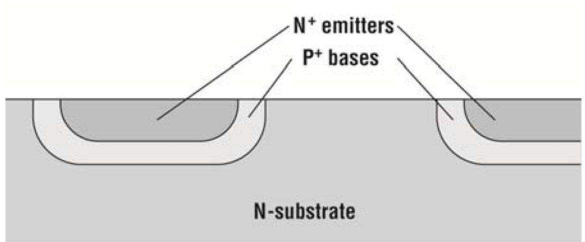


Figure 15.1 Simple junction isolation in a bipolar transistor technology with a common collector.

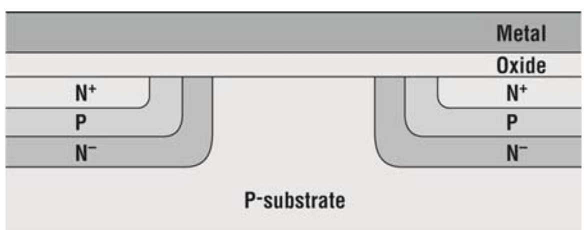


Figure 15.3 Cross section of simple bipolar technology with a metal line crossing the junction isolation region, forming a parasitic MOSFET.

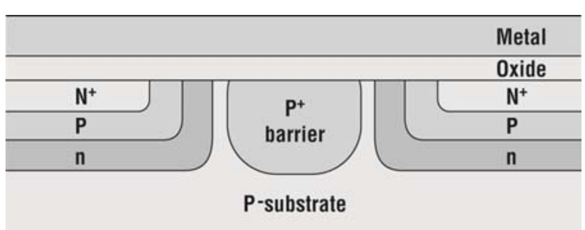


Figure 15.5 Guard ring isolation for the bipolar technology from Figure 15.3.

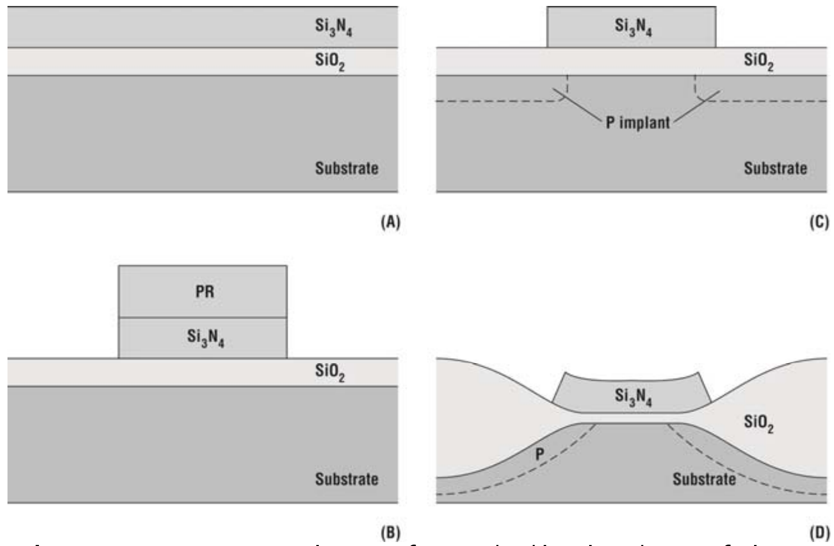


Figure 15.6 Cross-sectional views of a standard local oxidation of silicon (LOCOS) process.

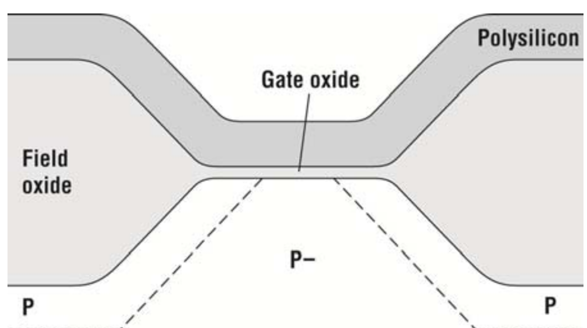


Figure 15.8 Cross-sectional schematic of an MOS transistor cut along its width, illustrating the origin of narrow channel effects.

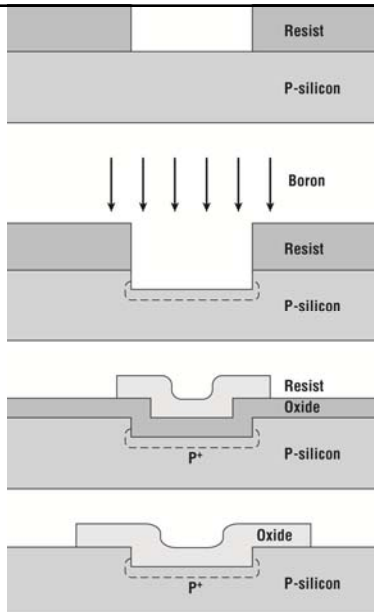
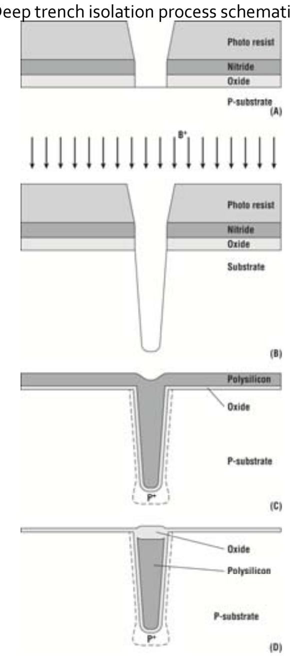


Figure 15.9 Simple shallow trench isolation process.

Figure 15.10
Deep trench isolation process schematic.



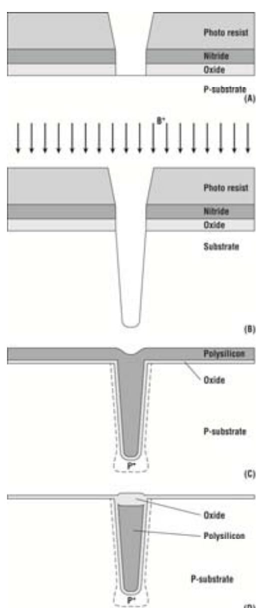


Figure 15.10 Deep trench isolation process schematic.

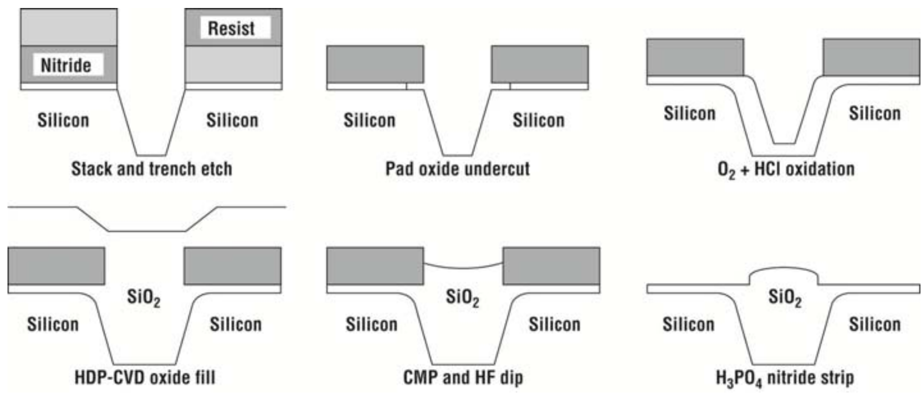
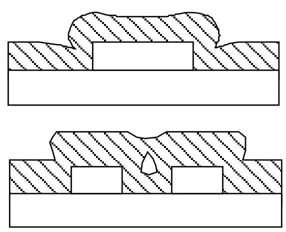
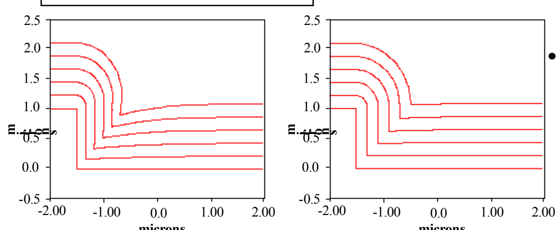


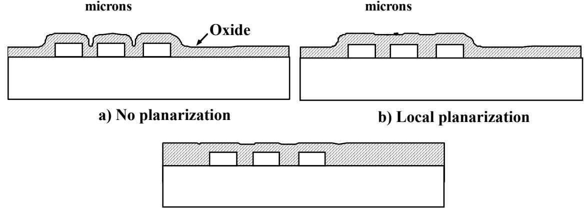
Figure 15.12 Schematic of a shallow trench isolation module *(after Chaterjee et al., used with permission, APS, 1997).*



- Intermetal dielectrics also made primarily of SiO_2 today, but cannot do reflow or densification anneals on pure SiO_2 because of T limitations.
- Two common problems occur, cusping and voids, which can be minimized using appropriate deposition techniques.



- SPEEDIE simulations of silicon dioxide depositions over a step for silane deposition (S_c = 0.4) and TEOS deposition (S_c = 0.1) showing less cusping in the latter case.

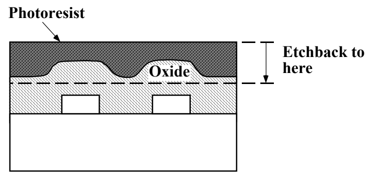


a) No planarization

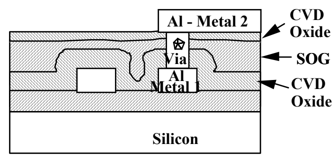
b) Local planarization

c) Global planarization

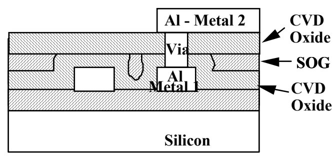
- However planarization is also usually required today.



- One simple process involves planarizing with photoresist and then etching back with no selectivity.



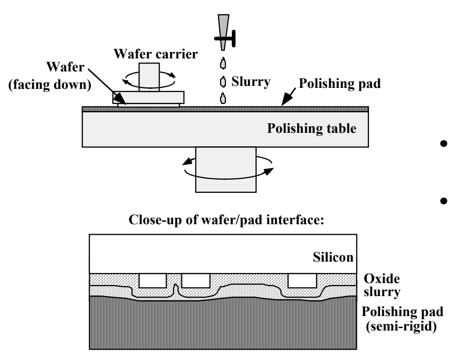
without etchback



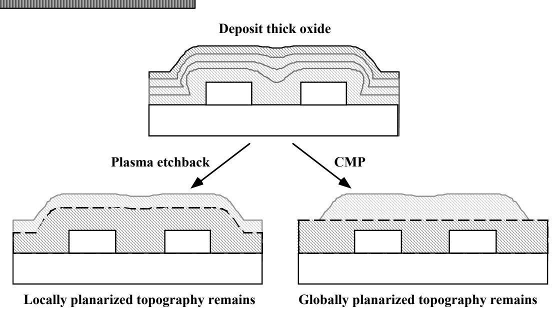
with etchback

- Spin-on-glass (SOG) is another option:
  - Fills like liquid photoresist, but becomes SiO_2 after bake and cure.
  - Done with or without etchback (with etchback to prevent poisoned via - no SOG contact with metal).
  - Can also use low-K SOD's. (spin-on-dielectrics)
  - SOG oxides not as good quality as thermal or CVD oxides
  - Use sandwich layers.
- A final deposition option is HDPCVD which provides angle dependent sputtering during deposition which helps to planarize.

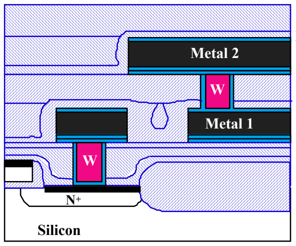
29



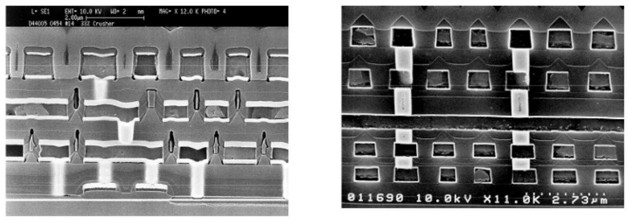
- The most common solution today is CMP which works very well.
- It is capable of forming very flat surfaces as shown in the example below.



30



• Backend structure showing one possible dielectric multi-structure scheme. Other variations include HDP oxide or the use of CMP.



• Two backend structures. Left: three metal levels and encapsulated BPSG for the first level dielectric; SOG (encapsulated top and bottom with PECVD oxide) and CMP in the intermetal dielectrics. The multilayer metal layers and W plugs are also clearly seen. Right: five metal levels, HDP oxide (with PECVD oxide on top) and CMP in the intermetal dielectrics.

31

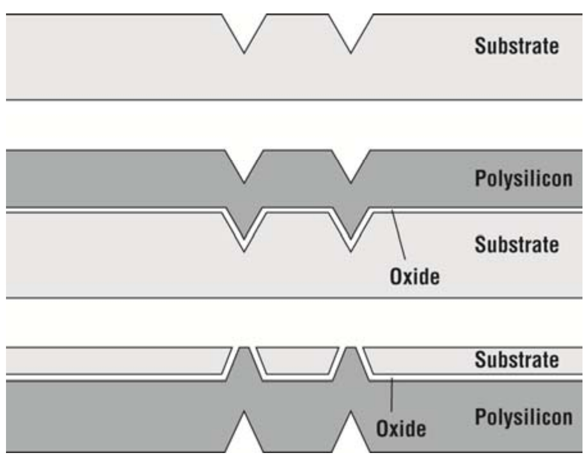


Figure 15.13 Dielectric isolation (DI) process for forming silicon on insulator.

Fabrication Engineering at the Micro- and Nanoscale Campbell

THE FUTURE OF BACKEND TECHNOLOGY

- **Remember:** $$\tau_L = 0.89RC = 0.89 \cdot K_I K_{ox} \varepsilon_0 \rho L^2 \left(\frac{1}{Hx_{ox}} + \frac{1}{WL_S} \right) \quad (1)$$

Year of Production	1998	2000	2002	2004	2007	2010	2013	2016	2018
Technology Node (half pitch)	250 nm	180 nm	130 nm	90 nm	65 nm	45 nm	32 nm	22 nm	18 nm
MPU Printed Gate Length		100 nm	70 nm	53 nm	35 nm	25 nm	18 nm	13 nm	10 nm
Min Metal 1 Pitch (nm)				214	152	108	76	54	42
Wiring Levels - Logic				10	11	12	12	14	14
Metal 1 Aspect Ratio (Cu)				1.7	1.7	1.8	1.9	2.0	2.0
Contact Aspect Ratio (DRAM)				15	16	>20	>20	>20	>20
STI Trench Aspect Ratio				4.8	5.9	7.9	10.3	14	16.4
Metal Resistivity (μohm-cm)	3.3, 2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2
Interlevel Dielectric Constant	3.9	3.7	3.7	<2.7	<2.4	<2.1	<1.9	<1.7	<1.7

- **Reduce metal resistivity - use Cu instead of Al.**
- **Aspect ratio - advanced deposition, etching and planarization methods.**
- **Reduce dielectric constant - use low-K materials.**

33

Material class	Material	Dielectric constant	Deposition technique
Inorganic	SiO_2 (including PSG and BPSG)	3.9-5.0	CVD/Thermal ox./Bias-sputtering/HDP
	Spin-on-glass (SiO_2) (including PSG, BPSG)	3.9-5.0	SOD
	Modified SiO_2 (e.g. fluorinated SiO_2 or hydrogen silsesquioxane - HSQ)	2.8-3.8	CVD/SOD
	BN (Si)	>2.9	CVD
	Si_3N_4 (only used in multilayer structure)	5.8-6.1	CVD
Organic	Polyimides	2.9-3.9	SOD/CVD
	Fluorinated polyimides	2.3-2.8	SOD/CVD
	Fluoro-polymers	1.8-2.2	SOD/CVD
	F-doped amorphous C	2.0-2.5	CVD
Inorganic/Organic Hybrids	Si-O-C hybrid polymers based on organo-silsesquioxanes (e.g. MSQ)	2.0-3.8	SOD
Aerogels (Microporous)	Porous SiO_2 (with tiny free space regions)	1.2-1.8	SOD
Air bridge		1.0-1.2	

- **All of these approaches are beginning to appear in advanced process flows today.**

34

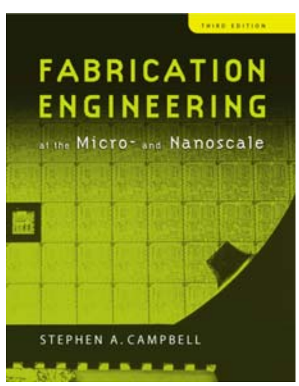
Summary of Key Ideas

- Backend processing (interconnects and dielectrics) have taken on increased importance in recent years.
- Interconnect delays now contribute a significant component to overall circuit performance in many applications.
- Early backend structures utilized simple Al to silicon contacts.
- Reliability issues, the need for many levels of interconnect and planarization issues have led to much more complex structures today involving multilayer metals and dielectrics.
- CMP is the most common planarization technique today.
- Copper and low-K dielectrics are now found in some advanced chips and their use will likely be common in the future.
- Beyond these materials changes, interconnect options in the future include architectural (design) approaches to minimizing wire lengths, optical interconnects, electrical repeaters and RF broadcasting. All of these areas will see significant research in the next few years.

35

Chapter 20

Integrated Circuit Manufacturing



Fabrication Engineering at the Micro- and Nanoscale   Campbell

Lecture Topics & Objectives

- Contamination, Defects & Distributions
- Yield
  - uniform defects
  - non-uniform defects
  - multiple defect types
- Reliability & Failure
  - distributions and plotting
  - target failure rates
  - accelerated testing
- Objective: Can use reliability statistics with various models

Cause of Defects

Faulty circuits by:

- Processing faults:
- thickness variations, oxide, poly-Si
- residues
- Circuit tolerances may be exceeded by device variation by minor processing variations.
- Dirt etc:
  - during lithography or metallization etc.
  - may have missing parts of circuit.
  - importance of clean room, water, etc.

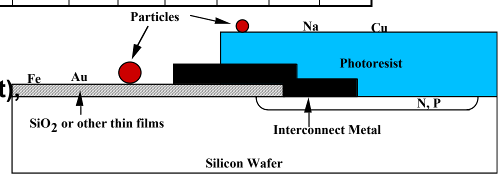
SEMICONDUCTOR MANUFACTURING: YIELD, CLEAN ROOMS, WAFER CLEANING, GETTERING, SPC, DOE, CIM

- **Modern IC factories employ a three tiered approach to controlling unwanted impurities: 1. clean factories 2. wafer cleaning 3. gettering**

Year of Production	1998	2000	2002	2004	2007	2010	2013	2016	2018
Technology Node (half pitch)	250 nm	180 nm	130 nm	90 nm	65 nm	45 nm	32 nm	22 nm	18 nm
MPU Printed Gate Length		100 nm	70 nm	53 nm	35 nm	25 nm	18 nm	13 nm	10 nm
DRAM Bits/Chip (Sampling)	256M	512M	1G	4G	16G	32G	64G	128G	128G
MPU Transistors/Chip ($x10^6$)				550	1100	2200	4400	8800	14,000
Critical Defect Size	125 nm	90 nm	90 nm	90 nm	90 nm	90 nm	65 nm	45 nm	45 nm
Starting Wafer Particles (cm^{-2})				<0.35	<0.18	<0.09	<0.09	<0.05	<0.05
Starting Wafer Total Bulk Fe (cm^{-3})	$3x10^{10}$	$1x10^{10}$	$1x10^{10}$	$1x10^{10}$	$1x10^{10}$	$1x10^{10}$	$1x10^{10}$	$1x10^{10}$	$1x10^{10}$
Metal Atoms on Wafer Surface After Cleaning (cm^{-2})	$5x10^9$	$1x10^{10}$	$1x10^{10}$	$1x10^{10}$	$1x10^{10}$	$1x10^{10}$	$1x10^{10}$	$1x10^{10}$	$1x10^{10}$
Particles on Wafer Surface After Cleaning (#/wafer)				75	80	86	195	106	168

2003 ITRS Front End processes

- **Contaminants may consist of particles, organic films (photoresist), heavy metals or alkali ions.**



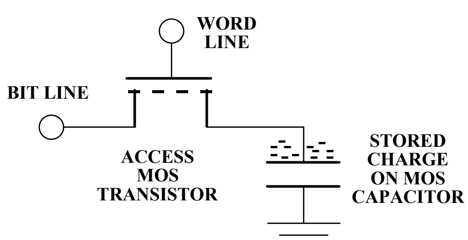
39

Contamination

Example #1: MOS V_{TH} is given by $$V_{TH} = V_{FB} + 2\phi_f + \frac{\sqrt{2\varepsilon_S qN_A(2\phi_f)}}{C_O} + \frac{qQ_M}{C_O} \quad (1)$$

If t_{ox} = 10 nm, then a 0.1 volt V_{th} shift can be caused by Q_M = 6.5 x 10^{11} cm^{-2} (< 0.1% monolayer or 10 ppm in the oxide).

Example #2: MOS DRAM



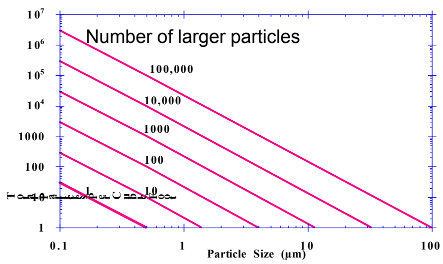
- **Refresh time of several msec requires a generation lifetime of**

$$\tau = \frac{1}{\sigma v_{th} N_t} \geq 100\ \mu sec \quad (2)$$

- **This requires trap density $N_t \geq 10^{12}$ cm^{-3} or $\leq$ 0.02 ppb.**
- **Traps typically Au, Fe, Cu impurities**

40

Level 1 Contamination Reduction: Clean Factories



- Air quality is measured by the "class" of the facility.

(Photo courtesy of Stanford Nanofabrication Facility.)

Factory environment is cleaned by:
- Hepa filters and recirculation for the air,
- "Bunny suits" for workers.
- Filtration of chemicals and gases.
- Manufacturing protocols.

41

Level 2 Contamination Reduction: Wafer Cleaning

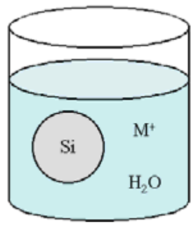
Step	Conditions	Purpose
H_2SO_4/H_2O_2 1:1 to 4:1	120 - 150ÞC, 10 min	Strips organics especially photoresist
HF/H_2O 1:10 to 1:50	Room T, 1 min	Strips chemical oxide
DI H_2O Rinse	Room T	
$NH_4OH/H_2O_2/H_2O$ 1:1:5 to 0.05:1:5 SC-1	80 - 90ÞC, 10 min	Strips organics, metals and particles
DI H_2O Rinse	Room T	
$HCl/H_2O_2/H_2O$ 1:1:6 SC-2	80 - 90ÞC, 10 min	Strips alkali ions and metals
DI H_2O Rinse	Room T	

- RCA clean is "standard process" used to remove organics, heavy metals and alkali ions.
- Ultrasonic agitation is used to dislodge particles.

42

Modeling Wafer Cleaning

- **Cleaning involves removing particles, organics (photoresist) and metals from wafer surfaces.**
- **Particles are largely removed by ultrasonic agitation during cleaning.**
- **Organics like photoresists are removed in an O_2 plasma or in H_2SO_4/H_2O_2 solution**
- **The "RCA clean" is used to remove metals and any remaining organics.**
- **Metal cleaning can be understood in terms of the following chemistry.**



$$Si + 2H_2O \leftrightarrow SiO_2 + 4H^+ + 4e^- \quad (5)$$

$$M \leftrightarrow M^{z+} + ze^- \quad (6)$$

- **If we have a water solution with a Si wafer and metal atoms and ions, the stronger reaction will dominate.**
- **Generally (6) is driven to the left and (5) to the right so that SiO_2 is formed and M plates out on the wafer.**
- **Good cleaning solutions drive (6) to the right since M^+ is soluble and will be desorbed from the wafer surface.**

43

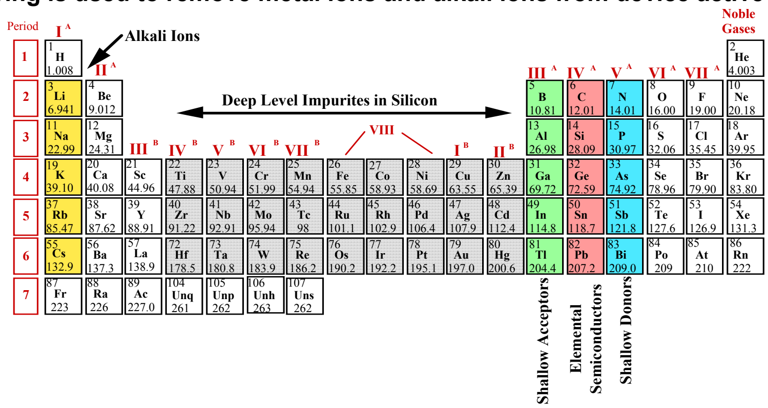
Oxidant/ Reductant	Standard Oxidation Potential (volts)	Oxidation-Reduction Reaction
Mn^{2+}/Mn	1.05	$Mn \leftrightarrow Mn^{2+} + 2e^-$
SiO_2/Si	0.84	$Si + 2H_2O \leftrightarrow SiO_2 + 4H^+ + 4e^-$
Cr^{3+}		$Cr \leftrightarrow Cr^{3+} + 3e^-$
Ni^{2+}		$Ni \leftrightarrow Ni^{2+} + 2e^-$
Fe^{3+}		$Fe \leftrightarrow Fe^{3+} + 3e^-$
H_2SO_4		$H_2O + H_2SO_3 \leftrightarrow H_2SO_4 + 2H^+ + 2e^-$
Cu^{2+}		$Cu \leftrightarrow Cu^{2+} + 2e^-$
O_2		$2H_2O \leftrightarrow O_2 + 4H^+ + 2e^-$
Au^{3+}		$Au \leftrightarrow Au^{3+} + 3e^-$
H_2O_2		$2H_2O \leftrightarrow H_2O_2 + 2H^+ + 2e^-$
O_3		$O_2 + H_2O \leftrightarrow O_3 + 2H^+ + 2e^-$

- **The strongest oxidants are at the bottom (H_2O_2 and O_3). These reactions go to the left grabbing e^- and forcing (6) to the right.**
- **Fundamentally the RCA clean works by using H_2O_2 as a strong oxidant.**

44

Level 3 Contamination Reduction: Gettering

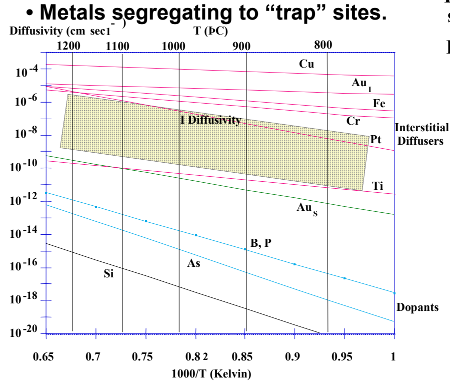
- Gettering is used to remove metal ions and alkali ions from device active regions.

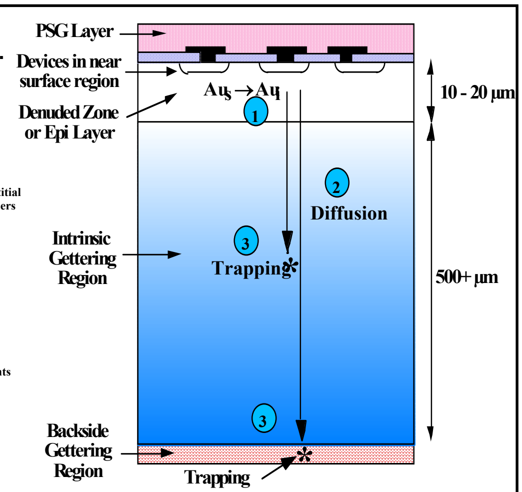


- For the alkali ions, gettering generally uses dielectric layers on the topside (PSG or barrier Si_3N_4 layers).
- For metal ions, gettering generally uses traps on the wafer backside or in the wafer bulk.
- Backside = extrinsic gettering. Bulk = intrinsic gettering.

45

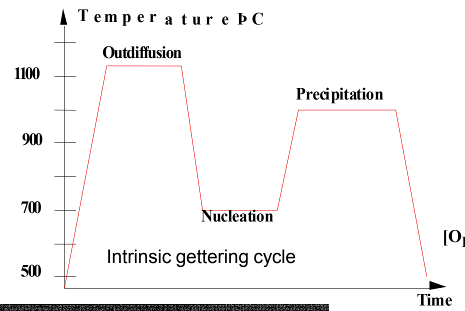
- Heavy metal gettering relies on:
  - Metals diffusing very rapidly in silicon.
  - Metals segregating to "trap" sites.

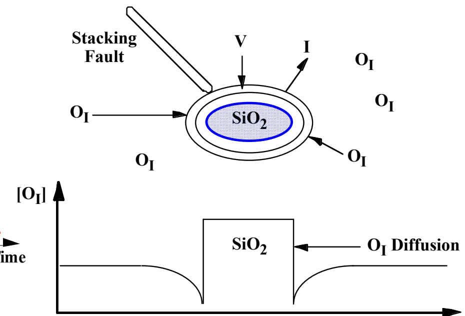




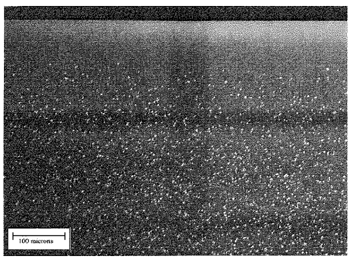
- Gettering consists of
  1. Making metal atoms mobile.
  2. Migration of these atoms to trapping sites.
  3. Trapping of atoms.
- Step 1 generally happens by kicking out the substitutional atom into an interstitial site. One possible reaction is:
- Step 2 usually happens easily once the metal is interstitial since most metals diffuse rapidly in this form.
- Step 3 happens because heavy metals segregate preferentially to damaged regions or to N^+ regions or pair with effective getters like P (AuP pairs).
- In intrinsic gettering, the metal atoms segregate to dislocations around SiO_2 precipitates.

46





(See Chapter 3 class notes)



SiO_2 precipitates (white dots) in bulk of wafer.

- **"Trap" sites can be created by SiO_2 precipitates (intrinsic gettering), or by backside damage (extrinsic gettering).**
- **In intrinsic gettering, CZ silicon is used and SiO_2 precipitates are formed in the wafer bulk through temperature cycling at the start of the process.**

Defects & Yield

Typical "learning curve"

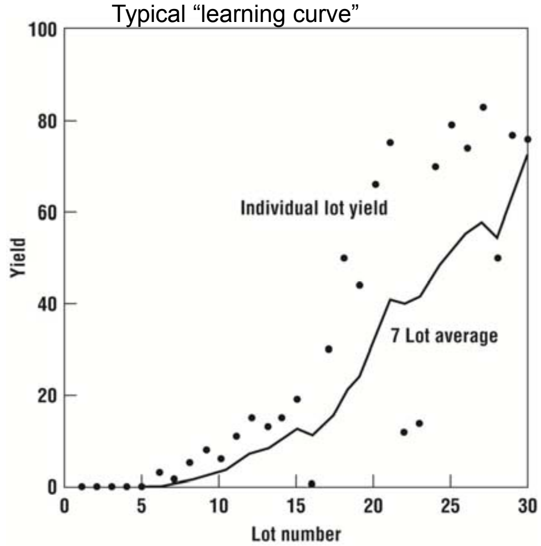


Figure 20.1 A plot of yield versus lot number for a particular technology. Also shown is the running average for the last seven lots.

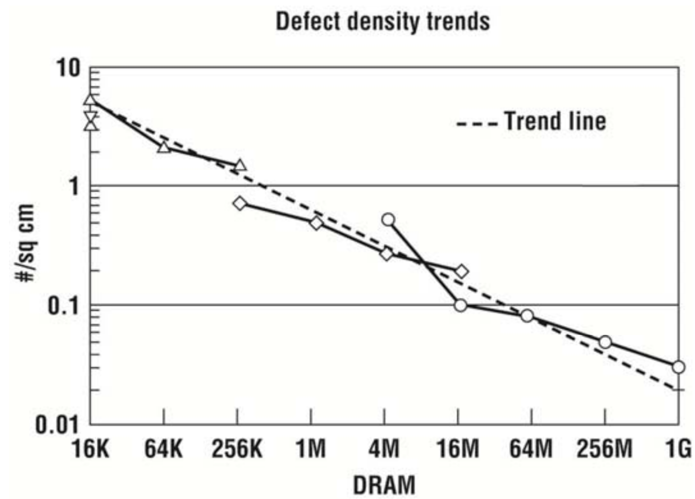


Figure 20.2 Defect–density trend with generation of DRAMs *(after Dance and Gildersleeve).*

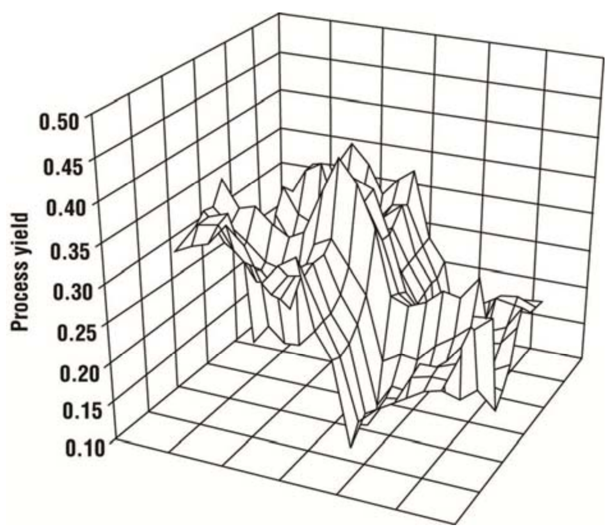


Figure 20.3 Yield plot as a function of die position on the wafer. Yield is often higher near the center of the wafer, but the right side of the wafer shows a strong yield falloff that bears investigation.

Process control modules: Highlight (contamination) defects

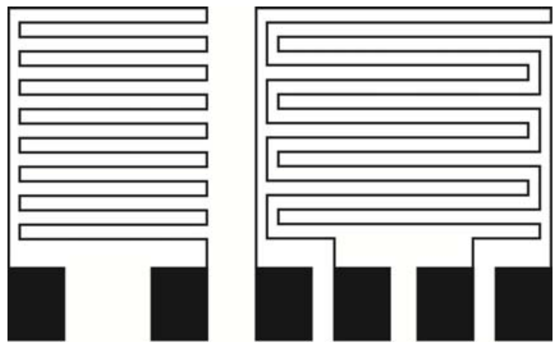


Figure 20.4 Meander lines and interdigitated comb structures used to determine a layer defect density.

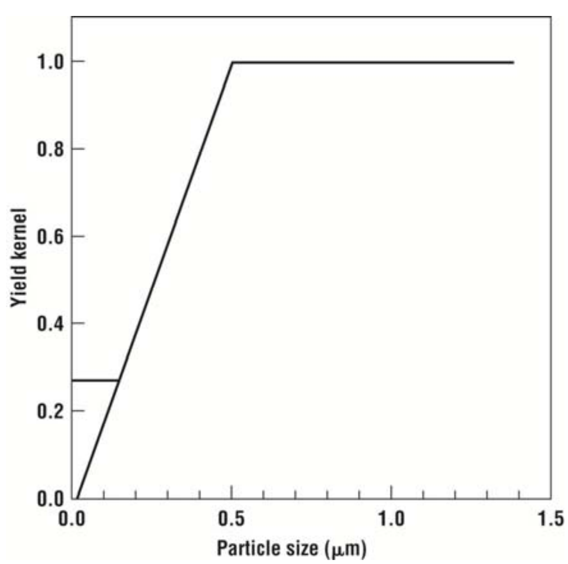
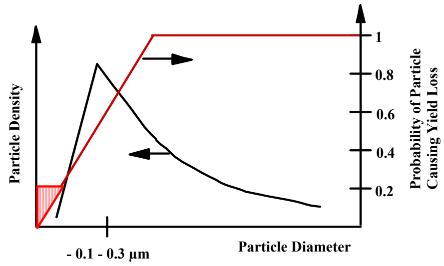


Figure 20.6 Two typical fault kernel curves

Modeling Particle Contamination and Yield



- ≈ 75% of yield loss in modern VLSI fabs is due to particle contamination.
- Yield models depend on information about the distribution of particles.
- Particles on the order of 0.1 - 0.3 µm are the most troublesome:
  - larger particles precipitate easily
  - smaller ones coagulate into larger particles

- Yields are described by Poisson statistics in the simplest case.

$$Y = \exp^{-A_C D_O} \qquad (3)$$

where A_C is the critical area of the chip (i.e. sensitive to defects) and D_O the defect density.

$$Y \rightarrow (1-G)\exp^{-A_C D_O} \text{ if G = fraction of wafer where all circuits fail (edges, test, etc)}$$

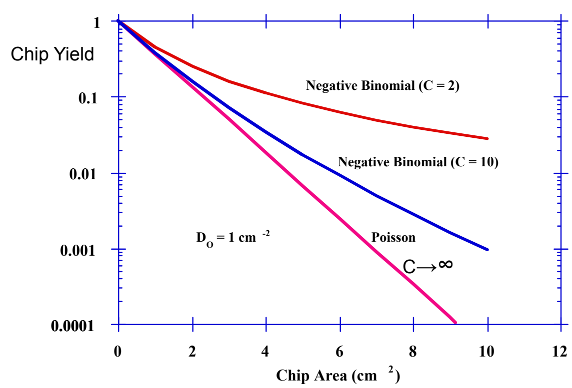
- This model assumes independent randomly distributed defects and often under-predicts yields.

53

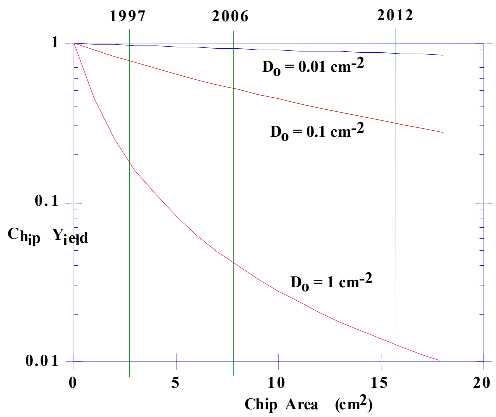
- Use of negative binomial statistics eliminates these assumptions and is more accurate.

$$Y = \frac{1}{\left(1 + \frac{A_C D_O}{C}\right)^C} \qquad (4)$$

where C is a measure of the particle spatial distribution (clustering factor).



54



- Vertical lines are estimated chip sizes (from the ITRS).
- Note that defect densities will need to be extremely small in the future to achieve the high yields required for economic IC manufacturing.
- (See particle defect densities)

$$\textit{Overall yield } Y = \prod_{i=1}^{levels}\left(1 + \frac{A_{C_i} D_{0_i}}{C_i}\right)^{-C_i}$$

55

Defect sizes

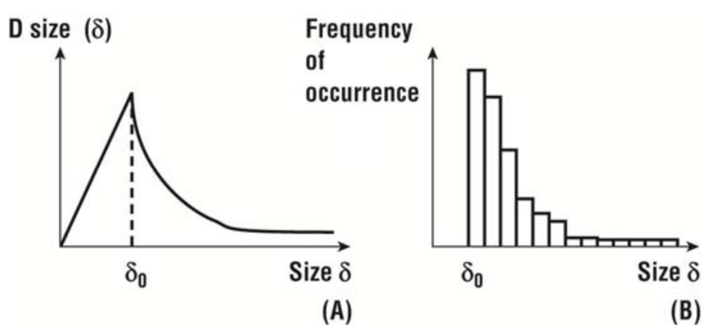
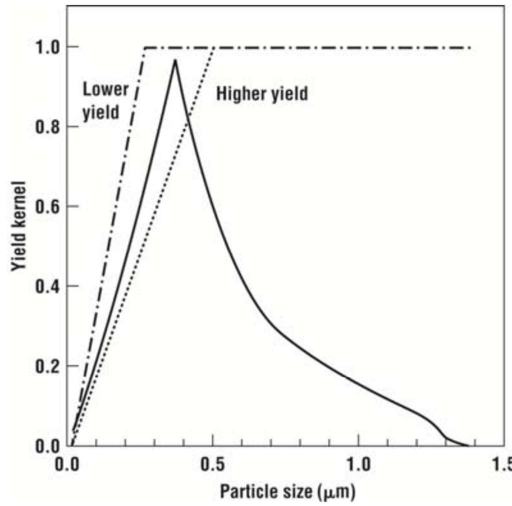


Figure 20.5 Theoretical and measured particle density as a function of particle diameter *(after deGyvez).*

$$D_{size} = C\frac{\delta^q}{\delta_0^{q+1}} \text{ for } 0 \le \delta \le \delta_0$$

$$D_{size} = C\frac{\delta_0^{p-1}}{\delta^p} \text{ for } \delta_0 \le \delta \le \delta_{\max}$$

$$D_{size} = 0 \qquad \text{for } \delta > \delta_{\max}$$



$$Y = (1 - G)\exp^{-D\phi A}$$

for ϕ = averaged probability that mean size defect is fatal

$$\phi = \int_0^\infty D_{size}(x)K(x)dx,$$

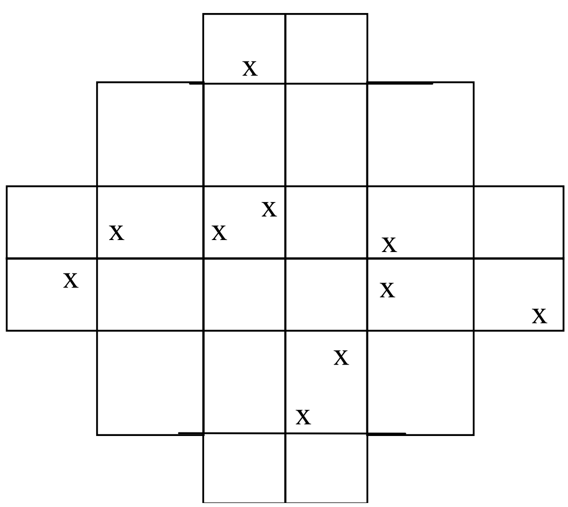
$K(x)$ = fault probability kernel

$$Y = \prod_{i=1}^{levels} Y_i = \prod (1 - G_i)e^{-D_i\phi_i A}$$

Figure 20.7 An overlay of the fault kernel (broken lines) and the defect distribution (solid lines). Small shifts in the fault kernel curve can lead to dramatic changes in the yield.

Fabrication Engineering at the Micro- and Nanoscale Campbell

Uniform Defects Distribution



24 chips,
10 defects
--> 8 chip fails,
16 good

Uniform Defects: Statistics

Problem is:

Place n balls in N cells.

Calculate probability that a given cell contains k balls.

For n defects spread over N chips on wafer, probability that given chip contains k defects

$= P_k = n! / (k! (n-k)!)\ N^{-n}\ (N-1)^{n-k}$

(Binomial distribution)

$\approx e^{-m} m^k / k!$ (Poisson distribution)

for n & N large, n/N = m finite

Uniform Defects Yield

Yield = Probability that chip contains no defects

$= Y_1 = P_o = e^{-m}$

ie. P_oxN good chips on wafer

Prob. that chip contains one defect = $P_1 = me^{-m}$

If chip area = A, total number of chips = N,

then total area = NA, & Defect density $D_0 = n/NA$

Avg no of defects/chip= $n/N = m = D_0NA/N = D_oA$

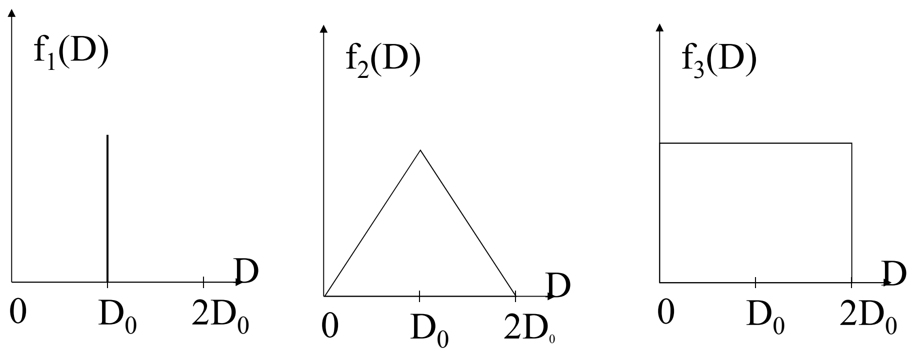
$\therefore\ Y_1 = P_0 = e^{-D_0A}$

In practice this value is much too low. Fallacy lies in random distribution of defects. In practice, processing problems tend to cluster in given area.

Non-uniform Defect Distributions

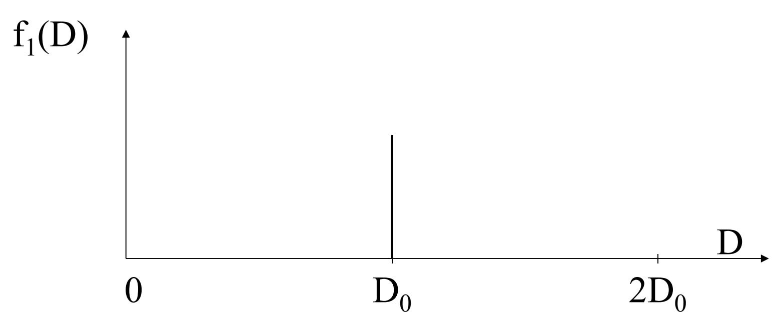
Rewrite yield = $Y = {}_0\int^{\infty} e^{-DA} f(D)\, dD$

where ${}_0\int^{\infty} f(D)\, dD = 1$



Average defect density = D_0 in all cases.

Delta Function

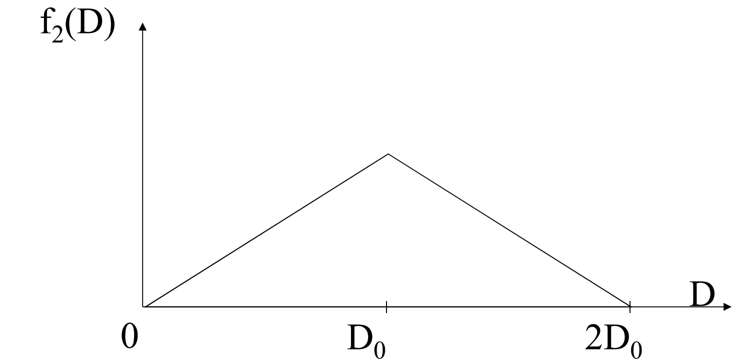


$D = D_0$, i.e. uniform distribution

$Y_1 = e^{-D_0 A}$

and for $D_0 A >> 1$, $Y_1 = e^{-D_0 A}$

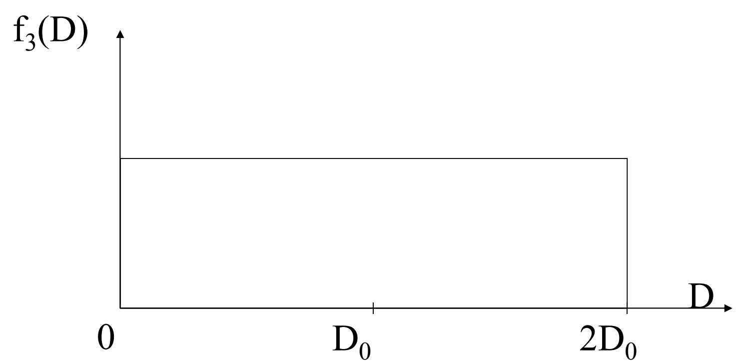
Triangular Function



$Y_2 = [\ (1 - e^{-D_0A}) / (D_0A)\ ]^2$

and for $D_0A \gg 1$, $Y_2 \approx 1 / (D_0A)^2$

Rectangular Function



$Y_3 = (1 - e^{-2D_0A}) / (2D_0A)$

and for $D_0A \gg 1$, $Y_3 \approx 1 / 2D_0A$

Comparison

$Y_1 = \exp - D_0A$

$Y_2 = (D_0A)^{-2}$

$Y_3 = (2D_0A)^{-1}$

& $Y_2, Y_3 >> Y_1$

ie. not as pessimistic as Y_1

Gamma Distribution #1

$f_4(D) = (\Gamma(\alpha)\,\beta^{\alpha})^{-1}\; D^{\alpha -1}\, e^{-D/\beta}$

Average density $D_0 = \alpha\beta$

variance in D = $\alpha\beta^2$

coeff. of variation = $(\mathrm{var}(D))^{1/2} / D_0$

$= (\alpha\beta)^{1/2} / \alpha\beta$

$= 1 / \alpha^{1/2}$

Probability of chip having k defects,

$P_k = {}_0\int^{\infty} e^{-m} (m^k / k!)\, f(D)\, dD$

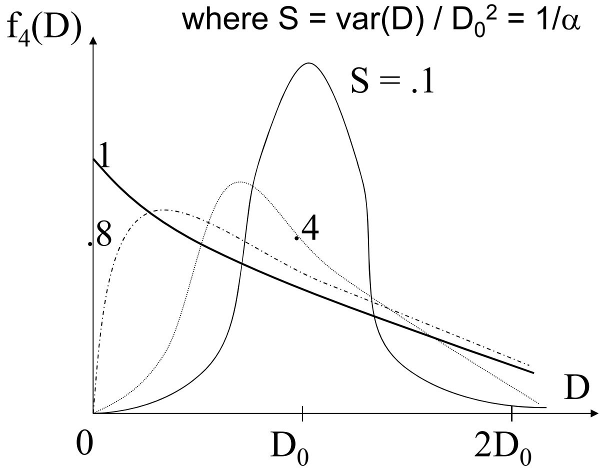
$= [\Gamma(k+\alpha) / (k!\,\Gamma(\alpha))].[(A\beta)^k / (A\beta+1)^{k+\alpha}]$

Gamma Distribution #2

Probability of no defects on given chip

$Y_4 = P_0 = 1 / (A\beta + 1)^{\alpha} = 1 / (1+SAD_0)^{1/s}$

where $S = var(D) / D_0^2 = 1/\alpha$



For s=1, α=1:

$f_4(D) = \beta^{-1} e^{-D/B}$,

$Y_4 = (1+AD_o)^{-1}$

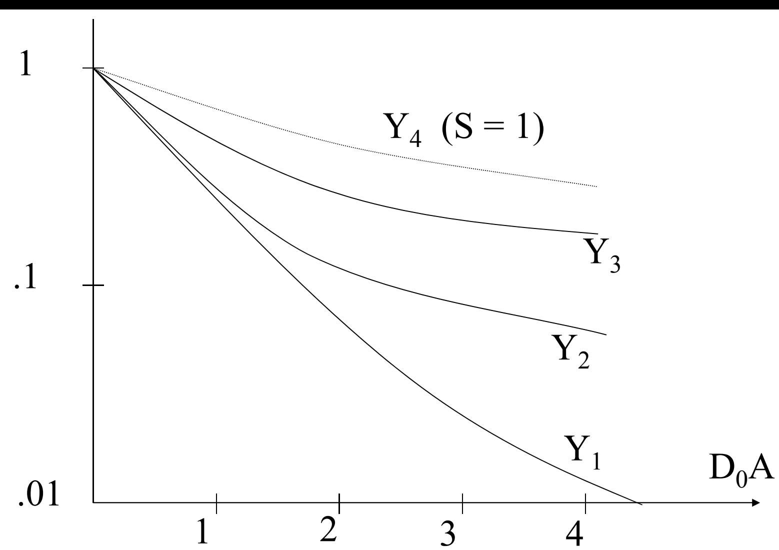
For s=0, $\alpha=\infty$:

$f_4(D) \dashrightarrow \delta(D-D_0)$

$= f_1(D)$

$Y_4 \dashrightarrow Y_1$

Yield for Various Models



Yield with Redundancy

Yield with redundant circuit designed into chip:

$$Y_1 = P_0 + \eta P_1$$

P_0 is defect probability
P_1 is probability of one defect
η is probability that one defect can be "repaired" by redundancy.

Various Defects #1

In practice, each <u>type</u> of defect has its own distribution function

eg. Thin gate oxide, metal opens, etc.

i.e. each defect "n"

---> mean defect density D_0 --> D_{no}
---> distribution shape factor S --> S_n
---> total chip area susceptible to defect A --> A_n

$\therefore$ for each type of defect n, $Y_n = (1+S_nA_nD_{no})^{-1/Sn}$

$\therefore$ overall yield $Y = \prod_{n=1}^{N} Y_n = \prod_{n=1}^{N} (1+S_nA_nD_{no})^{-1/Sn}$

Various Defects #2

If $S_n A_n D_{no} << 1$,

ie. y_n --> 1 for all individual defect types,

ie. process has been well characterized, defects minimized, controlled etc.,

$$\ln Y = {}_{n=1}\Sigma^{N} - S_n^{-1} \ln(1+S_n A_n D_{no})$$
$$\approx {}_{n=1}\Sigma^{N} - S_n^{-1}(S_n A_n D_0) = {}_{n=1}\Sigma^{N} - A_n D_{no}$$

$$\therefore Y \approx \exp(-{}_{n=1}\Sigma^{N} A_n D_{no}) = \exp - AD_m$$

where $D_m = A^{-1}\ {}_{n=1}\Sigma^{N} A_n D_{no}$

ie. yield is exponential, independent of shape parameters S_n provided $S_n A_n D_{no} << 1$.

D_m function of circuit type due to A_n

$\therefore D_m$ varies with circuit type

Radial Dependence

Many defect types have radial dependence, especially handling, misalignment, photoresist residue, etc.

$\therefore D(r) = D_0 + D_R e^{(r-R)/L}$, D_0 defect density at center, D_R increase at edge, r radial coordinate, R wafer radius, L characteristic length for edge defects

$\therefore Y_R = (\pi R^2)^{-1}\ {}_0\int^R Y\, dA$, $(A=\pi r^2,\ dA=2\pi r\, dr)$

--> integrate Poisson yield factor over wafer

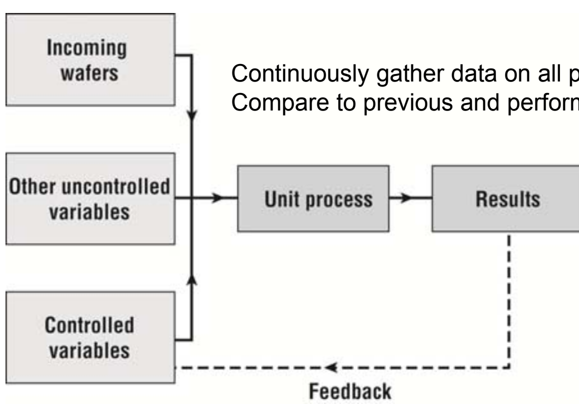
$= 1 / (\pi R^2)\ {}_0\int^R e^{-D(r)A}\, 2\pi r\, dr$

$= 2R^{-2}\ {}_0\int^R e^{-D(r)A}\, r\, dr$

Summary of Key Ideas

- A three-tiered approach is used to minimize contamination in wafer processing.
- Particle control, wafer cleaning and gettering are some of the "nuts and bolts" of chip manufacturing.
- The economic success (i.e. chip yields) of companies manufacturing chips today depends on careful attention to these issues.
- Level 1 control - clean factories through air filtration and highly purified chemicals and gases.
- Level 2 control - wafer cleaning using basic chemistry to remove unwanted elements from wafer surfaces.
- Level 3 control - gettering to collect metal atoms in regions of the wafer far away from active devices.
- The bottom line is chip yield. Since "bad" die are manufactured alongside "good" die, increasing yield leads to better profitability in manufacturing chips.

Statistical Process Control (SPC)



Continuously gather data on all processes.
Compare to previous and performance.

Figure 20.9 A schematic of a semiconductor process. Both controlled and uncontrolled variables must be considered.

Full Factorial experiments (ANOVA)

2 variables. In practice, many more, and data (number of tests) rapidly escalates.

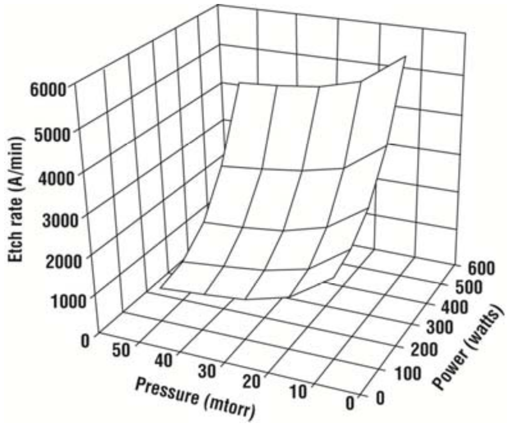
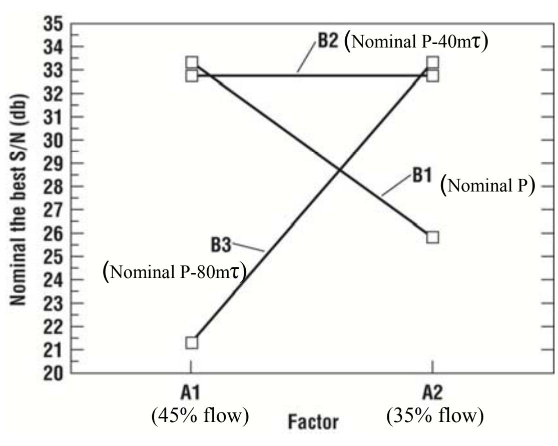


Figure 20.10 Etch rate of a plasma system as a function of power and pressure.

Design of experiments (DOE) for more complex (limits data)

1. Identify all variables. 2. "Screening" experiments (to identify "weak" effects.)
3. 2nd, 3rd experiment sequences on significant parameters.



Note variation with P …. "confounded"

Interpretation difficult with few experiments

Figure 20.11 Particle counts as a function of silane flow and chamber pressure for the low pressure deposition of polysilicon *(after DePinto).*

Computer Integrated Manufacturing (CIM)

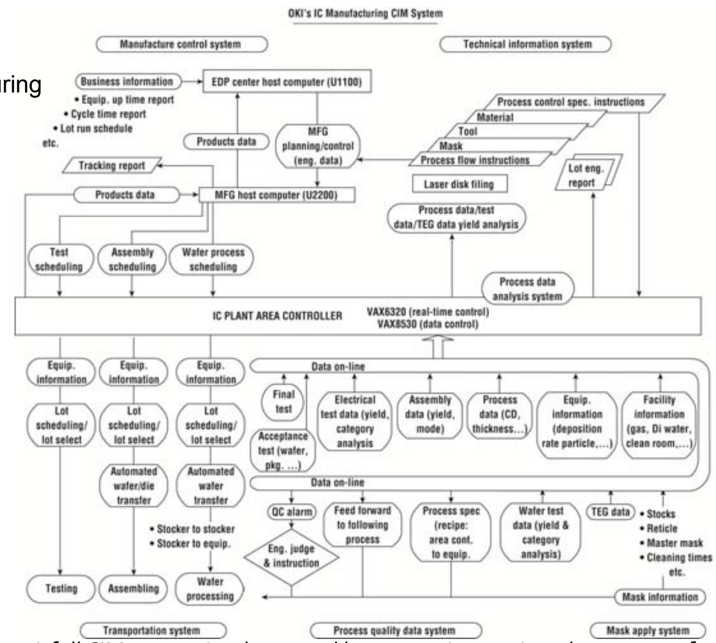


Figure 20.12 A full CIM system implemented by one major semiconductor manufacturer *(from Mizokami).*

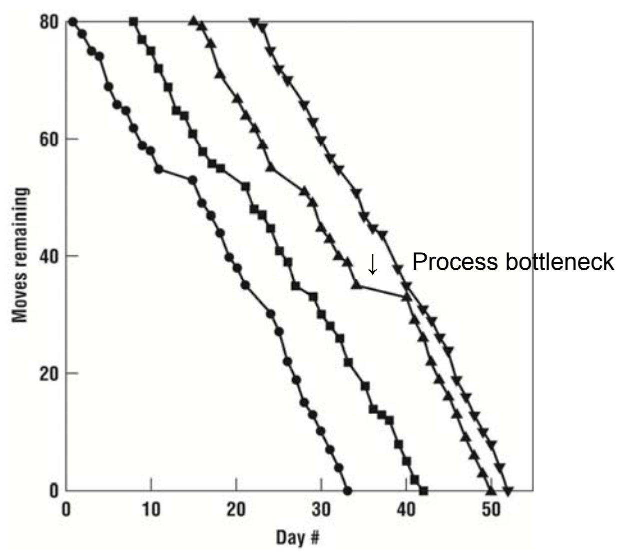


Figure 20.13 Charts showing the remaining lot moves as a function of time are often used to point out bottlenecks in the line. These diagrams are sometimes called chicken scratches.

Target Failure Rates

Say 100,000 components in system
Say < 1 failure/month
$\therefore \lambda < 1$ failure / 10^5 devices x 720 hours
= 14×10^{-9} failures / device hr

Define 1 Failure unIT (FIT)
= 1 failure/10^9 device hrs.

10 FIT --> < 1 service call/month (target)
100 FIT --> < 1 service call/ 4-5 days (OK)
1000 FIT --> ~ 2-3 service calls/ day (unacceptable)

Target Failure Rates

FIT	Fails/month	% failures in 10yr life
10	0.7	0.1%
100	7.0	1%
1000	70	10%

10,000 devices

FIT	Fails/month	% systems failing/month
10	0.07	1%
100	0.7	10%
1000	7.0	65%

Target Failure Rates

200 devices

FIT	MTTF (years)	% systems failing/month
10	51	0.16%
100	5	1.6%
1000	0.5	16%

100 devices on test

FIT	Time to 1st failure
10	114 yrs
100	11 yrs
1000	1 yr

Run 500 devices for 6 months

Confidence level (%)	Failure rate (FIT)
99	2100
95	1400
90	1100
60	430
Best estimate	325

Accelerated Testing

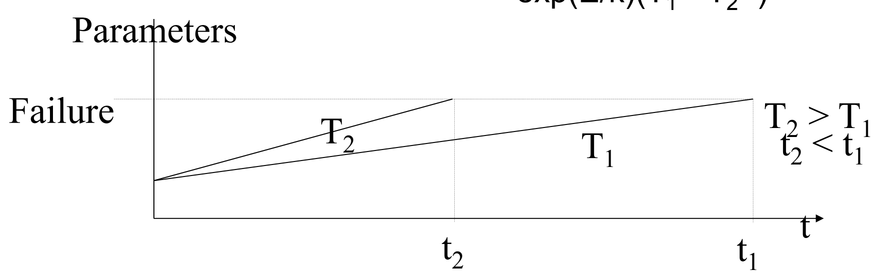
For 100 FIT
must wait 114 years for 1 device from 100 to fail.

Need 10^5 - 10^{11} hrs (depends on σ) for median life.

∴ Obviously need to speed up failure rates
& relate back.

Temperature Acceleration

Failure mechanisms chemical etc: Rate $R=R_0 \exp{-E/kt}$
Speed up by increasing T Accel = $R_2/R_1 = t_1/t_2$
$=\exp(E/k)(T_1^{-1}-T_2^{-1})$



Find E_a for time to failure & T
$t_F = \text{const} * \exp E_a/kT$
$\ln t_F = \text{const} + E_a/kT$
Then can extrapolate times back to normal T

Acceleration Factors

Incr T (°C)	Acceleration factor		Time equiv to 40 yrs (hrs)	
	E_a=1eV	E_a=0.5eV	E_a=1eV	E_a=0.5eV
85	11.5	3.4	30,000	103,000
125	300	17	1200	20,200
150	1700	41	200	8,526
200	31,000	176	11	2,000
250	320,000	570	1.1	616
300	2,200,00	1500	0.2	233

Miscellaneous

Voltage/Current

Rate $R(T, V) = R_0(T).V^{\gamma(T)}$ $\gamma \sim 1 \rightarrow 4.5$

Cannot get much acceleration with practical device voltages.

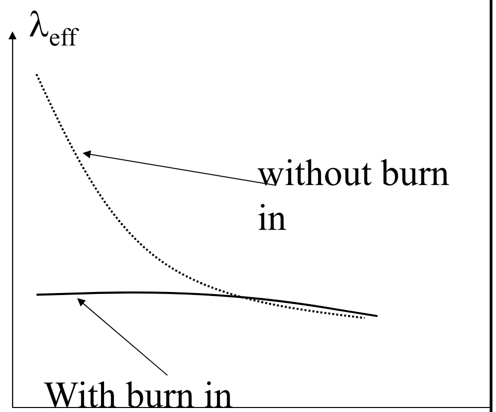
Dielectric breakdown

---> more burn in

Humidity

Burn in

$\lambda_{EFF}(t) = \lambda(t+\tau)$



Summary

- Contamination
- Defect distributions
- Yield statistics for various defect models
- Modeling, SPC, DOE, & CIM
- Failure statistics & Accelerated testing
- (Reliability Theory)

Assignment #8

- 15.4
- 15.7
- 15.10
- 15.11
- 20.1
- 20.2

Reliability & Failure: CDF

Cumulative Distribution Function (cdf)

Device or system operating at t=0

$F(t)$ = probability that device will fail by time t

$F(t) = 0$, $t<0$

$0 \leq F(t) \leq F(t') \text{ for } 0 \leq t \leq t'$

$F(t) \to 1$ as $t \to \infty$

Define Reliability Function $R(t) = 1 - F(t)$

= probability that device will survive until time t

Reliability & Failure: PDF

Probability Density Function (pdf)

$f(t) = dF(t)/dt$

ie. $F(t) = \int_0^t f(x)\,dx$

$\therefore R(t) = 1 - F(t)$

$= \int_0^\infty f(x)\,dx - \int_0^t f(x)\,dx$

$= \int_t^\infty f(x)\,dx$

$f(t) = d/dt\,(1 - R(t)) = -\,d/dt\,R(t)$

Reliability & Failure: Hazard Rate

Failure Rate (Hazard Rate)

i.e. instantaneous failure rate (not average)

$F(t+\Delta t) - F(t) = R(t) - R(t+\Delta t)$

=fraction of devices good at t which fail by $t+\Delta t$

Average failure rate during Δt

$= (1/\Delta t)\,[\,(R(t) - R(t+\Delta t))/R(t)\,] = \lambda(t)$,

because divide by number left at t

$\therefore$ as $\Delta t \to 0$,

$\lambda(t) \to R(t)^{-1}\,dR(t)/dt$

$= -\,d/dt\,\ln R(t) = f(t)/R(t) = f(t)/(1-F(t))$

ie. $R(t) = \exp\,[-\int_0^t \lambda(x)\,dx\,]$

instantaneous failure rate

Reliability & Failure: MTTF

Mean time to failure (MTTF)

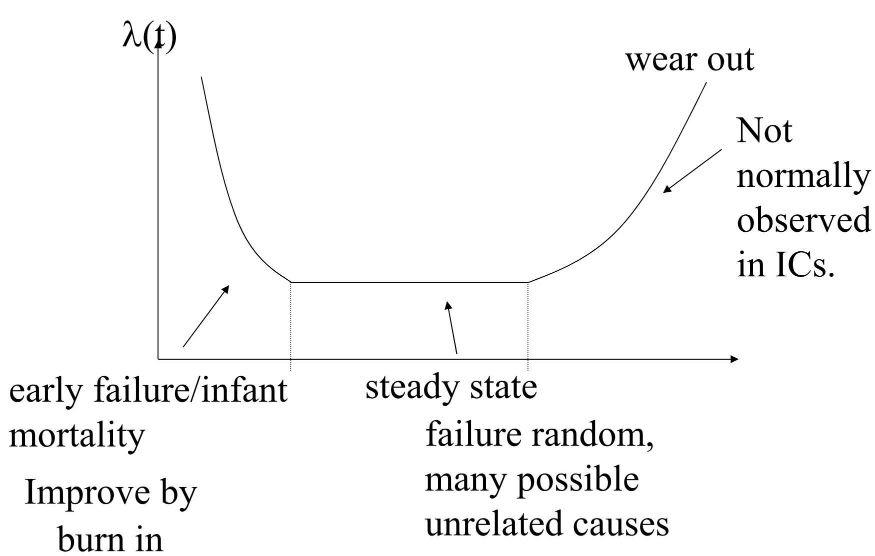
= mean time between failures MTB

if repair assumed

MTTF = $\int_0^\infty t\, f(t)\, dt$

ie. average age at failure

Failure Rate Distributions



A: Constant Failure Rate

$\lambda(t) = \lambda_0$ constant

$\therefore$ $R(t) = e^{-\lambda o\, t}$ & $F(t) = 1 - e^{-\lambda o\, t}$

$\therefore$ $f(t) = d/dt\, F(t) = \lambda_0\, e^{-\lambda o\, t}$

& MTTF $= {}_0\!\int^{\infty} t\, \lambda_0 e^{-\lambda o\, t}\, dt$

$= \lambda_0 [\, t\, e^{-\lambda o\, t} / -\lambda_0 - {}_0\!\int^{\infty} e^{-\lambda o\, t} / -\lambda_0\, dt]$

$= [\, -te^{-\lambda o\, t} + e^{-\lambda o\, t} / -\lambda_0\,]_0^{\infty} = 1 / \lambda_0$

B: Weibull Failure Rate Distribution

Weibull Distribution function Failure rate $\propto t^{power}$

$\lambda(t) = (\beta/\alpha)\, t^{\beta-1}$

$\beta < 1$ failure rate decr with t (early failures)
$\beta > 1$ failure rate increases with t (wear out)
$\beta = 1$ failure rate constant

$R(t) = \exp - {}_0\!\int^{t} (\beta/\alpha)\, t^{\beta-1} dt$
$= \exp -\beta/\alpha .\ t^{\beta}/\beta \quad = \exp - t^{\beta}/\alpha$

$F(t) = 1 - e^{-t^{\beta}/\alpha}$

$f(t) = (-e^{-t^{\beta}/\alpha})((- \beta t^{\beta-1})/ \alpha) = (\beta/\alpha)\, t^{\beta-1} \exp - t^{\beta}/ \alpha$

Weibull Notes: Weibull Plots

Notes (I) If some time τ of device life expired due to device test, etc,

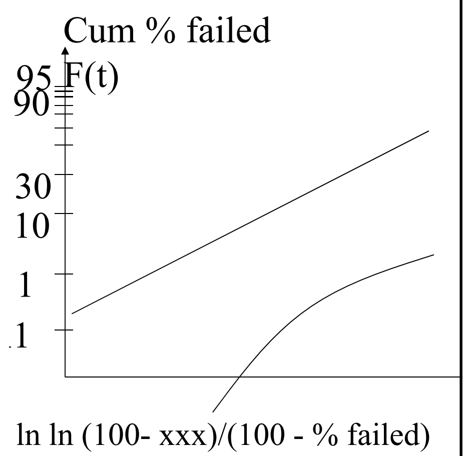
Write $t = t' + \tau$

(ii) Weibull plotting paper:

$1 - F(t) = e^{-t^{\beta}/\alpha}$

$\ln(1 - F(t)) = -t^{\beta}/\alpha$

$\ln\{\ln[1 / (1-F(t))]\} = \beta \ln t - \ln \alpha$



Duame Plotting

Plot average failure rate (AFR) vs log t

AFR = Fraction of failed devices/t=F(t)/t

If plot is straight line, negative slope:

$\ln AFR = -s \ln t + \ln k$

implies $AFR = F(t)/t = k\, t^{-S}$

$\therefore F(t) = k\, t^{1-S}$, $f(t) = k(1-s)t^{-S}$

$\lambda(t) = f(t)/(1-F(t)) = (k(1-s)t^{-S}) / (1-kt^{1-S})$

If only few devices have failed $F(t) << 1$

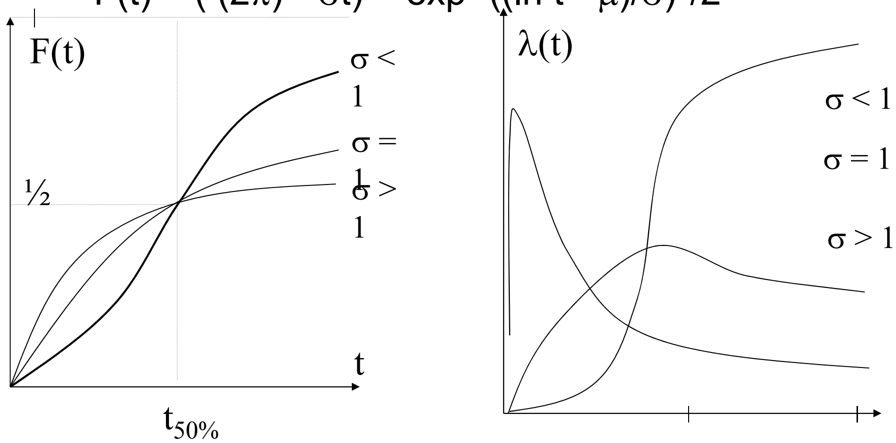
$\lambda(t) \approx k\,(1-s)\, t^{-S} = (1-s)\, AFR$

Compare Weibull with $\beta = 1-s$, $\alpha = 1/k$,

but cannot extrapolate shorter Duame plot for $F(t) << 1$ to long term where inequality fails.

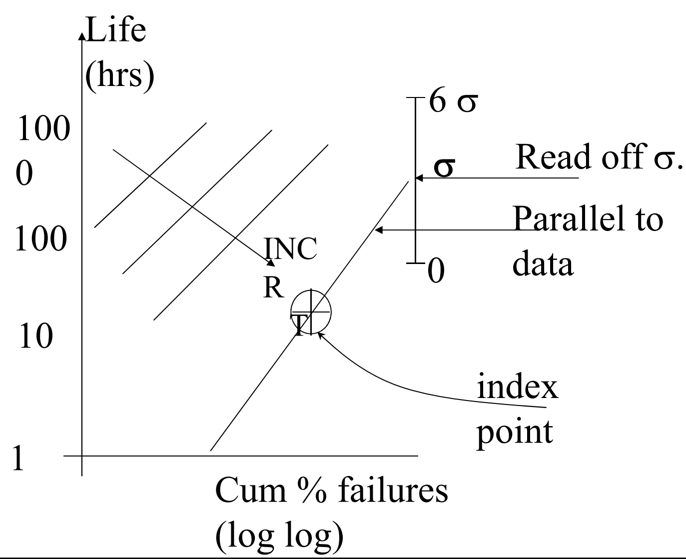
C: Log Normal Distribution Function

$$F(t) = ((2\pi)^{1/2}\sigma t)^{-1} \exp -((\ln t - \mu)/\sigma)^2/2$$



Can describe statistics of device failure over wide periods

Log Normal Plotting



Note: Bimodal distributions

Bimodal - log normal

$\lambda_T(t) = \lambda_S(t)$ * fraction of sports

+ $\lambda_M(t)$ * fraction of main population