A Statistical Model for Electromigration Failures†

Gilbert Yoh and Farid N. Najm

Integrated Circuits Business Division
Agilent Technologies
Fort Collins, Colorado 80525, USA

ECE Department
University of Toronto
Toronto, Ontario M5S 3G4, Canada

Abstract

The lognormal has been traditionally used to model the failure time distribution of electromigration failures. However, when used to estimate the failure of large metal layers, it leads to a clear disagreement with established empirical data. To resolve this problem, we propose to use a *shifted lognormal* (SLN) as a better model of the failure time of individual wires. We will show that the SLN is well justified because it matches other more detailed and more physical models, such as the multilognormal [1]. We will also show that the SLN exhibits the right behavior for long wires. Finally, we will provide an estimation methodology by which the parameters of the SLN can be estimated from failure data. Finally, the analysis will be extended to large metal layers where the advantages of using SLN over LN will be clearly demonstrated.

1. Introduction

Electromigration in metal lines is the movement of metal atoms due to the momentum exchange between electrons and the metal atoms [2]. The resulting atomic diffusion occurs mainly at the grain-boundaries between grain regions [1, 3–5]. Over time, the metal diffusion causes *depletion* of enough material so as to create an open circuit in a wire. Material accumulation can also occur and cause a short circuit to a neighboring wire, but in an IC this is suppressed somewhat by the layers of other material around and above the wire. Failures due to depletion have therefore received more attention, and we will be focusing only on them.

The rate of material depletion depends on the grain boundary structure of the individual metal line, and varies from wire to wire. As a result, the failure time of a metal line due to electromigration is modeled as a random variable. Even though other distributions have been proposed, the commonly used distribution for the time-to-failure of a metal line has been the *lognormal* distribution. The lognormal is easy to work with and seems to fit the data well.

However, there is a problem. If one attempts to estimate the time-to-failure distribution for a large metal layer (or a very long line) using the lognormal, then it will be found that the time to failure goes to zero. This contradicts empirical data which shows that the long wire time to failure asymptotically approaches a non-zero value. To resolve this problem, we propose to use a *shifted lognormal* (SLN) as a better model of the failure time of individual wires. We will show that the SLN is well justified because it matches other more detailed and more physical models, such as the multilognormal [1]. We will also show that the SLN exhibits the right behavior for long wires and provides a better match to chip failure times. We do not have access to actual failure times of real chips - instead, we have used a detailed physical model in which samples of the grain boundary structure are generated at random in order to investigate the failure time distributions. Finally, we will provide an estimation methodology by which the parameters of the SLN can be estimated from failure data.

The paper is organized into 6 sections. Section 2 is an overview of electromigration and introduces the shifted lognormal. Section 3 describes the multilognormal distribution. Section 4 introduces the truncated multilognormal and compares it to the shifted lognormal. Section 5 describes the parameter estimation methodology for the SLN. Section 6 gives a case study.

2. Electromigration

For an electromigration failure to occur, *flux divergence* of metal atoms must occur [3, 6]. A flux divergence occurs in a region when the flow of metal atoms into the region is not equal to the flow of atoms out of the region. Depletion occurs when flow out of a region is greater than flow into it. Flux divergence commonly occurs at *triple points*. A triple point is a site where three grain-boundaries meet, as shown in Fig. 1.

The vulnerability of a triple point to high current densities depends on the geometric structure of the triple point [4, 7]. Thus, we can view a wire

† This work was performed while both authors were with the ECE Dept. and Coordinated Science Lab., University of Illinois at Urbana-Champaign

as a series of triple points, each of which represents a possible failure site [8, 9, 1, 4]. The worst case triple point determines the failure time of the wire. Longer wires are more likely to contain worst-case triple points and thus fail sooner.

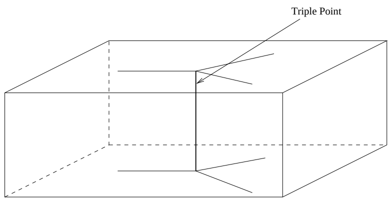

Figure 1. A triple point in a wire.

It has been known for a while that electromigration failure times seem to have a good fit to a *lognormal* (LN) distribution, i.e., its logarithm has a *normal* (Gaussian) distribution. If $\mathbf{t_{wire}}$ is the (random) failure time for a wire, then its *cumulative distribution function* (CDF) is:

$$F(t) = \Pr\{\mathbf{t_{wire}} < t\} = \Phi\left(\frac{\ln t - \mu}{\sigma}\right) \quad (1)$$

when $t > 0$ (otherwise $F(t) = 0$), where $\Phi(t)$ is the CDF for the *standard normal distribution*, and μ and σ^2 are the mean variance of $\ln(\mathbf{t_{wire}})$. In general, μ and σ depend on technology as well as the wire length, width, & current density.

The failure time for a group of wires is the minimum of the failure times of the individual wires. The wire failure times are typically assumed to be independent, so that the CDF for the group failure time is [1]:

$$F_{grp}(t) = 1 - \prod_{i=1}^{n}\left\{1 - \Phi\left(\frac{\ln t - \mu_i}{\sigma_i}\right)\right\} \quad (2)$$

If the number of wires is very large, it is easy to show that the CDF approaches a unit step function while the *probability density function* (PDF) approaches a unit impulse function. Therefore, a large group of wires should fail immediately after power up! This is in contradiction with common experience and with empirical data. It is well-known that long wires (which can be considered as large collections of unit length wires) have failure times whose mean decreases with length but does not go to zero. Instead, the long wire *mean time-to-failure* (MTF) approaches a certain minimum non-zero value. We will refer to this as the *long wire asymptote* (LWA).

2.1. The Shifted Lognormal

One obvious way to fix the lognormal distribution is by adding a shift parameter δ to the model. We call this new distribution the *shifted lognormal* (SLN) distribution:

$$F(t) = \Phi\left(\frac{\ln(t - \delta) - \mu}{\sigma}\right) \quad (3)$$

when $t > \delta$ (otherwise $F(t) = 0$), where μ and σ^2 are the mean and variance of $\ln(\mathbf{t} - \delta)$. Using the SLN distribution, the failure time distribution for a large group of wires can be shown to have a distribution whose CDF approaches the unit step function at δ and whose PDF approaches the unit impulse function centered at δ.

The parameter δ represents the minimum failure time for a wire under electromigration stress. In practice it is equal to the long wire failure time, for a given current density and temperature. But is this enough to discard the traditional lognormal in favor of the SLN? In the following, we will show that there is, in fact, sufficient physical reasons to support the use of the SLN distribution instead of the lognormal.

3. The Multilognormal

The lognormal distribution is not a *physical* model, i.e., it is not derived from the underlying physics of failure of metal wires. In [1], a more physical model was introduced, called the *multilognormal* (MLN) distribution. Since the model [1] is derived for narrow metal lines, so called "fine lines" in [1], this is also a feature of the work presented in this paper. Future extensions of this work will aim to extend these results to the general case.

Not all triple points are potential failure sites either. A lot depends on the distance between triple points, and there are differing opinions on this [1]. Nevertheless, only a finite number of points along the wire length are potential failure sites, which may be separated by distances of tens to hundreds of microns [1]. These potential failure sites create a partitioning of a wire into a series of *segments*, also called *characteristic elements* or *failure units* in [1]. Failures occur only at the junctions between segments.

For a given segment in a wire, one can define an *effective diffusion constant* as the sum of the *diffusion constants* for the grain boundaries in that segment, $D^{(n)} = \sum_{i=1}^{n} D_0 \exp\left(-\frac{\Delta H_i}{kT}\right)$, where n is the number of grain boundaries in the segment, T is the absolute temperature, k is Boltzmann's constant, and ΔH_i is the activation energy for the ith grain-boundary. The activation energy is different for different grain boundaries depending on their geometry and structure. Indeed, in [1] ΔH was modeled as a normally distributed random variable, with mean $\mu_{\Delta H}$ and variance $\sigma^2_{\Delta H}$. It reflects the random nature of grain boundaries, with variations due to angle mismatch between grains and orientation of the grain boundary with respect to current flow. Assuming a fine line, and if the wire has N segment pairs (potential failure sites), then the failure time CDF is found [1] to be:

$$F(t;N) = 1 - \left[1 - \Phi\left\{\frac{\ln t - \mu_{tot}}{\sigma_{tot}}\right\}\right]^{Np_1} \tag{4}$$

where $\mu_{tot} = \ln\left[\left(\gamma T^2/j^2\right)\exp\left(\mu_{\Delta H}/kT\right)\right]$ & $\sigma_{tot} = \sigma_{\Delta H}/kT$, and where N is the number of potential failure sites, j is the current density, p_1 is the probability that one grain-boundary is present in the segment, which depends on the distribution of grain sizes, and γ is a constant.

According to [1], N is a linear function of length. By letting N go to infinity, it is easy to show that the failure time goes to zero so that the MLN, as is the case with the LN, does not exhibit the correct LWA behavior.

4. Truncated MLN & the SLN

The MLN distribution of [1] was based on an assumption that ΔH is normal. This in effect results in a lognormal distribution for the individual grain-boundary diffusion constants, since $D^{(1)} \propto e^{-\Delta H}$. We propose that a *truncated normal* is a better distribution to use for ΔH because it is reasonable that the activation energy should have a non-zero minimum value and because it leads directly to a failure time distribution that shows the LWA behavior. Note that a lower bound on ΔH leads to an upper bound on $D^{(1)}$. There are precedents for this, such as [10], which suggests that a truncated distribution be used for the grain-boundary activation energy. In addition, in [4] the diffusion constant has a maximum value, due to a maximum that is imposed on the angle mismatch between neighboring grain lattices. Finally, in [7] the atomic flux is proportional to a value called the *structural factor* ΔY, which also has an upper bound due to bounds on the angles that are part of the grain boundary structure and due to bounds on the activation energy. Since a truncated distribution for the diffusion constant (and therefore for the activation energy) is justified, we have used the model of Lloyd and Kitchin [1] but have based it on a *truncated normal* distribution of the activation energy, rather than a normal distribution, with PDF:

$$\phi_\Delta(x) = \begin{cases} 0, & \text{if } \frac{x-\mu}{\sigma} < -\Delta \\ \frac{1}{\alpha}\phi\left(\frac{x-\mu}{\sigma}\right), & \text{otherwise} \end{cases} \tag{5}$$

where $\alpha = 1 - \Phi(-\Delta)$ and ϕ is the standard normal PDF. If we replace the normal distribution assumed in [1] with the truncated normal distribution given above and then use the same approximation and derivation as [1] (we validated that all approximations are still valid for the case of the truncated normal), the result is the following CDF for what we will call the *truncated MLN* distribution:

$$F(t;N) = 1 - \left[1 - \Phi_\Delta\left\{\frac{\ln t - \mu_{tot}}{\sigma_{tot}}\right\}\right]^{Np_1} \tag{6}$$

It is easy to show that the truncated MLN is consistent with the LWA.

We have found that the shifted lognormal and the truncated MLN are very similar. For one thing, we can show that $\delta = \exp(\mu_{tot} - \Delta\sigma_{tot})$. In Fig. 2, we show CDF plots for the shifted lognormal (SLN) and truncated MLN distributions. Since the distributions are similar and since the SLN is a simpler expression and therefore easier to work with (for one thing, it does not require one to resolve the controversy regarding the meaning of N in the truncated MLN), we propose to use the SLN to model the failure time distribution of metal lines.

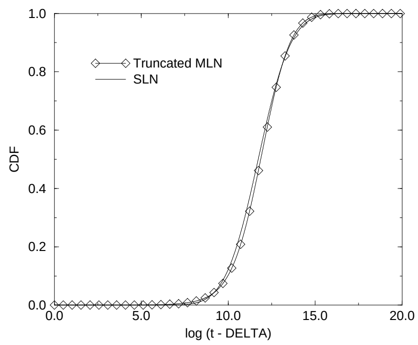

Figure 2. The SLN and the truncated MLN.

5. SLN Parameter Estimation

To get a complete SLN distribution for a wire, a direct approach would be to collect enough data in order to estimate values for the shift factor, δ, as a function of width and current density, and the mean, μ, and standard deviation, σ as functions of length, width, and current density. The first step would be to find values for $\delta(w,j)$, because it is needed to compute the random variable $\ln(\mathbf{t} - \delta)$. To measure the value of δ, we must test long wires until they fail. Then, one would find the mean and variance using standard methods of statistical estimation. This procedure would require a large number of wires to be tested in order to predict the dependence on width, length, and current density.

We propose a simplification of this based on an underlying physical failure model given in [1] and [7], as follows.

In [1], the failure time for a single triple point was found to be:

$$t_1 = \begin{cases} \frac{\gamma^* T^2}{j^2\left(D^{(2)}-D^{(1)}\right)}, & \text{for } D^{(2)} > D^{(1)} \\ \infty, & \text{otherwise} \end{cases} \tag{7}$$

where γ^* is a constant. The term $(D^{(2)} - D^{(1)})$ represents the diffusion constant for material depletion out of the triple point, but does not take into account full information about the grain-boundary microstructure of the metal. The structural factor ΔY of [7] provides a better measure of the diffusion constant at the triple point because it factors in the various angles of the geometry around the triple point. Actually, ΔY is proportional to the sum of the diffusion constants at a triple point, which necessitates a change of the proportionality constant (from γ^* to γ). Based on [7], we can write that $D^{(2)} - D^{(1)} \propto \exp\left(-H_0/kT\right) \Delta Y$ where H_0 is the average activation energy for a grain-boundary. In [1] $H_0 = 0.75eV$ is used. Therefore, the failure time of a wire with N failure sites is given by:

$$t = \begin{cases} \frac{\gamma T^2 \exp\left(\frac{H_0}{kT}\right)}{j^2 \max\limits_N\{\Delta Y\}}, & \text{for } \max\limits_N\{\Delta Y\} > 0 \\ \infty, & \text{otherwise} \end{cases} \tag{8}$$

This physical model constitutes our **golden model**. In the absence of actual failure times, we will use this model to *randomly generate* failure data for individual wires, whole metal layers, etc. To do this for a given wire, N is assumed to be Poisson distributed as in [8]. As for the ΔY, we have augmented the model of [7] by including in it the dependence on the angle mismatch between neighboring grain lattices, as was done in [4]. The several angles around a triple point are then assumed random, and uniformly distributed, as in [7].

Given the above physical model, it becomes possible to establish relationships between parameters of the SLN distribution and the underlying physical model. To begin with, we have:

$$\delta = \frac{\gamma T^2 \exp\left(\frac{H_0}{kT}\right)}{j^2 \max\limits_\infty\{\Delta Y\}} \tag{9}$$

Notice that $\max\limits_\infty\{\Delta Y\}$ is fixed while $\max_N\{\Delta Y\}$ is dependent on wire length. If we use $E[\cdot]$ to denote the expected value (mean) operator, then using (8) and (9) we get:

$$\mu = E\left[\ln(\mathbf{t} - \delta)\right] = \ln\left(\frac{\gamma T^2 \exp\left(\frac{H_0}{kT}\right)}{j^2}\right) + \hat{\mu}(l) \tag{10}$$

where $\hat{\mu}(l) = E\left[\ln\left(\frac{1}{\max\limits_N\{\Delta Y\}} - \frac{1}{\max\limits_\infty\{\Delta Y\}}\right)\right]$. We can also derive an expression for σ^2 as follows:

$$\sigma^2 = E\left[\left(\ln(\mathbf{t} - \delta) - \mu\right)^2\right] = \hat{\sigma}^2(l) \tag{11}$$

where $\hat{\sigma}^2(l) = V\left[\ln\left(\frac{1}{\max\limits_N\{\Delta Y\}} - \frac{1}{\max\limits_\infty\{\Delta Y\}}\right)\right]$ and $V[\cdot]$ denotes the variance.

Thus we see that we can relate μ and σ to the mean and standard deviation of a new random variable $\mathbf{Y_N}$, defined as follows:

$$\mathbf{Y_N} = \ln\left(\frac{1}{\max\limits_N\{\Delta Y\}} - \frac{1}{\max\limits_\infty\{\Delta Y\}}\right) \tag{12}$$

so that $\hat{\mu}(l)$ & $\hat{\sigma}^2(l)$ are the mean & variance of $\mathbf{Y_N}$.

With this, a parameter estimation methodology can be proposed, as follows. For one thing, δ can be be computed from empirical measurements of lifetimes for long wires. Then, given knowledge of the grain-boundary microstructure for a certain metal technology, one can model the angles around a triple point as uniform random variables and proceed to compute $\max\limits_\infty\{\Delta Y\}$ in a way similar to [10]. This value is required only in order to estimate γ from (9). With this, we can use random sampling (Monte Carlo simulation) to construct tables for $\hat{\mu}(l)$ and $\hat{\sigma}(l)$ as functions of l. To do this, requires knowledge of the distribution of $\mathbf{Y_N}$. Using a standard normal scores test, we have found $\mathbf{Y_N}$ to have an approximately normal distribution, so that standard techniques from [11] can be used. An important feature of this methodology is that $\hat{\mu}(l)$ and $\hat{\sigma}(l)$ are independent of j and w. Therefore, in order to generate tables for $\hat{\mu}(l)$ and $\hat{\sigma}(l)$, it is enough to test a batch of wires that have the same value of width and current density. Only the length needs to be varied.

6. Case Analysis

We will demonstrate our approach method on the power and ground interconnects for circuit C17, from the ISCAS'85 benchmark suite [12]. The circuit with the power and ground grids was laid out in a $0.35\mu m$ CMOS technology [13].

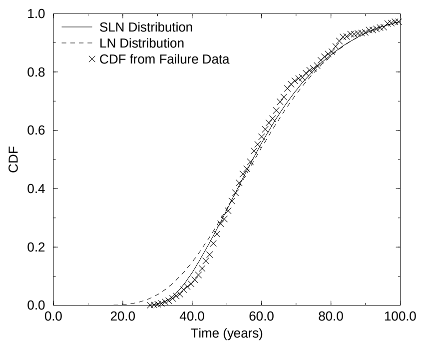

Figure 3. CDF comparisons for the failure time of the power and ground gids in C17.

To start with, we have built tables for $\hat{\mu}(l)$ and $\hat{\sigma}(l)$ as described above using our golden model. In [5], Kinsbron gives data for a 250 μm wire of varying widths. For the 8 μm width, the mean time to failure (MTF) is given to be 65 hours. We will assume that this is the LWA time δ. The current density is approximately $2\times10^6 A/cm^2$ in [5] and the temperature is $473K$. As for $\max_{\infty}\{\Delta Y\}$, we found it to be 1.137. Using these values, we found that $\gamma = 1.132 A^2/K^2 cm^4$. With this value and the $\hat{\mu}(l)$ and $\hat{\sigma}(l)$ tables, we can use (10) to get values for μ as a function of length, width, and current density for every wire. To compute the current density in every wire we used [14] to get the average gate currents then simulated the power and ground grids to find the average current density (it is justified to use just the average [15, 16]).

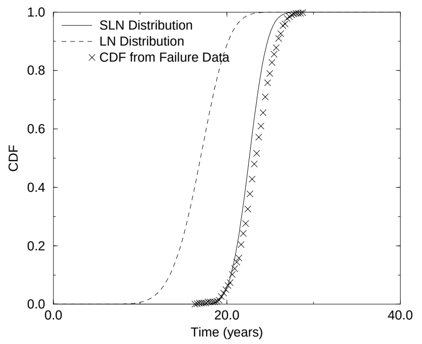

Figure 4. CDF comparisons for the failure time of a 6000 wire circuit.

Finally, we compared the CDFs resulting from three types of analyses: 1) Having characterized metal wires, using a LN distribution for failure time of each wire, combine them to get the CDF for the circuit, 2) Having characterized metal wires, using an SLN distribution for failure time of each wire, combine them to get the CDF for the circuit, and 3) Collect actual failure times for the circuit using our golden model, build the CDF for the circuit failure time. Fig. 3 compares these for circuit C17 with 60 interconnects. We see in this figure that using the SLN distribution has little effect on the resulting predicted distribution because the circuit is very small. Fig. 4 plots the same for a 6000 wire circuit. In this case, the predicted distribution using the SLN distribution is much more accurate than the CDF obtained assuming the LN distribution. This fits our expectation that the LN distribution is less accurate than the SLN distribution for larger circuits. Both simulations were at 333 K.

References

[1] J. R. Lloyd and J. Kitchin, "The electromigration failure distribution: The fine-line case," *J. Appl. Phys.* , vol. 69, pp. 2117-2127, Feb. 1991.

[2] J. R. Black, "Electromigration - a brief survey and some recent results," *IEEE Trans. on Electron Devices*, vol. ED-16, no. 4, 1969.

[3] B. N. Agarwala, M. J. Attardo, and A. P. Ingraham, "Dependence of electromigration-induced failure time on length and width of aluminum thin-film conductors," *J. Appl. Phys* , vol. 41, no. 10, pp. 3954-3960, Sept. 1970.

[4] J. T. Trattles, A. G. O'Neill, and B. C. Mecrow, "Computer simulation of electromigration in thin-film metal conductors," *J. Appl. Phys* , vol. 75, no. 12, pp. 7799-7804, Sept. 1970.

[5] E. Kinsbron, "A model for the width dependence of electromigration lifetimes in aluminum thin-film stripes," *Appl. Phys. Lett.*, vol. 36, no. 12, pp. 968-970, 1980.

[6] D. G. Pierce and P. G. Brusius, "Electromigration: a review," *Microelectron. Reliab*, vol. 37, no. 7, pp. 1053-1072, 1997.

[7] A. Christou, *Electromigration and Electronic Device Degradation*, pp. 27-77, 1994.

[8] T. Smy, S. S. Winterton, S. K. Dew, and M. J. Brett, "Simulation and analysis of electromigration failure distributions," *Microelectron. Reliab.*, vol. 34, no. 6, pp. 1047-1056, 1994.

[9] M. J. Attardo, R. Rutledge, and R. C. Jack, "Statistical metallurgical model for electromigration failure in aluminum thin-film conductors," *J. Appl. Phys* , vol. 42, pp. 4343-4349, Oct. 1971.

[10] M. Shatzkes and J. R. Lloyd, "A model for conductor failure considering diffusion concurrently with electromigration resulting in a current exponent of 2," *J. Appl. Phys.*, vol. 50, no. 11, pp. 3890 - 3893, 1986.

[11] I. R. Miller, J. E. Freund, and R. Johnson, *Probability and Statistics for Engineers*, 2nd Edition. Englewood Cliffs, NJ: Prentice Hall, 1977.

[12] S. Yang, "Logic Synthesis and Optimization Benchmarks User Guide Version 3.0," Technical Report, Microelectronics Center of North Carolina, 1991.

[13] HP CMOS10QA Design Reference Manual, 1996.

[14] M. Xakellis and F. N. Najm, "Statistical estimation of the switching activity in digital circuits," *ACM/IEEE Design Automation Conference*, pp. 728-733, 1994.

[15] L. M. Ting, J. S. May, W. R. Hunter, and J. W. McPherson, "AC electromigration characterization and modeling of multilayered interconnects," *Int. Rel. Phys. Symp.*, pp. 311-316, 1993.

[16] B. Liew, N. W. Cheung, and C. Hu, "Projecting interconnect electromigration lifetime for arbitrary current waveforms," *IEEE Trans. on Electron Devices*, vol. ED-37, no. 5, 1990.