



Semiconductor Reliability

1. Semiconductor Device Failure Region

Below figure shows the time-dependent change in the semiconductor device failure rate. Discussions on failure rate change in time often classify the failure rate into three types of early, random and wear-out failure regions (the so-called “bathtub” curve). However, there is no clear definition for determining the boundary between these regions.

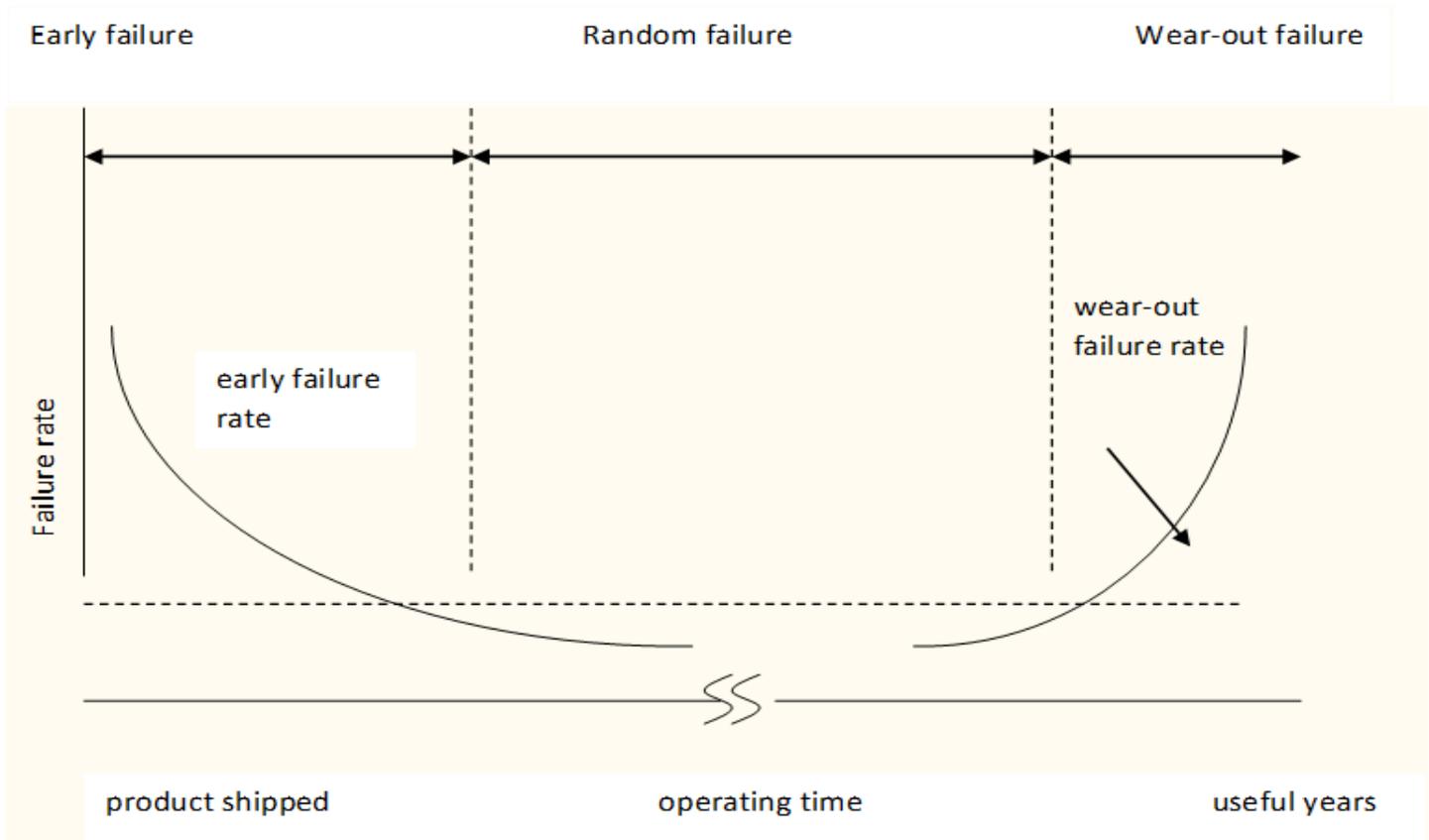


Figure 1: Time-Dependent Changes in Semiconductor Device Failure Rate

1.1 Early Failures

The failure rate in the early failure period is called the early failure rate (EFR), and exhibits a shape where the failure rate decreases over time. The vast majority of semiconductor device's initial defects belong to those built into devices during wafer processing.

While most of these defects will be eliminated in the final sorting process, a certain percentage of devices with relatively insignificant defects may not have failed and may be shipped as passing products. These types of devices are inherently defective from the start and apt to fail when stress (voltage, temperature, etc) is applied for a relatively short period, and exhibit a high failure rate in a short time within the customer's mounting process. However, these inherently defective devices fail and are eliminated over time, so the rate at which early failures occur decreases.

This property of semiconductor devices where the failure rate decreases over time can be used to perform screening known as "burn-in" where stress is applied for a short time in the stage before shipping to eliminate devices containing early defects. Products screened by burn-in not only improve the early failure rate in the market, but also make it possible to maintain high quality over a long period as long as these products do not enter the wear-out failure region.

1.2 Random Failures

When devices containing early defects have been eliminated to a certain degree, the early failure rate becomes extremely small, and the failure rate exhibits a constant level over time. In this state, the failure distribution is close to an exponential distribution, and this is called the random failure period.

The device failure rate is normally at a level that can be ignored for the most part. Most of the failures are due to devices containing relatively insignificant early defects (dust or crystal defects) that fail after a long time or random failures such as the memory soft error by alpha particles and other high-energy radioactive rays.

1.3 Wear-out Failures

Wear-out failures are failures rooted in the durability of the materials comprising semiconductor devices and the transistors, metal lines, oxide films and other elements. In this region, the failure rate increases with time until ultimately all the devices fail or suffer characteristic defects.

The main wear-out failure mechanisms for semiconductor devices are as follows:

- 1) Time-dependent dielectric breakdown (TDDB)
- 2) Hot carrier-induced characteristics fluctuation
- 3) Electromigration

Semiconductor device life is defined as the time at which the cumulative failure rate for the wear-out failure mode reaches the prescribed value, and is often determined by the reliability of each element comprising the device during the process development stage. These evaluation results are incorporated into design rules in the form of allowable stress limits to suppress wear-out failures in the product stage and ensure long-term reliability.

2. Failure Rate Calculation

Two functions are often used in the evaluation of reliability: probability density function (pdf) of failure $f(t)$ and failure rate $\lambda(t)$.

$f(t)$ denotes the probability of a device failing in the time interval dt at time t . It is related to the Cumulative Distribution Function (CDF), $F(t)$, as $f(t) = dF(t)/dt$.

On the other hand, failure rate $\lambda(t)$ is defined as the instantaneous failure rate of devices having survived to time t . Using the concept of conditional probability, $P(B|A) = P(B \text{ and } A \text{ both occur})/P(A)$, it can be derived that $\lambda(t)$ equals $f(t)/R(t)$ as shown below.

$$\text{instantaneous failure probability} = \frac{\text{fail in next } \Delta t}{\text{survive to } t} = \frac{F(t + \Delta t) - F(t)}{R(t)}$$

$$\text{instantaneous failure rate} = \lambda(t) = \frac{\lim_{\Delta t \rightarrow 0} \left[\frac{F(t + \Delta t) - F(t)}{\Delta t} \right]}{R(t)} = \frac{f(t)}{R(t)}$$

In the following discussion, the failure rate calculation is described according to the stages of product lifespan.

2.1 Methods for Estimating the Early Failure Rate

Weibull distribution is applied to approximate the CDF of early failure period; it can exhibit a shape where the failure rate decreases over time.

Weibull distribution is characterized by two important parameters, scale factor (α) and shape factor (β). They are defined as:

$$F(t) = 1 - \exp \left[- \left(\frac{t}{\alpha} \right)^\beta \right] = 1 - R(t)$$

where t : life cycle or life time (EFR duration to failure)

α : scale factor or characteristic function

β : shape factor or shape parameter

Rearranging the equation, one obtains:

$$1 - F(t) = \exp \left[- \left(\frac{t}{\alpha} \right)^\beta \right]$$

$$\ln[1 - F(t)] = - \left(\frac{t}{\alpha} \right)^\beta$$

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

When plotted in $\ln[-\ln(1 - F(t))]$ against t on log scale (Below figure), the data should approximately fall on a straight regression line. Scale factor α can be obtained from the intercept of the straight line; it is constant for a fixed test condition. Shape factor β is the slope of the straight line and its value is less than one for early failure period.

Using the relationships that $f(t) = dF(t)/dt$ and $\lambda(t) = f(t)/R(t)$, one can derive the failure rate as:

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

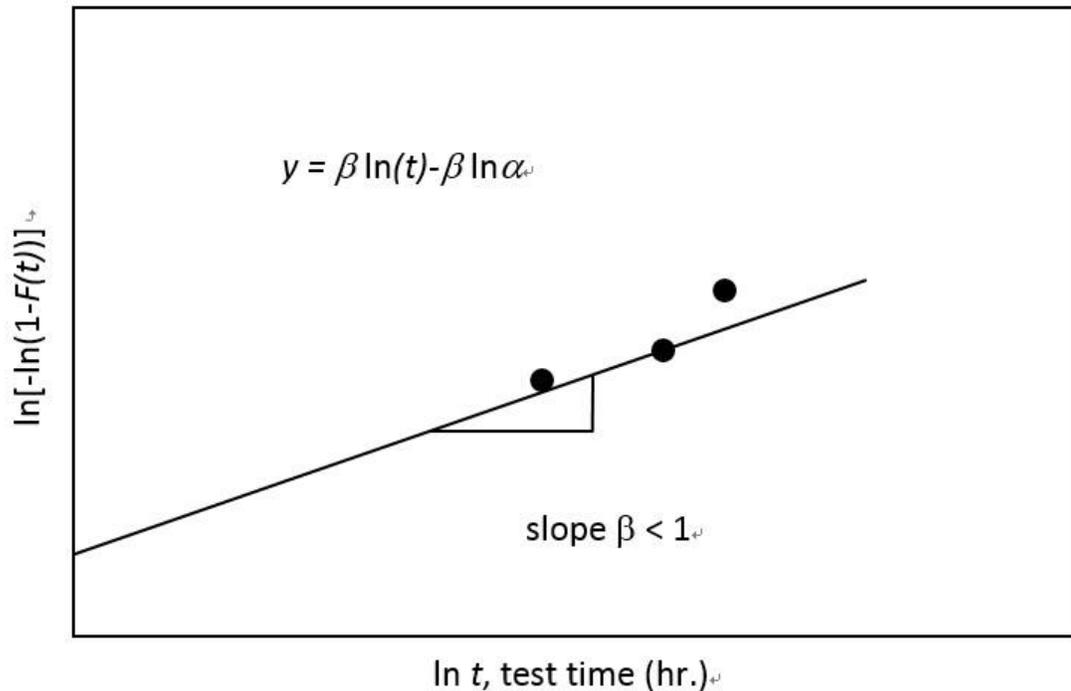


Figure 2 - $\ln t, \text{ test time (hr.)}$ VS $\ln[-\ln(1-F(t))]$

2.2 Methods for Estimating the Useful Life Failure Rate

When devices containing early defects have been eliminated to a certain degree, the initial failure rate becomes extremely small, and the failure rate exhibits a gradually declining curve over time. The failure rate at this period is obtained by dividing the number of failures observed by the device hours, usually expressed as failures per billion device hours (FITs). This is called point estimate because it is obtained from observation on a portion (sample) of the population of devices. In order to determine the unknown population parameter from known sample statistics, it is necessary to make use of specific probability distribution. The chi-square distribution (χ^2) that relates observed and expected frequencies of an event is frequently used. The relationship between failure rate at stress conditions and the chi-square distribution is shown in the following formula:

$$\lambda_{stress} = \frac{\chi^2(\alpha, n)}{2t}$$

where:

λ_{stress} = failure rate at stress conditions

χ^2 = chi-square function

α = (100 - confidence level) /100

n = degree of freedom = $2r + 2$

r = number of failures

t = device hours

The chi-square values for α equal to 60% and 90% at r between 0 and 12 are shown in below table.

Chi-Square Distribution Function			
60% Confidence Level		90% Confidence Level	
No. Fails	χ^2 Quantity	No. Fails	χ^2 Quantity
0	1.833	0	4.605
1	4.045	1	7.779
2	6.211	2	10.645
3	8.351	3	13.362
4	10.473	4	15.987
5	12.584	5	18.549
6	14.685	6	21.064
7	16.780	7	23.542
8	18.868	8	25.989

Since all the reliability tests are performed under accelerated stress condition, it is important to evaluate the acceleration factor of different stresses.

1) Thermal Acceleration Factor

Acceleration factor for thermal stress is calculated using the Arrhenius equation:

$$AF_t = e^{\left[\frac{E_a}{k} \left(\frac{1}{273 + T_{use}} - \frac{1}{273 + T_{stress}} \right) \right]}$$

where: AF_t = thermal acceleration factor

E_a = activation energy in electron Volts (eV)¹,

k = Boltzmann's constant 8.62×10^{-5} eV / °K

T_{use} = junction temperature at normal use condition in °C

T_{stress} = the stress temperature in °C

2) Voltage Acceleration Factor

High electrical field can cause physical damage in the oxide layers. The acceleration factor due to voltage stress is a function of both the stress voltage, V_{stress} , and the wafer process. AF_v can be derived from Eyring model as:

$$AF_v = e^{\gamma(V_{stress} - V_{use})}$$

¹ The activation energy is defined as the excess free energy over the ground state that must be acquired by an atomic or molecular system in order that a particular process can occur. Examples are the energy needed by the molecule to take part in a chemical reaction, by an electron to reach the conduction band in a semiconductor, or by a lattice defect to move to a neighboring site. (Ref. *Van Nostrand's Scientific Encyclopedia*) ISSI's failure rate calculations are based on acceleration from high temperature where thermal activation energy, E_a , is typically identified in JEDEC Standard.

where: AF_v = voltage acceleration factor

γ = constant in $1/V$

V_{stress} = stress voltage

V_{use} = use voltage

3) Humidity Acceleration Factor

For humidity acceleration test, the acceleration factor can be estimated by

$$AF_h = \left(\frac{RH_{stress}}{RH_{use}} \right)^m$$

where : RH_{stress} = relative humidity in stress

RH_{use} = relative humidity in use (typical value 17.6%)

m = experimentally determined (typical value $m = 3$)

If temperature is included in the humidity test, both temperature and humidity acceleration factor need to be considered.

In THB test (Temp/Humidity/Bias), the voltage acceleration factor must be added. The junction heating effect can reduce the relative humidity. For example, a 5°C junction heating effect by bias can reduce the RH from 85% to 73%.

4) Temp Cycling Acceleration Factor

The acceleration factor can be estimated by

$$AF_{tc} = \left(\frac{T_{\max, stress} - T_{\min, stress}}{T_{\max, use} - T_{\min, use}} \right)^n$$

where : $T_{\max, stress}$ = high temp in cycling stress

$T_{\min, stress}$ = low temp in cycling stress

$T_{\max, use}$ = high temp in field application, usually 70°C

$T_{\min, use}$ = low temp in field application, usually 0°C

n = experimentally determined, usually $n = 5$

5) MTTF at Use Conditions

The mean-time-to-failure (MTTF) is defined as the average time-to-failure (or expected time-to-failure) for a population of devices, when operating its required function under the specified conditions for a stated period of time. It can be expressed by:

$$MTTF = \int_0^{\infty} tf(t)dt$$

The overall relationship of $F(t)$, $R(t)$, $f(t)$, $\lambda(t)$ and MTTF can be depicted as below figure.

$$R(t) + F(t)$$

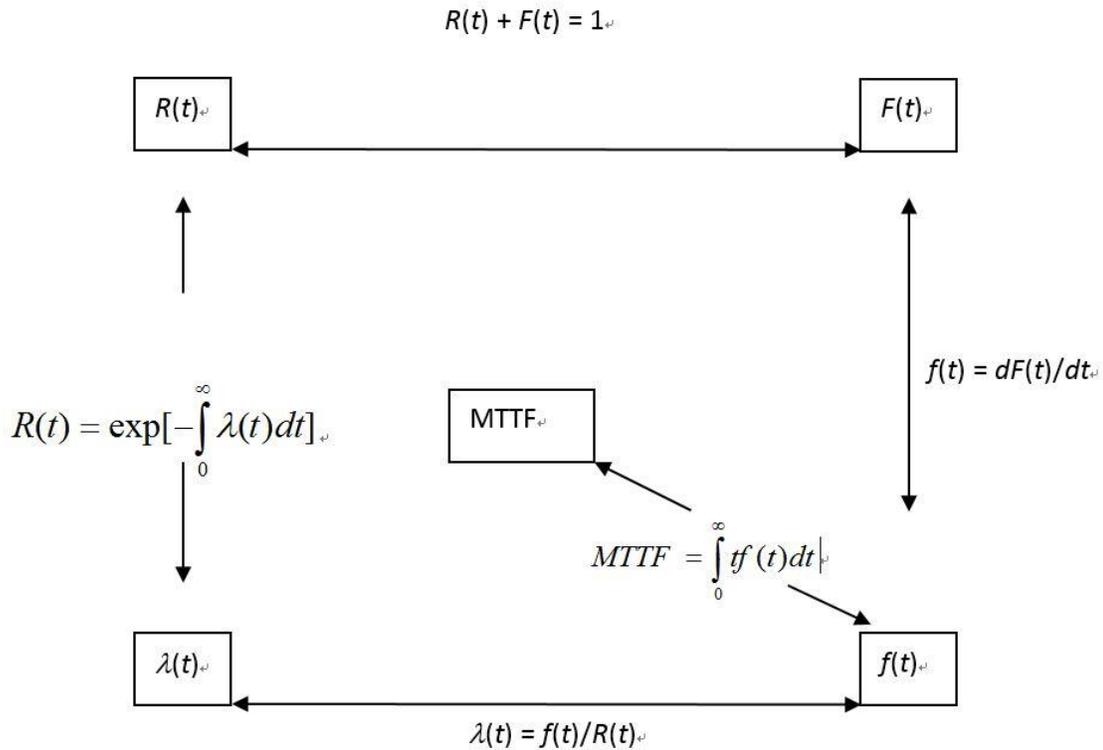


Figure 3 - The overall relationship of $F(t)$, $R(t)$, $f(t)$, $\lambda(t)$ and MTTF

For the case of constant failure rate ($\lambda(t) = \text{constant}$), $R(t) = \exp(-\lambda t)$, $F(t) = 1 - \exp(-\lambda t)$ and $f(t) = dF(t)/dt = \lambda \exp(-\lambda t)$

$$MTTF = \int_0^{\infty} tf(t) dt = \int_0^{\infty} t\lambda \exp(-\lambda t) dt = \frac{1}{\lambda_{use}}$$

where the point estimate of the failure rate at use conditions is calculated as:

$$\lambda_{use} = \frac{\lambda_{stress}}{AF_t \times AF_v}$$