CMOS Reliability Integration and Engineering (Part-1)

Introduction to Reliability

Dr. Eitan Shauly,
Tower Semiconductor
(Tel) 972-4-6506570, eitansh@Towersemi.com

1

Topics

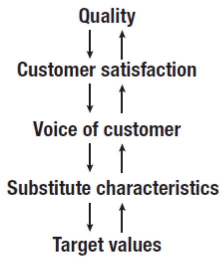
1. Quality and Reliability
2. The Reliability bathtub
3. Failure in time and the acceleration factors (Temp, Voltage stress)
4. MTTF, MTBF, FIT
5. Materials and device degradation vs time – modeling
6. Competing degradation mechanisms
7. Definition of quality and reliability
8. Yield vs Reliability
9. Scaling, MOSFET operation
10. Physical failure mechanisms: HCI NBTI, EM, SM, ESD, Latchup, Soft error

2

Quality and Reliability

Quality: The American Society for Quality (1983) defines quality as the "totality of features and characteristics of a product or service that bear on its ability to satisfy a user's given needs."

Quality function deployment (QFD) means of translating the "voice of the customer" into substitute quality characteristics, design configurations etc, through the whole organization: marketing, engineering, purchasing, etc.

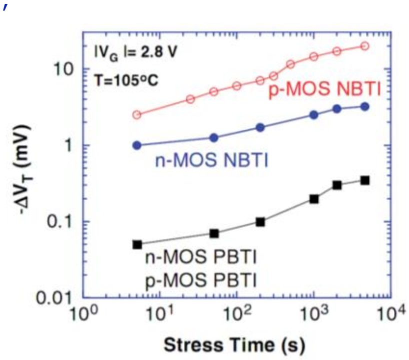


After: Reliability Engineering, Kailash C. Kapur, Michael Pecht (2014)

Quality and Reliability

Reliability: time-oriented quality (2000).

- The probability that a device, product, or system will not fail for a given period of time under specified operating conditions (Shishko 1995).
- Reduction of things gone wrong (2003),
- The capability of a product to meet customer expectations of product performance over time (1997),



After: A. Campos-Cruz et al., "On the Prediction of the Threshold Voltage Degradation in CMOS Technology Due to Bias-Temperature Instability," Electronics, 2018.

Semiconductor Reliability – the bathtub

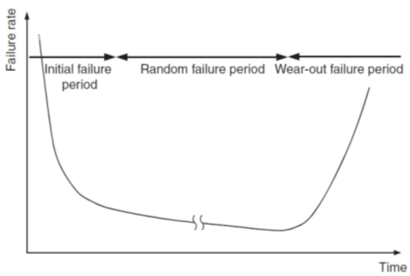
The reliability of semiconductor devices is represented by the failure rate curve (called the "bathtub curve").

The curve can be divided into three regions:

(1) initial failures, which occur within a relatively short time after a device starts to be used,

(2) random failures, which occur over a long period of time,

(3) wear-out failures, which increase as the device nears the end of its life.

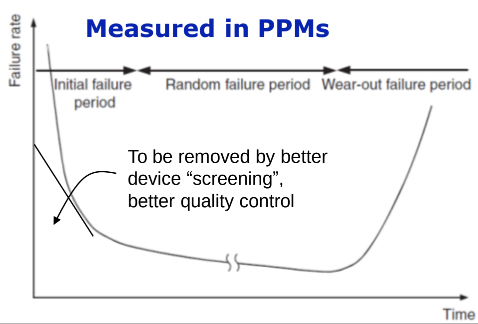


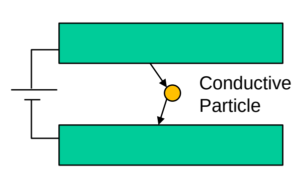
5

The bathtub – Initial failure (early failure)

"Initial failures" are considered to occur when a latent defect is formed, for example, during the device production process and then becomes manifest under the stress of operation. For example, a defect can be formed by having tiny particles in a chip in the production process, resulting in a device failure later. The failure rate tends to decrease with time because only devices having latent defects will fail, and these devices are gradually removed.

Measured in PPMs

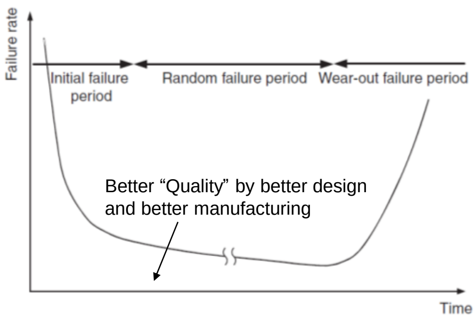




6

The bathtub – Random failure (normal Lifetime)

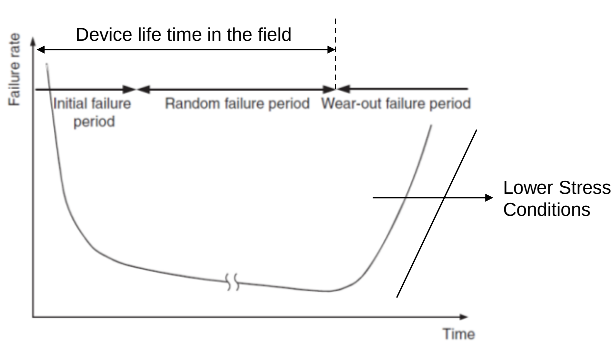
"Random failures" occur once devices having latent defects have already failed and been removed. In this period, the remaining high-quality devices operate stably. The failures that occur during this period can usually be attributed to randomly occurring ***excessive stress***, such as power surges. This group also includes devices susceptible to remains of initial failures (long-life failures).



7

The bathtub – wear-out failure

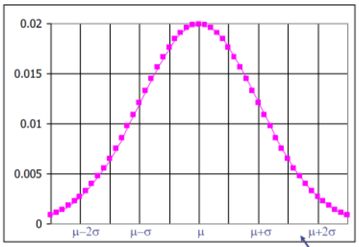
"Wear-out failures" occur due to the aging of devices from wear and fatigue. The failure rate tends to increase rapidly in this period. Semiconductor devices are therefore designed so that wear-out failures will not occur during their guaranteed lifetime.



8

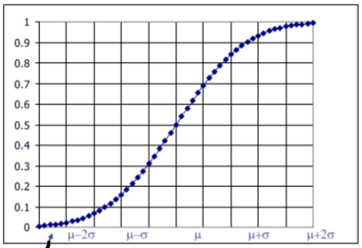
Introduction – Gaussian and normal CDF

PDF: Probability Distribution Function



$$PDF\ f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(t-\mu)^2/2\sigma^2}$$

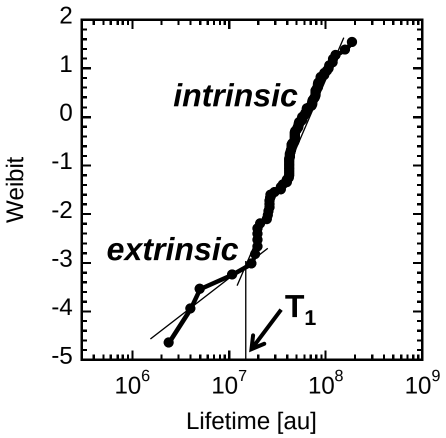
CDF: Cumulative Distribution Function



$$CDF\ F(t) = \int_{-\infty}^{t} f(x)dx = \Phi\left(\frac{t-\mu}{\sigma}\right)$$

TOWER SEMICONDUCTOR LTD.

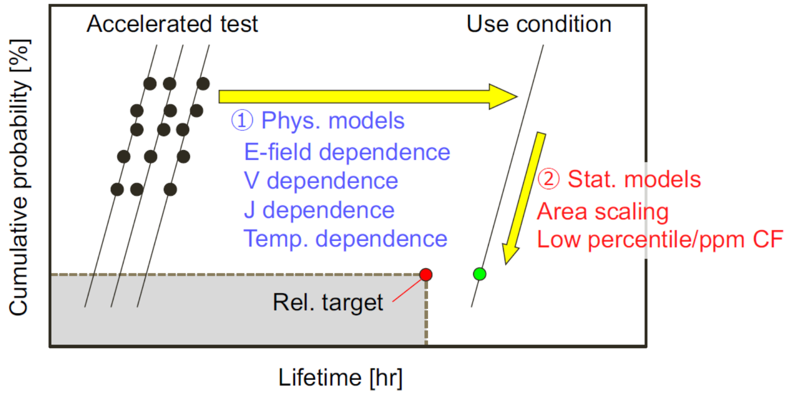
Device LT dependency on Intrinsic and Extrinsic Failure



An example of the typical relationship between the cumulative failure rate and the lifetime - the Weibull plot. The extrinsic and intrinsic failure regions can be easily observed.

TOWER SEMICONDUCTOR LTD.

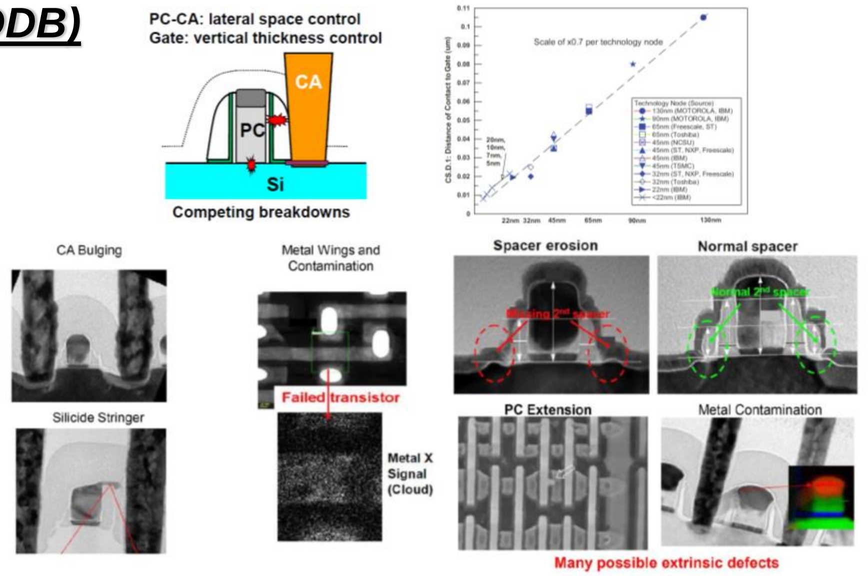
Conventional testing: Stress and extrapolation



① Describe extrapolating on use conditions.
② Describe extrapolating to chip area and target ppm.

12

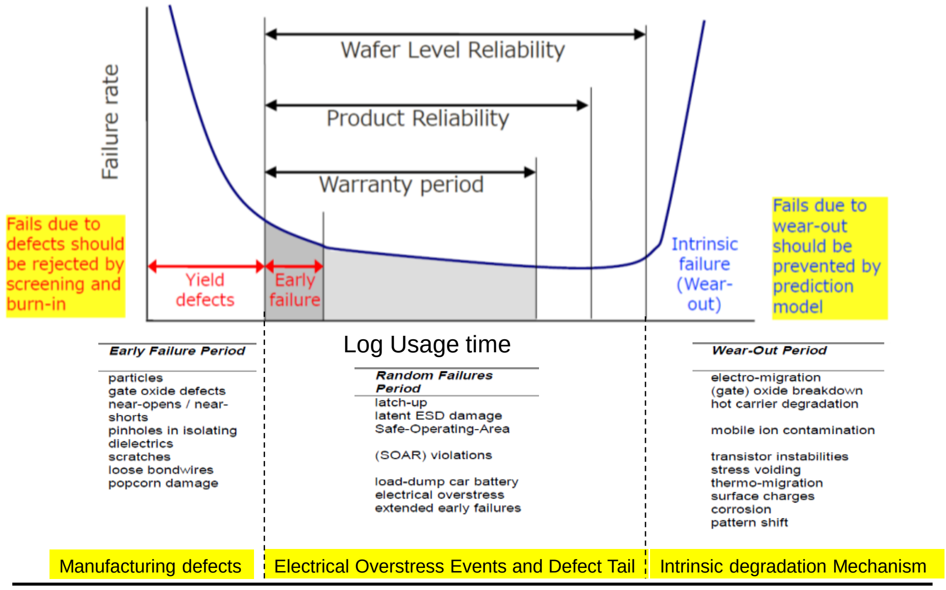
Extrinsic defect examples (for Poly-Gate TDDB)



14

Summary:

After: *JEITA EDR-4708C*, "Guideline for IC Reliability Qualification Plan" (2022.10).



15

Failure in Times and the Acceleration Factor

For *n* number of products, that fail in service after time t1, t2, t3…tn

The mean time to failure is defined as:

$$MTTF = \frac{t_1 + t_2 + t_3 + t_4 + t_5 + \cdots + t_n}{n}$$

The MTTF is different vs the median time, that marked as t_{50}. t1% is the time that the first 1% of the population failed. T0.1%, is the time that the first device from 1000 tested devices is failed.

The ***Temperature acceleration factor*** (TAF) is defined as the ratio of a degradation rate or times to failure at an elevated temperature T2 relative to that at a lower base temperature T1,

$$(\text{Temp})AF = \frac{MTTF(T_1)}{MTTF(T_2)} = exp\left[\frac{E_A}{k_B}\left(\frac{1}{T_1} - \frac{1}{T_2}\right)\right]$$

17

Failure Times and the Acceleration Factor

The ***voltage acceleration factor*** (VAF) is defined as the ratio of a degradation rate or times to failure at an higher voltage V2, relative to that at a lower base voltage V1,

$$(\text{Voltage})AF = \frac{MTTF(V_1)}{MTTF(V_2)} = exp[-\delta(V_2 - V_1)]$$

We will talk more on AF later.

Important note: the acceleration values to be used (Temp, voltage, pressure, humidity and others), must NOT lead to a new working regime.

Quantification by Statistical Parameters

- MTTF = Mean Time To Failure (for non-repairable items)

$$MTTF = \frac{\sum_{i=1}^{n} t_i}{n}$$

- MTBF = Mean Time Between Failure (for repairable items)

$$MTBF = \frac{\sum_{i=1}^{n-1} (t_{i+1} - t_i)}{n}$$

- Failure rate: λ, given in FIT = Failure-In-Time. MTTF=1/ λ

1 FIT = 1 failure/10^9 device hours

Example

Consider a system with 100,000 devices

Question-1: What is the number of failures in 1 month, if λ=10 FIT ?

1 month=30×24hr=720hr 10FIT=10 failures/10^9 device ×hours

$$\frac{10\ failures}{10^9\ dev \cdot hrs} \times 720hr \times 100000dev = 0.72\ failures$$

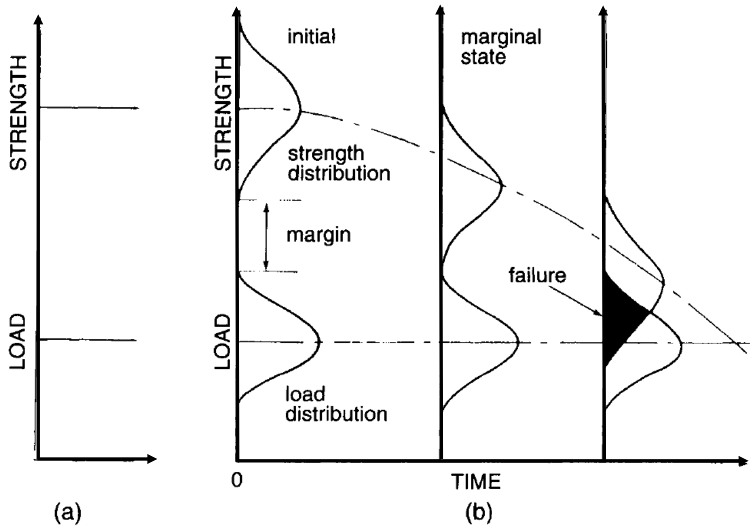
Question-2: What is the test time for 100 devices, to detect one failure:

$$t = \frac{1\ failure}{\left(\frac{10\ failures}{10^9\ dev \cdot hrs}\right) \times 100dev} = 10^6 hr = 114\ years$$

λ	Field failures (out of 100,000 devices)	Test time to detect 1 failure in 100 devices
10FIT	0.72	114yr
100FIT	7.2	11.4yr
1000FIT	72	1.14yr

→ Accelerated testing is required

Degradation vs time



Materials and Device degradation

S is a *critical* device or material parameter,

$$S(t) = S_{t=0} + \left(\frac{\partial S}{\partial t}\right)_{t=0} t + \frac{1}{2}\left(\frac{\partial^2 S}{\partial t^2}\right)_{t=0} t^2 + \dots .$$

The higher order terms in the expansion can be approximated by simply introducing a power-law exponent *m* and writing the above expansion in a shortened form:

$$S = S_o\left[1 \pm A_o(t)^m\right],$$

A_o is a material/device-dependent coefficient,
m is the power-law exponent.
Both Ao and m are adjustable parameters that can be extracted experimentally. Ao must have the units of reciprocal-time to the m-th power
So is the parameter value at t (time)=0

Materials and Device degradation vs time

Assuming reduction in critical device S:

$$S = S_o[1 - A_o(t)^m].$$

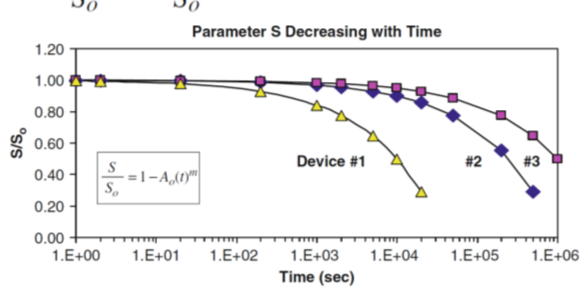
Taking the LAN of both sides, we get:

$$\ln(S^*) = m\,\ln(t) + \ln(A_o)$$

Where:

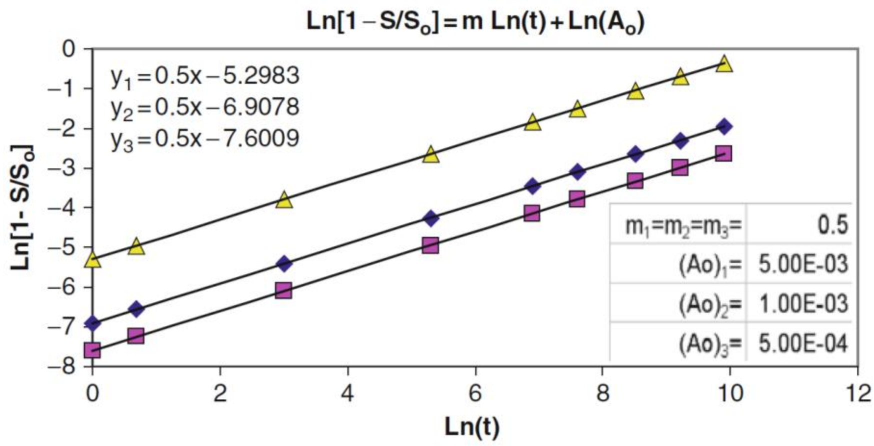
$$S^* = 1 - \frac{S}{S_o} = \frac{S_o - S}{S_o}.$$

We can easily extract So, but not Ao and *m*.



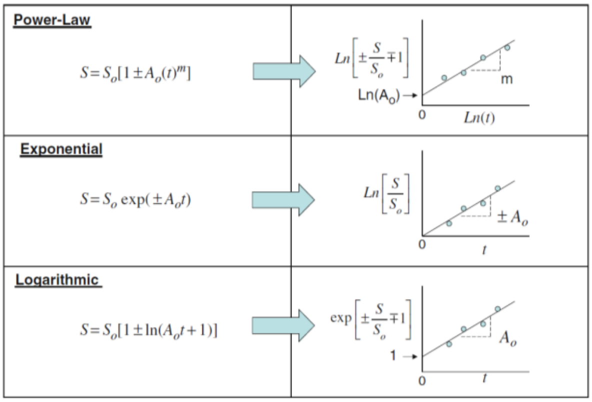
Materials and Device degradation vs time

If we used the log definition of S for drawing the degradation, we can easily extract *m* and Ao

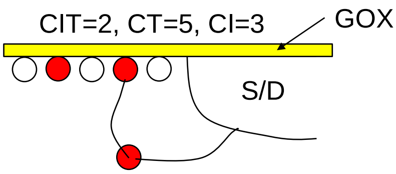


General Time-dependent Degradation Models

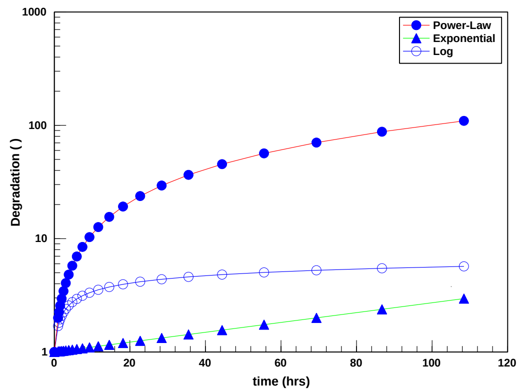
- There are many time-dependent forms for degradation.
- The following three forms is generally used:
  - The power-law: more frequently used
  - The exponential,
  - logarithmic.



General Time-depended Degradation Models



$$\frac{dC_{IT}}{dt} = K_r C_{IT} \text{-} K_C C_I C_T$$



CIT: concentration of filled sites
Kr – release rate, KC – capture rate

A0	1	0.01	1
m	1	1	1
So	1	1	1
Time	Power-Law	Exponential	Log

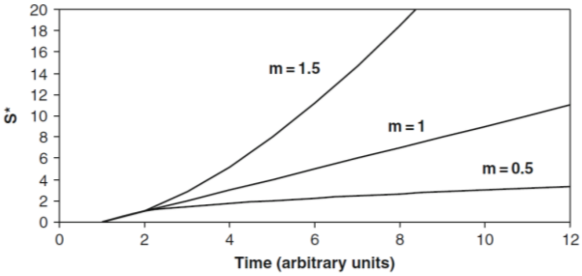
Degradation rate modeling

For power-law modeling:

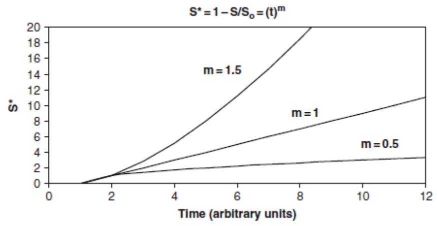
- Assume that the critical parameter S is decreasing with time
- Ao=1,
- S* is the relative change of the critical parameter,
- R is the degradation rate,

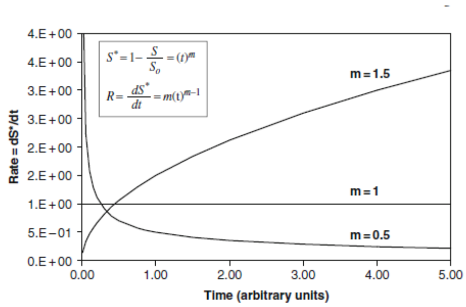
$$S^* = 1 - \frac{S}{S_o} = (t)^m. \longrightarrow R = \frac{dS^*}{dt} = m(t)^{m-1}.$$

m>1 means catastrophic degradation



Degradation rate modeling





33

Delay in the start of Degradation

At some cases, materials/devices will be remarkably stable for a period of time t0 and then show relatively rapid degradation with time.

Example: a stable resistance of a metal conductor until a void starts to form. We will learn more on that at the Electromigration.

+ means: "S increase with time"

$$S = S_0 \qquad (for\ t \le t_0)$$
$$S = S_o[1 \pm A_0(t - t_0)^m] \qquad (for\ t \ge t_0).$$

$$R_1 = \frac{dS}{dt} = 0 \qquad (for\ t \le t_0)$$
$$R_2 = \frac{dS}{dt} = (\pm)\, mS_oA_0(t - t_0)^{m-1} \qquad (for\ t \ge t_0).$$

If m>1, than R2 goes to zero at t=t0

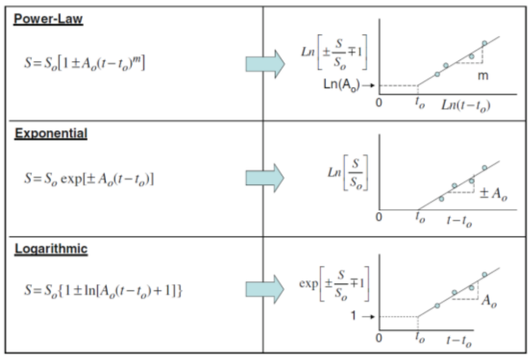
If m=1, than R2 is a constant,

If m<1, than R2 goes to infinity at t=t0

35

Delay in the start of Degradation

- The rate/slope R of the degradation can be used to find the time delay t0.
- Plot the degradation rate R versus time t → the time at which the rate R goes to zero, or R goes to infinity, is t = t0.
- If R goes to zero, or infinity, at t=0, then t0=0 and a time delay is **not needed** in the degradation equation.



36

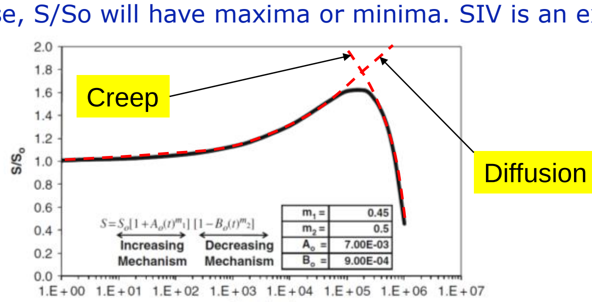
Competing Degradation Mechanisms

- One mechanism INCREASE the critical parameter vs time, and the other DECREAE the parameters vs time,

$$S = S_o\left[1 + A_o(t)^{m_1}\right]\left[1 - B_o(t)^{m_2}\right],$$

Increase vs time   Decrease vs time

- In such case, S/So will have maxima or minima. SIV is an example



39

Competing Degradation Mechanisms

- Below is a useful method for separating the problem into two regimes.

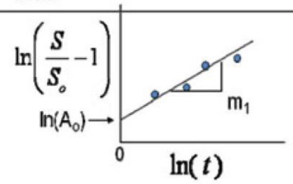
$$S = S_o\left[1 + A_o(t)^{m_1}\right]\left[1 - B_o(t)^{m_2}\right]$$

Limiting Conditions: Strengthening Term / Weakening Term

1. **During Early Stages:**

$$S \cong S_o\left[1 + A_o(t)^{m_1}\right]$$

$$\Rightarrow \ln\left(\frac{S}{S_o} - 1\right) = \ln(A_o) + m_1 \ln(t)$$

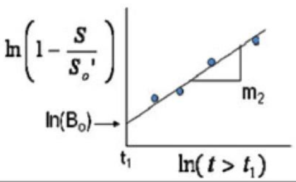


2. **During Later Stages:**

$$S \cong S_o'\left[1 - B_o(t)^{m_2}\right]$$

$$\text{where}: S_o' = S_o\left[1 + A_o(t_1)^{m_1}\right]$$

$$\Rightarrow \ln\left(1 - \frac{S}{S_o'}\right) = \ln(B_o) + m_2 \ln(t)$$



40

Quality and Reliability - definitions

Quality

- Guarantee that the IC performs its function (meeting IC specifications) at t=0,
- Driven by defects/faults during manufacturing
- Follow manufacturing procedures

Reliability

- Guarantee that the IC performs its function for $0 \le t \le$ lifetime
- Meeting specifications over time → time-dependent, aging
- Driven by changing materials/device properties, application (mission) profile, environmental conditions

44

Reliability - terminology

- Failure = Termination of the ability of an item to perform a required function within previously specified limits.
- Failure mechanism – The physical, electrical, chemical, thermal or any other process which leads to a failure.
- **Examples:** Hot carrier injection, Electromigration, stress voiding, oxide breakdown, corrosion, mechanical stress build-up, negative bias thermal instability,
- Failure mode – Consequence of the mechanism through which the failure occurs
- **Examples:** Internal short or open, threshold voltage shift, junction leakage

45

Reliability-related Terms

Intrinsic failures

- Inherent in the design and materials used
- Managed by ensuring that they occur beyond useful life of a product, by limiting loads that drive the failure mechanisms (e.g. maximum field, maximum current density, thermal management)

Extrinsic failures:

- Due to process or manufacturing defects or to misapplications such as overload, EOS, ESD, etc.
- Managed by improving manufacturing process and reducing defect density.
- Usually, product reliability is determined by extrinsic failures.

46

Reliability-related Terms

Sudden failures:

- Sudden change in a characteristics
- Examples: Oxide breakdown, interconnect open due to electromigration

Gradual failures (parametric drift):

- Caused by a parameter that drifts out of specification
- Examples: Threshold voltage or transconductance shift of a MOSFET due to hot carrier degradation

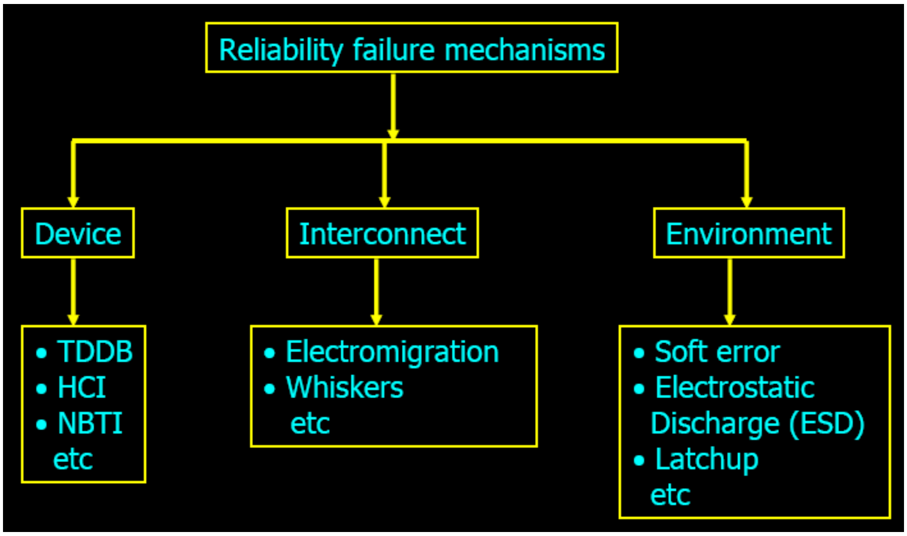
Reliability-related Terms

Redundancy - The duplication of critical components or functions of a system with the intention of increasing the reliability of the system, usually in the form of a backup or fail-safe, or to improve actual system performance.

- Redundancy sometimes produces less, instead of greater reliability – it creates a more complex system which is prone to various issues, it may lead to human neglect of duty, and may lead to higher production demands which by overstressing the system may make it less safe

Technical data contained herein are proprietary information of TOWER SEMICONDUCTOR LTD / Eitan Shauly which shall be treated confidentially, and shall not be furnished to third parties or made public without prior written permission by TOWER SEMICONDUCTOR LTD / Eitan Shauly.

Reliability Failure Mechanisms - Classification



TOWER SEMICONDUCTOR LTD.

51

Yield vs. Reliability

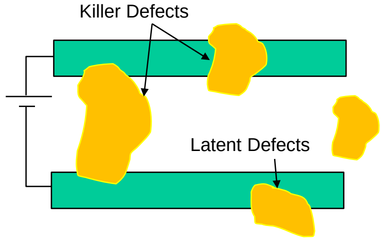
- Yield = probability of failure of an as-processed device, $t=0$
- Reliability is defined as functional failure of the device during its operation ($t>0$),
  - Yield loss is caused by KILLING defects.
  - Reliability loss is caused by LATENT defects
- A process with low yield (due to various extrinsic defects) is unacceptable to begin with, but even a process with high yield (low initial defects) but relatively large degradation rates (poor reliability) is unacceptably expensive in the long term.
- For microelectronics systems, reliability of various components is an issue of major interest since the microprocessors or memories are expected to function without failure for a long period of time (e.g. 10 years) under extreme operating conditions.

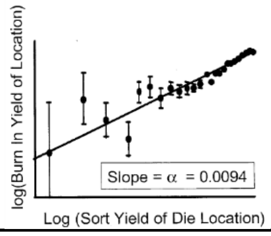
TOWER SEMICONDUCTOR LTD.

52

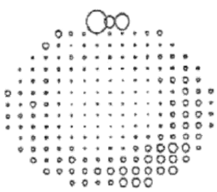
Reliability vs. Yield

- Particles can have Yield, Reliability, or no signature depending on their size and position.
- This usually exhibits correlation of good yield to reliability
- Fail map at sort identical to Burn in failure map. Failure Analysis found same failure mechanism

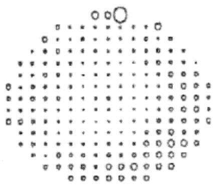




Sort Wafer Map for specific Fail Bin



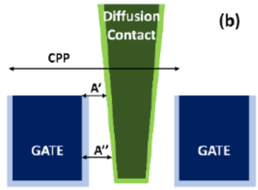
Burn In Failures Wafer Map



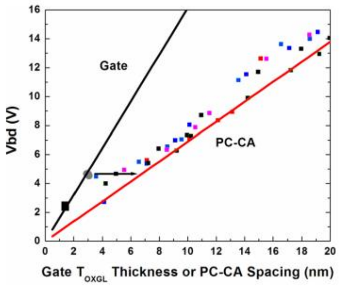
53

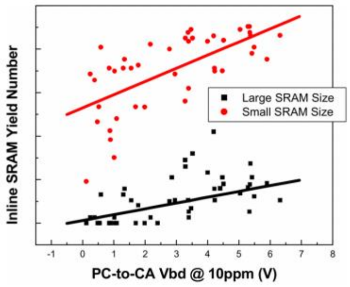
Reliability vs. Yield (example: CS.D.1)

- CS.D.1: Distance from CS (over AA) to related Gate
- CS.D.1 scale by ~x0.7 for any new generation
- Sensitive to: CS size, Gate size, Dielectric material, CA/Poly Misalignment



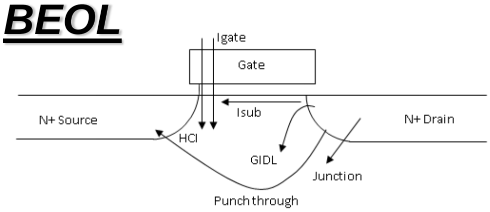
[79] Chen, *et al.*, IRPS 2012

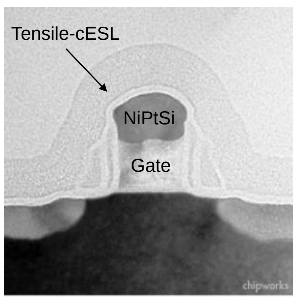




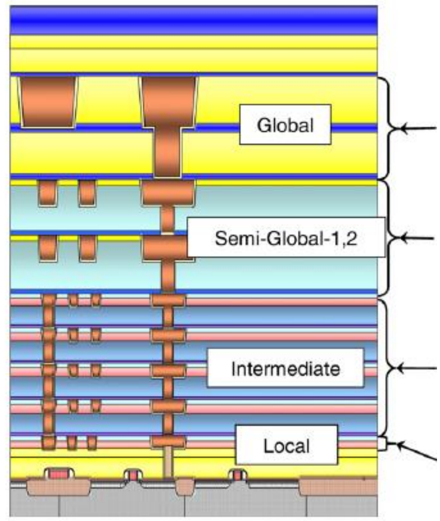
54

Typical x-section of Planar MOSFET, Typical BEOL



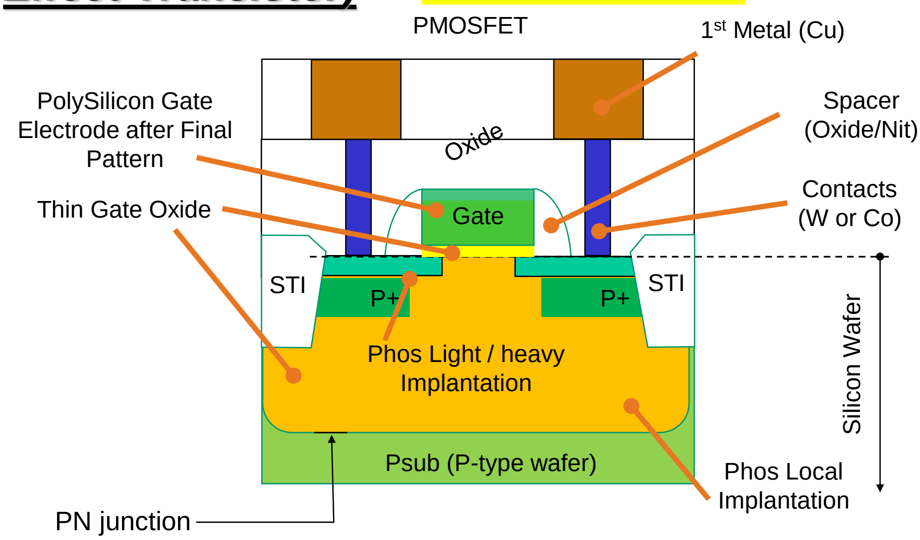


TSMC 28nm LP TEM of NMOSFET

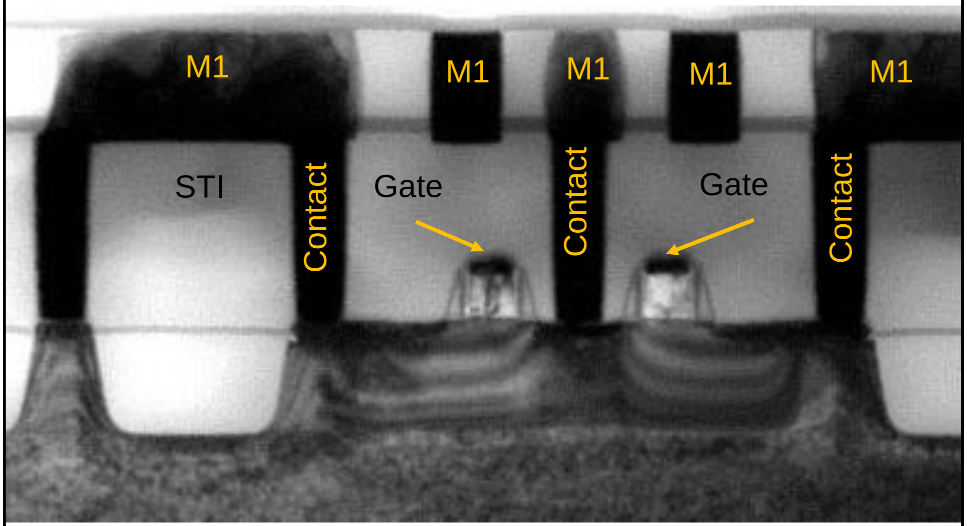


MOSFET (Metal-Oxide-Semiconductor-Field-Effect-Transistor)

Technologies: 10um ~ 20nm

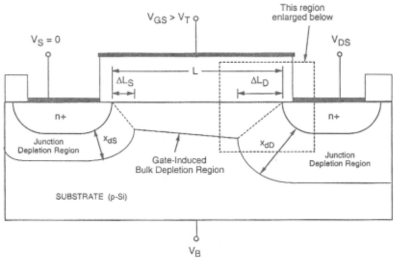


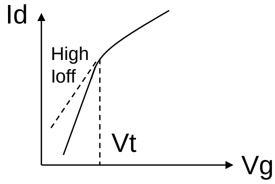
MOSFET (Metal-Oxide-Semiconductor-Field-Effect-Transistor)

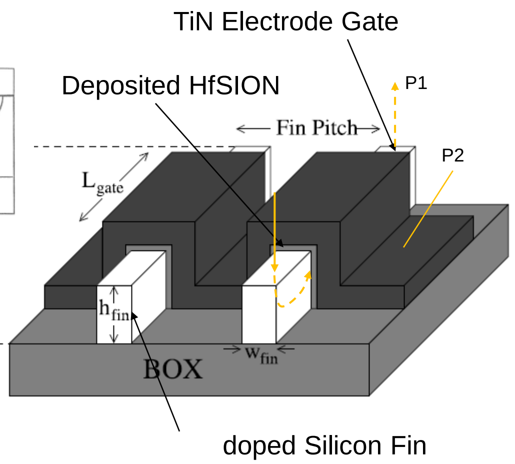


FinFET – Overall Structure (1/2)

Technologies: 16nm to 3nm



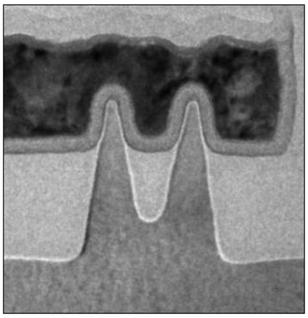


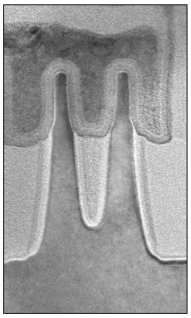


- SCE (Short Channel Effects), SS, GIDL, DIBL, SILC limitations

Technical data contained herein are proprietary information of TOWER SEMICONDUCTOR LTD / Eitan Shauly which shall be treated confidentially, and shall not be furnished to third parties or made public without prior written permission by TOWER SEMICONDUCTOR LTD / Eitan Shauly.

Intel 22nm and 14nm FinFETs – "Scaling"





After: Rani Borkar et al., "Advancing Moores Law on 2014 (Intel, 2014)

22 nm 1st Generation Tri-gate Transistor

14 nm 2nd Generation Tri-gate Transistor

(1) Tighter Fin Pitch for higher density

(2) Taller and thinner Fins to increase drive current and performances

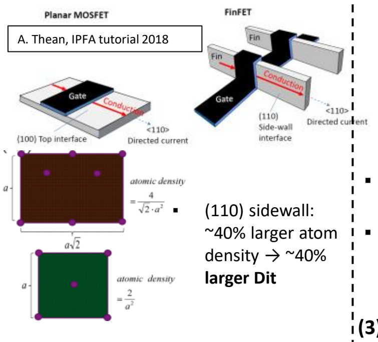
(3) Reduced number of fins for improved Density and Lower Capacitance

TOWER SEMICONDUCTOR LTD.

64

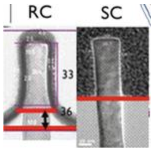
FinFET main reliability concerns

(1) Fin orientation



- (110) sidewall: ~40% larger atom density → ~40% **larger Dit**

(2) Fin shape/width/height

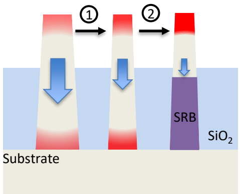


- Fin width/height has an impact on BTI/HCD?
- Round or sharp corner make an impact on reliability (field crowding)?

(3) Variability (time$_0$ and time-dependent)

- Smaller area, full depleted devices → impact on variability?

(4) Self-Heating Effects

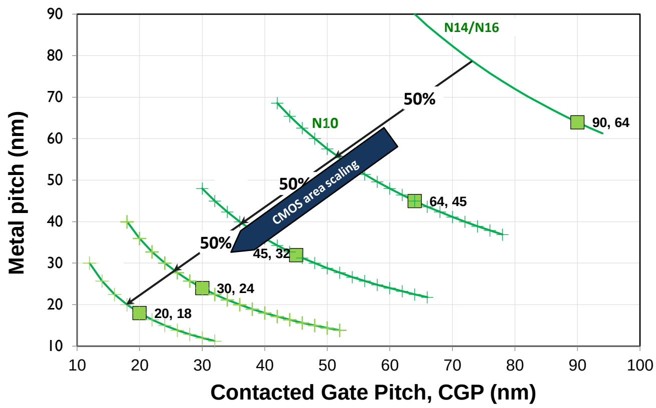


- R_{TH} (thermal resistance) increases due to smaller fin slab for heat dissipation towards bulk
  - Significant impact if alternative channel materials are chosen

TOWER SEMICONDUCTOR LTD.

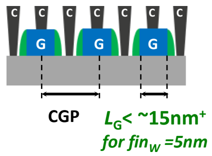
65

A view on the CMOS scaling roadmap



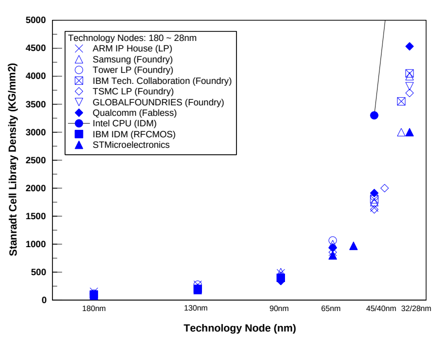
Main challenges:
(1) Photo lithography
(2) CGP:

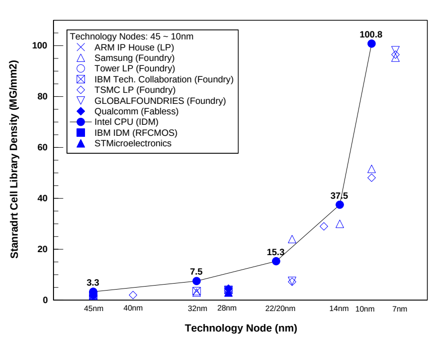
- Gate Lenth
- Spacer width
- Contact width
- → GAA



- Each of these lines represent constant area (=logic density) defined as: metal pitch x CGP
- From generation-to-generation, both MP and CGP needs to scale by x0.5

FEOL and M1 Scaling – Std cell gate density

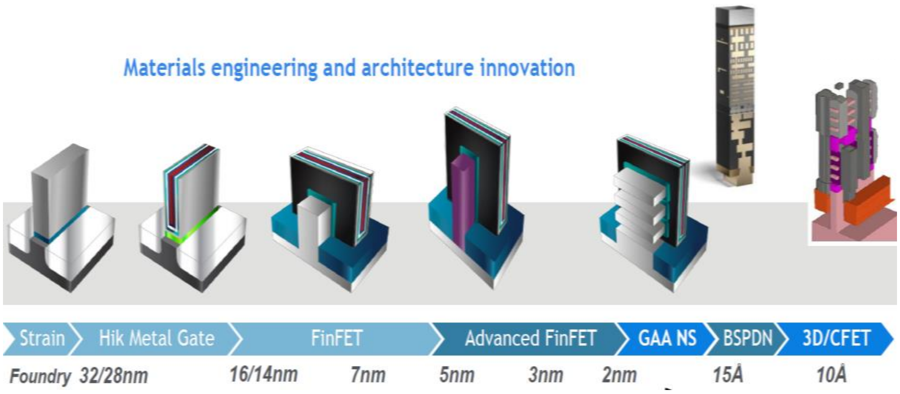




After: *Design Rules in a Semiconductor Foundry* Edited by Eitan N. Shauly, 2022 Jenny Stanford Publishing

Logic CMOS Device Evolution

- ~20 Years of innovations



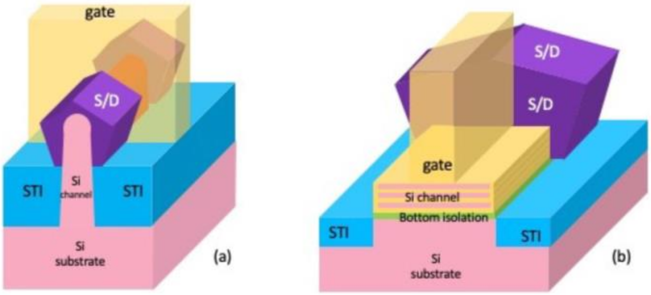
After: Naomi Yoshida, SC-1 VLSI 2023.

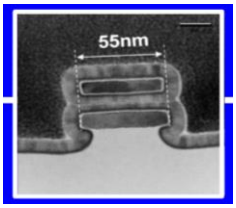
68

Gate-All-Around

Technologies: <3nm

- Stack of Nanosheets (NS) deliver the current from source to drain





- FinFET "tri-gate" to GAA:
  - Improved electrostatic control
  - Stacked devices offer improved performance per footprint

69

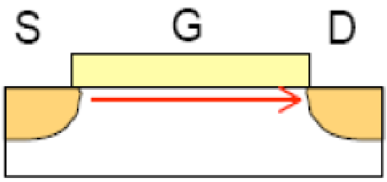
Material Reliability

- Semiconductor Substrate – Sustain vs Strong Electrical fields, that can leads to junction breakdowns, very high currents and avalanche breakdown.
- Insulators (Dielectrics) - Sustain vs Strong Electrical fields, that can leads to dielectric leakage and dielectric breakdown. For example, Gate oxide breakdown due to string Vgs.
- Metals (in BEOL) - Metal layers are used as contacts, electrodes, resistors, and interconnect. High current densities can cause to electromigration.

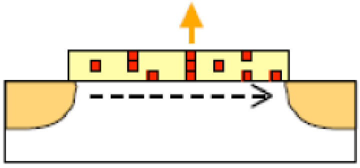
70

MOS Failure Mechanisms - Overview

- Gate oxide degrade with the passage of time due to charge injection
  - NMOS and PMOS: Hot carrier degradation (HCI)
- Broken Si-H bonds (at the interface)
  - PMOS: Negative Bias Temperature Instability (NBTI)
- Broken Si-O bonds and creation of traps
  - Time Dependent Dielectric Breakdown (TDDB)



Initially

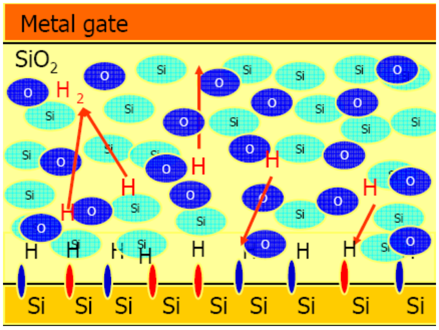


After some time

72

Device Failures at FEOL - NBTI

- NBTI degrades PMOS transistor at
  - Negative gate potential (high gate field)
  - Elevated temperature
- The gate potential breaks Si-H bonds and create positive charges at the interface
- The charges effect the threshold voltage, disrupting device characteristics
- Results
  - Threshold voltage increases
  - Drain current decreases

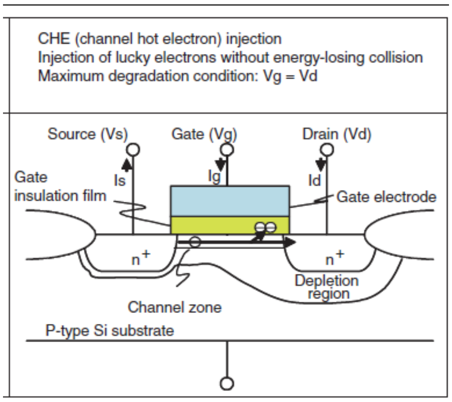


Curtsy: Dr. Efraim Aharoni

73

Device Failures at FEOL - HCI

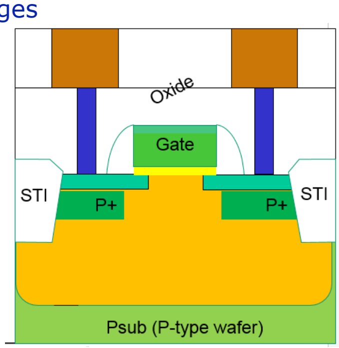
- HCI affects NMOS and PMOS transistors
- Channel carriers get energy by impact ionization
  - Most of the carriers are collected at the drain, but some are directed toward gate or substrate
  - Energetic carriers breaks bonds in
    - Gate oxide
    - Silicon substrate interface
- Results
  - Threshold voltage increases
  - Saturation current decreases
  - Leakage current increases



74

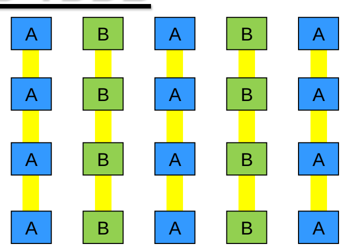
Device Failures at FEOL/BEOL - TDDB

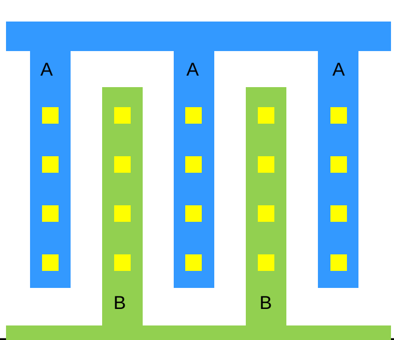
- TDDB affects both PMOS and NMOS transistors/capacitors,
- Gate dielectric suffers from short circuit - Percolation current increase with time, charges inside gate oxide align and results in a continuous path
- Direct tunneling - Charges jump directly across the oxide layer
- Accelerating factors: electric field (voltage), temperature
- Results
  - Oxide BD and loss of gate voltage control
  - Increase in gate leakage current

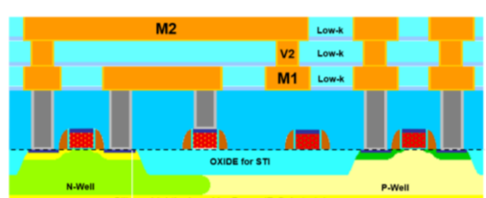


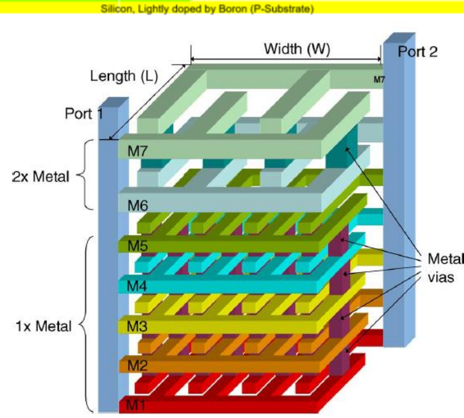
75

IMD-TDDB





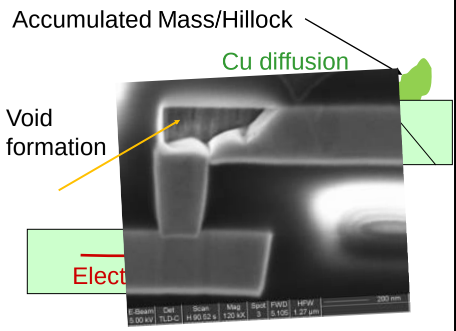




77

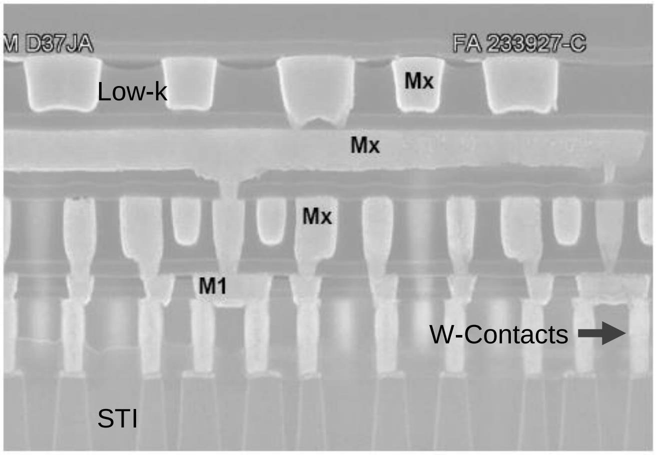
Interconnect Failures – Electromigration (EM)

- EM affect Al and Cu interconnects,
- Mass transport of metal atoms (migration) due to
  - Momentum exchange between electrons and ions, combined with ions vibration
  - Acceleration factors: Temperature, current density, grains size, mechanical stress
  - Prevention: Design rules, materials/process
- Results
  - Void formation
    - High resistance
    - Open circuits (disconnects)
    - Shorts



78

Example: 28nm Cu technology

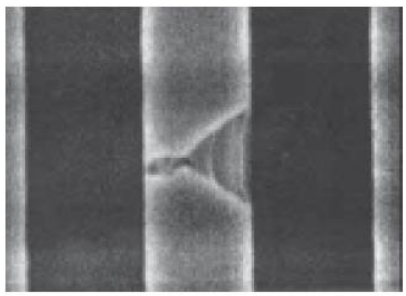


After: F. Arnaud et al., "Competitive and Cost Effective high-k based 28nm CMOS Technology for Low Power Applications," IEDM 2009 (IBM/ST/Infinion/Charter/GF/Samsung/Toshiba/NEC)

80

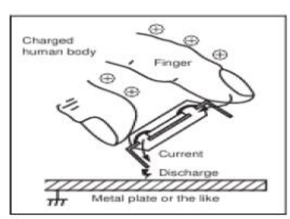
Interconnect Failures – Stress Migration (SM)

- SM affect Al and Cu interconnects
  - Happens without current (and if w/ current – so enhance EM)
  - Due to mismatch of thermal coefficients
  - Accelerating factor - temperature

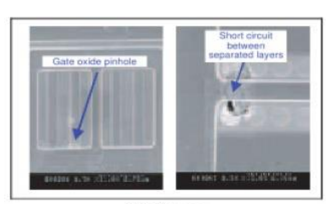


Electrostatic Discharge (ESD) – circuit failure

- ESD is a common phenomena,
  - observed by rubbing two object that creates charges (triboelectric effect)
- May occur at IC or system level
- ESD may damage devices or interconnects
- Sources of charges
  - Human body
  - Manufacturing machines
  - Charged devices
- Results
  - Transistor junctions burnout
  - Metallization burnout
- Prevention
  - Antistatic measures
  - ESD-protection circuits (within the IC)



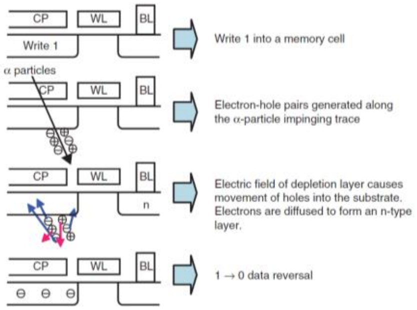
Discharge Model with Human Body



ESD Damage

Soft Errors

- Cause by alfa particles,
- Non-destructive change in CMOS based storage devices
- Ionization impacts a single/multiple cell/register/FF, causing change of state
- The change is
  - Random
  - Non recurring.
  - Single and multi-bit
- Sources of the damage are
  - Alpha particles
  - Neutrons
  - Radiations
  - Temperature



84