CMOS Processing - An Overview

- Silicon IC's (Integrated Circuits) are created on large circular sheets of silicon called "wafers".
- A silicon wafer can be up to 12" in diameter for modern processes.
- A wafer will contain a few 100 up to a few 1000 IC's, depending on the size of each IC.

Figure 4.1 Silicon wafer showing die sites

- The flat edge of the wafer provides a reference for the grid and the crystal orientation.
- The wafer is very thin - just thick enough to provide mechanical stability so that the wafer can be handled without breaking.
- The surface of the wafer is highly polished, and provides a "substrate" on which the IC's are built.
- IC's (or "die") are created by building up many layers on the wafer surface, one at a time; each stacked on top of the previous layers.
- Each layer is patterned by a photolithographic process, to create the structures needed for that layer, using "masks".
- Wafers are processed together in groups, or "lots", of 6-24 at a time.
CMOS Processing - An Overview (cont.)

- **Several copies of a chip are grouped together (typ. 4-9) on a "reticle"**

- **This multi-chip reticle is then used to expose the wafer surface using a "step-and-repeat" process with a "wafer stepper"**

- **Several "test sites" are included on the wafer, to allow the process engineers to test the wafer to insure quality**

![Image of wafer sites]

Figure 4.15 Wafer sites

- **Test sites (or "drop-ins" or "wat keys") are placed strategically on the wafer to check quality on different locations of the wafer using "wafer probes", giving "wat" or "e-test" data such as VT, R, IpSAT, etc.**

- **Geometric patterns called "registration marks" are used to align each layer to the next**

- **Wafers are processed in a "clean room" to reduce defects due to dust, etc.**

- **A "class X" clean room means < X particles per cubic foot in the air with diameters > 0.5 μm (e.g., class I clean rooms typical today)**

- **Many modern lines are automated using robots, others require workers to wear "bunny suits"**
THE ECONOMICS OF CMOS PROCESSES

- A STATE-OF-THE-ART CMOS PROCESS FACTORY (FABRICATION PLANT, OR "FAB") CAN COST MORE THAN $1 BILLION TO BUILD!

  → VERY EXPENSIVE, SO NOT MANY ARE BUILT

  → SOME FABS WORK AS "FOUNDRIES," WHICH ALLOW MANY DIFFERENT COMPANIES TO SUBMIT IC DESIGNS TO BE FABRICATED IN THEIR FACTORY

  → LARGE COMPANIES (E.G., INTEL) OWN MANY PLANTS OF THEIR OWN WHERE ONLY THEIR CHIPS ARE MADE ("CAPTIVE" FABS)

  → MASK SETS CAN COST UP TO $1 MILLION EACH, SO FOUNDRIES NOW OFFER "SHUTTLES," WHICH ARE MULTI-PROJECT WAFERS WITH MANY DIFFERENT CHIPS SUBMITTED BY DIFFERENT COMPANIES ON EACH WAFER (COST ≈ $100K)

- NOT ALL CHIPS FABRICATED FUNCTION CORRECTLY, AS MANUFACTURING DEFECTS CAUSE THEM TO FAIL

  → "YIELD" = \( Y = \frac{N_{\text{good}}}{N_{\text{total}}} \times 100\% \)

  → YIELD CAN BE ESTIMATED IN ADVANCE FOR A GIVEN CHIP, BASED ON:

  1) THE AREA OF THE CHIP, AND

  2) THE "DETECT DENSITY" FOR THE PROCESS

  \[ Y = e^{-DA} \times 100\% \] (OTHER MODELS EXIST FOR OTHER TYPES OF DEFECTS)

  WHERE: \( A = \text{AREA OF DIE IN CM}^2 \)

  \( D = \text{AVERAGE # OF DEFECTS PER CM}^2 \)

  → COST IS CALCULATED BASED ON THE EXPECTED # OF GOOD CHIPS PER WAFER

  → BIG CHIP'S YIELD < SMALL CHIP'S YIELD

    (E.G., PENTIUM) (E.G., A NAND GATE)
**Basic CMOS Process Techniques**

- **Silicon Dioxide (SiO₂) Growth/Deposition**
  
  → SiO₂ is used extensively because it can be easily grown and/or deposited; it is an excellent insulator; it adheres well to other materials; it is hard (glass!)
  
  → Grown SiO₂ takes longer to make, but gives a higher quality oxide
  
  ↓ Used for gate oxides, where quality is key
  
  ↓ Made by exposing Si wafer to oxygen at high temperatures to speed growth (~1000°C)
  
  ↓ Also known as a "thermal oxide" consumes some of Si from wafer

![Diagram of Thermal Oxide Growth](image1)

**Figure 4.2** Thermal oxide growth

→ Above the wafer's surface, SiO₂ is deposited using "chemical vapor deposition" (CVD)

↓ Faster to make thick layers, at lower temperatures

![Diagram of CVD Oxide Process](image2)

**Figure 4.3** CVD oxide process

↓ Used to insulate between metal interconnect layers
**Silicon Nitride Deposition**

\[ Si_3 N_4 \] forms a strong barrier to most atoms

- Makes an excellent final coating on a finished chip to seal it and protect it from impurities in the environment ("passivation" or "overglass" layer)
- Used to define the "active area" where MOSFETs are built (more on this later)
- High dielectric constant = 7.8
  - Makes good capacitors (high density)
  - E.g., DRAM
- Always deposited

**Polysilicon Deposition**

- Used primarily to form MOS gates
  - When deposited on top of \( SiO_2 \), forms many small silicon crystals instead of one big one, since no reference crystal exists
  - Can be doped either N or P type
    - TYPICALLY HEAVILY , to make low-resistance
  - TYPICALLY "coated" with a refractory metal (high melting point metal) such as Ti or Pt to reduce the sheet resistance even further (2-10 nΩ typ) ⇒ "silicidation"
  - Also used in analog CMOS processes to make high quality R & C (extra steps)
**Ion Implantation**

Used to dope Si with acceptor (Na) or donor atoms (Np) to form P-type and N-type regions, respectively.

- Atoms are first ionized and then accelerated to high energies (MV).
- A beam of ions are smashed into the Si surface, penetrating it → damages Si crystal.
- An "anneal" Si crystal t. ("Heals" the 1.

![Diagram of ion implanter](image1)

**Figure 4.5** Basic sections of an ion implanter

![Diagram of ion stopping process](image2)

**Figure 4.6** The ion stopping process

![Diagram of Gaussian implant profile](image3)

**Figure 4.7** Gaussian implant profile

\[ R_p = "\text{projected range}" \]
\[ \Delta R_p = "\text{straggle}" \] (standard deviation)
METALIZATION

- Traditionally aluminum (Al) used to make metal interconnect layers
- Copper (Cu) now emerging for 13 µm processes and below, due to lower resistivity (~1/2 Al)
  - But, tough to process! (more later...)
- Al is evaporated by heating in a vacuum chamber and deposited on wafer
- Patterned by "etching" and CMP ("chemical mechanical polishing") — more later on this...
- Problems can occur at high current densities due to "electromigration"

- High currents actually push atoms down metal lines, creating "voids" and "hillocks"
- Requires design rules limiting the maximum current density \( J = \frac{I}{A} \) allowed in metal lines

Figure 4.4 Visualization of electromigration effects in aluminum
PLANARIZATION

- When SiO₂ is deposited between layers to insulate them ("inter-layer dielectric"), the surface is not smooth due to the underlying features.

- To be able to stack many metal interconnect layers on top of each other, each layer is "planarized" before the next is added.

- Planarization is achieved using "CMP" ("Chemical-Mechanical Polishing"), which uses a combination of chemical etching and mechanical polishing.

Figure 4.8 Surface planarization

- For aluminum metal processes, CMP is applied to the oxide.

- For copper metal processes, CMP is applied to the metal (more on copper later...).
PHOTO LITHOGRAPHY

• ICS ARE 3-DIMENSIONAL, BUILT UP FROM MULTIPLE LAYERS ON TOP OF EACH OTHER

• EACH LAYER MUST BE PATTERNED TO DEFINE SHAPES OF WIRES, MOSFETS, ETC.

• PATTERNING IS DONE THROUGH A PHOTOGRAPHIC PROCESS, WHERE MASKS ARE MADE FOR EACH LAYER AND LIGHT IS PASSED THROUGH THEM TO TRANSFER THE PATTERN TO THE SILICON (LIKE A PHOTO NEGATIVE) (LIGHT = UV FOR SHORT I)

Figure 4.9 A reticle is a glass plate with a chromium pattern

• "PHOTORESIST" IS APPLIED TO THE WAFER SURFACE, AND ACT LIKE PHOTOGRAPHIC FILM

→ POSITIVE RESISTS HARDEN IN AREAS NOT EXPOSED TO LIGHT, THE REST IS REMOVED (MOST COMMON IN VLSI)

→ NEGATIVE RESISTS HARDEN IN AREAS THAT ARE EXPOSED TO LIGHT, THE REST IS REMOVED

Note! "POSITIVE" BECAUSE THE REMAINING RESIST FORMS THE SAME PATTERN AS SEEN ON THE MASK

→ THE REMAINING RESIST PROTECTS THE SILICON UNDER IT
Figure 4.10 Photoresist application

Figure 4.11 Exposure step

Figure 4.12 Characteristics of positive photoresist
PHOTOLITHOGRAPHY (CONT.)

- The hardened photoresist protects the areas of the silicon wafer underneath it during the following steps.

- Oxide layers are patterned by “etching”, which is used to remove the oxide not protected by photoresist.

![Diagram showing initial patterning of resist and after etching process.](image)

**Figure 4.13** Etching of an oxide layer

- "Wet etching" uses liquid chemicals (less common now).
- "Reactive-ion etching" (RIE) uses a gaseous plasma containing chemicals (most common now).

- We would like the etch to go straight down, but the chemicals also react & etch laterally.
- An "anisotropic" etch goes straight down (not possible in practice).

![Diagram showing resist pattern, pure anisotropic etch, and isotropic etch.](image)

**Figure 4.28** Etching profiles
Photolithography (cont.)

- **Doped patterns are created on the silicon surface by ion implantation through holes in the oxide**

![Diagram of ion implantation](image)

(a) Incoming ion beam  
(b) Doped n-type regions

Figure 4.14 Creation of doped silicon patterns

- Oxide is needed to pattern implants as the photosist is not tough enough by itself.
- Once implanted, the wafer is "annealed" by heating it to several hundred degrees C, to allow the dopant atoms to redistribute themselves in the crystal lattice through "diffusion".
- Diffusion occurs both vertically and laterally, so the resulting diffusion area is slightly larger than the oxide opening.
- Higher temperatures and longer oven times are used to "drive" some implants deeper (e.g., N-wells).
• For processes using Al metal, the metal interconnect layers are patterned similar to SiO₂ using etching to remove unwanted Al.

• Copper metal (Cu) processes use a different patterning technique, called "Damasocene" or "Dual-Damasocene", where Cu is inlaid into grooves etched in SiO₂ by deposition, and excess Cu removed by CMP.

• CMP is applied to the Cu itself, not the SiO₂!

("Damasocene" comes from the ancient process of inlaying gold in swords in the city of Damascus.)

Figure 4.22 Copper patterning using the Damascene process

Figure 4.23 Dual-Damascene structure with copper vias

• "Vias" – holes etched in SiO₂ and filled with metal to metal layers

• Are the same as vias, but connect to diffusions (e.g., source/drain) or poly (gates)
Figure 4.16 Initial sequences in the CMOS fabrication sequence

(a) Starting wafer with epitaxial layer

(b) Creation of n-well in p-epitaxial layer

(c) Active area definition using nitride/oxide

(d) Silicon etch

(e) Field oxide growth

(f) Surface preparation

*EPI LAYER IS NOT ALWAYS USED (E.G., ANALOG)*
Figure 4.17 Formation of nFETs and pFETs
CMOS PROCESS FLOW (CONT.)

(a) After anneal and CVD oxide

(b) After CVD oxide active contact, W plugs

(c) Metall1 coating and patterning

Figure 4.18 First metal interconnect layer

(a) Cross-section

(b) Layout

Figure 5.20 Metall1-Metall2 connection using a Via mask

- Oxide used as "inter-layer dielectric" for isolation

- Contacts filled with metal plugs of tungsten (W)

- Metal lines and vias look very planar, even on SEMs!

- Modern processes (e.g., 0.13 µm) use up to 8 metal layers
Figure 4.20 Sequence for creating a lightly doped drain nFET

- "LIGHTLY DOPED DRAINS" (LDD) REDUCE MAX ELECTRIC FIELD AND THEREFORE HOT-ELECTRON EFFECTS

Figure 4.21 LDD nFET with silicided gate and contacts

- SILICIDED DRAIN/SOURCE, GATES REDUCE RESISTANCE
Figure 1.5 Bonding pad frame for interfacing

Figure 4.19 Bonding pad structure

NEW APPROACH IS "FLIP-CHIP" (A.K.A. "CY")

"Bumps" placed in an array across entire surface of chip, not just at edges

Chip placed in package upside down to make contacts

Much lower resistance (power) and inductance (speed) than traditional pad ring; no "power busses" required from pad ring into core (lower R); no bond wires required (lower L)

BUT, expensive! (And hard to de-bug chip!)
EXAMPLE OF A FINISHED VLSI CHIP
**Design Rules for Layout**

- **MIN WIDTH**
- **MIN SPACE**
- "PITCH" = WIDTH + SPACE

**Figure 4.24** Design rule limits for two polysilicon lines

- (a) Side view
- (b) Surround rule

**Figure 4.25** Example of a surround design rule

- (a) Top view
- (b) Side view along A-A'

**Figure 4.26** Misalignment-induced defect

- (a) Gate overhang DR
- (b) Misalignment failure

**Figure 4.27** Example of an extend (gate overhang) design rule
Figure 4.29 Limits on n+ spacings

- **NEEDED TO PREVENT TWO ADJACENT DEPLETION REGIONS FROM MERGING = SHORT!**
- **WELLS HAVE BIGGER DEPLETION REGIONS, THUS BIGGER SPACES REQUIRED BETWEEN WELLS TIED TO DIFFERENT VOLTAGES**
Cu CMP Issue Definition

Isolated line

\[ \text{Cu Loss} = \text{GOL} + \text{Recess} + \text{Dishing} \]

Multi-lines

\[ \text{Cu Loss} = \text{GOL} + \text{Ox Erosion} + \text{Recess} + \text{Dishing} \]

- "DUAL DAMASCENE" process used for copper (Cu)
  instead of deposition and etching for aluminum (Al)
  \( \rightarrow \) Cu inlaid into grooves etched in SiO₂
  CMP applied directly to Cu to remove excess

  But, Cu is much softer than SiO₂, leading to "oxide erosion" and "dishing" problems
  \( \Rightarrow \) Requires strict control of "metal density" (typ. 20-80%)

- Cu also is a very mobile ion which can contaminate silicon \( \Rightarrow \) typ. encased in other metal to contain ions
**LATCH-UP**

(a) Structure
(b) Behavior

Figure 5.22 Characteristics of a 4-layer pnpn device

Figure 5.21 Latch-up current flow path

- Caused by close proximity of PMOS & NMOS FETS
- Parasitic SCR!
- Typically occurs when a "spike" happens at an I/O pad
- May require power to be turned off & on to recover
- May destroy chip!
- Epi helps
- I/O pads designed to prevent this

- Parasitic NPN formed by Nwell - Psubstrate - N+ S/D
- Parasitic PNP formed by Psubstrate - Nwell - P+ S/D
- Resistors from Nwell and Psubstrate contacts

* Positive feedback can cause this parasitic circuit to "latch" if loop gain > 1!

**MUST:** 1) Use low- R contacts to Nwell, Psub
2) Keep β of parasitic BJTs low!
Electro Static Discharge (ESD)

- Occurs when a machine or person has a static electric charge build up and then accidentally discharges to an IC
  - Same as the shock you get sometimes when you touch a metal doorknob on a cold, dry day
  - Can be 1000's of volts!
  - When this energy is discharged to a pin on an IC, the IC may be damaged or even destroyed (looks OK; damage is microscopic)

- Input/output (I/O) pins on ICs are typically protected by specially designed structures
  - ESD protection devices must "turn on" and divert this energy off-chip to protect IC

- Most common ESD protection devices are diodes and "snap-back" devices, resistors for inputs

- Industry standard protection expectations are at least >2kV (human-body model)

  "Human-body" model for ESD (~2kV or better)

  "Machine" model for ESD (~500V or better)

- ESD protection devices added to I/O pins must be big to handle large energy surges, which adds lots of capacitance and slows down high speed I/Os.