Loughborough
University

This item was submitted to Loughborough's Institutional Repository
(https://dspace.lboro.ac.uk/) by the author and is made available under the
following Creative Commons Licence conditions.

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/

Analysis of multistate models for electromigration failure

V. M. Dwyer

Additional information on J. Appl. Phys.

Analysis of multistate models for electromigration failure

V. M. Dwyer[a)]

Department of Electronic and Electrical Engineering, Loughborough University, Loughborough LE11 3TU, United Kingdom

The application of a multistate Markov chain is considered as a model of electromigration interconnect degradation and eventual failure. Such a model has already been used [Tan *et al.*, J. Appl. Phys. **102**, 103703 (2007)], maintaining that, in general, it leads to a failure distribution described by a gamma mixture, and that as a result, this type of distribution (rather than a lognormal) should be used as a prior in any Bayesian mode fitting and subsequent reliability budgeting. Although it appears that the model is able to produce reasonably realistic resistance curves R(t), we are unable to find any evidence that the failure distribution is a simple gamma mixture except under contrived conditions. The distributions generated are largely sums of exponentials (phase-type distributions), convolutions of gamma distributions with different scales, or roughly normal. We note also some inconsistencies in the derivation of the gamma mixture in the work cited above and conclude that, as it stands, the Markov chain model is probably unsuitable for electromigration modeling and a change from lognormal to gamma mixture distribution generally cannot be justified in this way. A hidden Markov model, which describes the interconnect behavior at time t rather than its resistance, in terms of generally observed physical processes such as void nucleating, slitlike growth (where the growth is slow and steady), transverse growth, current shunting (where the resistance jumps in value), etc., seems a more likely prospect, but treating failure in such a manner would still require significant justification. © *2010 American Institute of Physics*. [doi:10.1063/1.3262497]

I. INTRODUCTION

Electromigration (EM), the electric current-induced atom transport in very large scale integrated interconnect metallization, remains one of the most serious reliability concerns for modern integrated circuits. The standard failure model (e.g., Ref. 1) assumes that momentum transfer from the applied electron current j drives atoms toward the anode, increasing the tensile stress at the cathode. When this stress reaches a critical value, a void is nucleated. Further matter transport causes the void to grow, and when a critical void volume V_{cr} is reached, the line fails. It is clear however that the critical volume does depend on the location of the void, for example, in general, the value of V_{cr} required for a void growing beneath a via (via-above) is significantly smaller than that for a void above it (via-below); consequently the standard model must distinguish a variety of different EM failure modes.[2]

The overall atomic diffusion coefficient (D_A) plays a key role in any metal migration and its value at a point may be written in terms of a sum of local contributions as[3]

$$D_A = D_b + \frac{(\delta D)_{GB}}{d_{50}} + \frac{(\delta D)_{Cu:N}}{h} + (\delta D)_{Cu:b}\left(\frac{2}{w} + \frac{1}{h}\right), \quad (1)$$

where D_b is the diffusion coefficient in the bulk material, while $(\delta D)_{GB}$, $(\delta D)_{Cu:N}$, and $(\delta D)_{Cu:b}$ are the path-width $\times$ diffusion coefficient products, respectively, for grain boundary, copper/caplayer-interface and copper/barrier-interface diffusion. w and h are the linewidth and height, and d_{50} is the mean metal grain diameter. In copper interconnect, diffusion along the interface between the copper and the nitride caplayer generally dominates D_A (Refs. 4–6) so that voids tend to nucleate and grow along this interface. As the values of the diffusivity terms in Eq. (1) vary with local grain orientation, and hence microstructure, the failure times observed for a group of lines made using the same manufacturing process will be described according to some probability distribution. Lognormal statistics are almost universally assumed to apply for this failure mechanism as discussed in Ref. 7.

Empirical EM failure studies on the other hand, partially out of convenience, generally define the failure event to be a fixed (relative) increase in the line resistance R(t).[8] Typically, a figure of 10%–20% is chosen and this has been justified as being sufficient to cause enough interconnect delay in global wiring for timing errors to occur.[9] In some cases the resistance trace changes so rapidly at failure that the choice of 10% or 20% is not important. However, in other cases, $R(t)/R_0$ meanders somewhat and the times corresponding to 10% and 20% increases can be very different. Unfortunately, the link between these two pictures of failure (void volume and relative resistance change) remains somewhat unclear (see Doan *et al.*[10] for something of a review of this area for aluminum interconnect), and clearly there are some obvious differences. For example, all of the voids in a line contribute toward the increase in R(t), whereas only the largest determines critical volume failure. Generally one argues a link in the following manner.[11,12] If one ignores shunt layers, if the resistivity of the line is constant, and if the line can be treated

a)Tel.: +44(0)1509 227027. FAX: +44(0)1509 227014. Electronic mail: v.m.dwyer@lboro.ac.uk.

as a simple set of series and parallel resistors, then the overall fractional resistance change (FRC) for the line is roughly

$$\frac{\Delta R(t)}{R_0} = \frac{1}{L}\int_0^L \frac{A_{void}(x,t)}{A_{metal}(x,t)}dx$$

$$\approx \frac{1}{A_{line}L}\int_0^L A_{void}(x,t)dx = \frac{V_{void}(t)}{V_{line}}, \qquad (2)$$

where L and A_{line} are the length and cross-section of the line, $V_{void}(t)$ is the total volume of all voids along the line, $\Delta R(t)$ is the resistance change, R_0 is the initial line resistance, and $A_{void}(x,t)$ and $A_{metal}(x,t) = A_{line} - A_{void}(x,t)$ are the cross-sectional area given over to voids and metal at point $x \in [0, L]$ at time t. The approximation made in Eq. (2) is valid provided $A_{void}(x,t)/A_{line}$ is small. Thus for a single slitlike void (or for multiple such voids) with a small cross-section relative to the line, the fractional increase in line resistance will be a reasonable measure of the total void volume. As it is inconceivable that a slitlike void, nucleated at a random point along the line, will grow to a void volume of 0.1 V_{line}, such slitlike voids will not themselves lead to failure. For larger values of $A_{void}(x,t)/A_{line}$, the situation can be different. Suppose that a small prismatic void is situated between x and x+X along the line with an arbitrary cross-section $A_{void}(t)$, independent of x. The factional resistance change is

$$\frac{\Delta R(t)}{R_0} = \frac{1}{L}\int_x^{x+X} \frac{A_{void}(t)dx}{A_{line} - A_{void}(t)}$$

$$= \frac{V_{void}(t)}{V_{line}}\left(1 - \frac{A_{void}(t)}{A_{line}}\right)^{-1}, \qquad (3)$$

which includes a correction factor, relative to Eq. (2), dependent on the void shape. A short transverse void (small X, large A_{void}) will produce a larger $\Delta R(t)/R_0$ than a long slitlike void (large X, small A_{void}) of the same volume. Likewise, for the same volume growth rate, $\Delta R(t)/R_0$ will increase more rapidly for transverse than for slitlike growth. This case naturally includes the case of a slitlike void growing beneath a via, as this is transverse to the current flow.

Jump changes in resistance can occur if the current is diverted through shunt layers typically Ti/TiN (aluminum) or Ta/TaN (copper).[13,14] Including a shunt through a liner, Eq. (3) (for a void in the form of a prism) becomes

$$\frac{\Delta R(t)}{R_0} = \frac{V_{void}(t)}{V_{line}}\frac{1 - \dfrac{\rho_{Cu}}{\rho_{liner}}\dfrac{A_{liner}}{A}}{1 - \dfrac{A_{void}(t)}{A} + \dfrac{\rho_{Cu}}{\rho_{liner}}\dfrac{A_{liner}}{A}}$$

$$\underset{A_{void}\to A}{\longrightarrow} \frac{\rho_{liner}A}{\rho_{Cu}A_{liner}}\frac{X(t)}{L} = \frac{R_{liner}(t)}{R_0}. \qquad (4)$$

Since, in the case of a Ta, the electrical resistivity of the liner is around 20–70 times larger[11,15] than that of Cu, and the cross-section of the liner is around 10%–40% (Ref. 11) of the line cross-section, $\rho_{Cu}A_{liner}/\rho_{liner}A$ will be small (around 0.25/50 = 0.005 say), and will only contribute to Eq. (4) when $A_{void}(t) \sim A$, i.e., when the void extends across the line. At this point the line resistance rises suddenly to $\Delta R(t)/R_0$

$\approx R_{liner}(t)/R_0$ [equal to around 200X(t)/L] and will result in immediate EM failure, taken as $\Delta R(t)/R_0 = 0.1$, if $X(t) > L/2000$. For a 100 μm line, this corresponds to $X_{shunt}(t) > 50$ nm, while for 3000 μm this is $X_{shunt}(t) > 1.5$ μm; while smaller voids will cause shunting over a shorter distance and so cause smaller jumps in $\Delta R(t)/R_0$. In many ways this is the difference between the pictures, tying the definition of failure to a fixed *relative* increase in R(t) corresponds to a critical void volume V_{cr} that scales with line length, alternatively using a fixed critical void volume corresponds [to within the shape factor in Eq. (3)] to more of a fixed *absolute* increase in R(t).

This analysis is, of course, far too simplistic ignoring the differences for an interconnect tree between via-above failures and via-below failures; the differences between failures caused by cathode edge displacement and those due to cathode end thinning; the effects on $\Delta R(t)/R_0$ of multiple via contacts or multiple links; the feedback effects of current crowding around a void; and the possibility of resistance increase due to the formation of pancakelike voids in any solder joint. It also ignores void nucleation, which often plays a dominant role in solder and via-above failure and an important role in many of the other modes. In an M1/M2 test structure, if the void forms in the M1 layer below a via (via-above), a much smaller metal depletion is required to produce a slitlike void capable of creating an open circuit in the copper than if the void forms on the M2 topside (via-below). Clearly for a (via-below) void forming at the cathode, edge displacement and cathode thinning cause similar effects, reducing the line cross-section above the via in the M2 layer. Depending on geometry, microstructure, and whether or not there is a reservoir overlap, edge displacement may shunt the current through the liner sooner or later than by cathode end thinning. In multivia contacts, several jumps have been observed as the vias gradually become open circuit one by one and, progressively, more and more of the current is forced through the remaining contacts. Similarly, during the growth of the void, current crowding occurs as the current is forced through a smaller area of metal. This is likely to accelerate the failure process by increasing the EM force. This same process is at work at solder bumps where the geometry of the contact between the under bump metallization and the solder ball causes crowding, increased current densities and accelerated failure.[16–18]

A complete discussion of possible void behavior, and its influence on the measured resistance, is beyond the scope of this work; however some of the key features of the process of the nucleation and the growth may be demonstrated using a simple toy model similar to that used for classical condensation.[19] If the surface energy of a void (taken to be a constant here) is γ_S, then forming a small spherical void of radius r increases the free energy by $4\pi r^2\gamma_S$. The atoms removed to form that void must be relocated and, if σ is the local hydrostatic tensile stress, an amount of work (equal to the integral of the elastic strain energy density) of $-4\pi r^3\sigma/3$ is done by remote stresses. The resultant net increase in the free energy is $F = 4\pi r^2(\gamma_S - r\sigma/3)$. Consequently growing the void by increasing its radius from r to $(r+\delta r)$ alters the free energy by $\delta F = 4\pi r(2\gamma_S - r\sigma)\delta r$, which will be negative if r

$> r_{cr} = 2\gamma_S / \sigma$. As a result, it is energetically favorable for any void of radius greater than r_{cr} to grow provided that it is supplied with sufficient vacancies through the EM drift-diffusion mechanism. On the other hand all voids smaller than r_{cr} are unstable and would be expected eventually to evaporate. Growth of embryos smaller than the critical radius is possible provided that the random processes by which vacancies are captured and emitted provide sufficient net capture to overcome the tendency toward evaporation. The situation described is of course unrealistic in a number of ways, not least of which are the assumptions that the metal is in a hydrostatic state throughout the process and that the incorporation of the void into the bulk metal will not change the stress σ. Nevertheless, many of the conclusions are essentially valid, for example, it is intuitively correct that the critical radius for void growth should vary inversely with the hydrostatic stress $r_{cr} \sim \sigma^{-1}$. This implies that during the annealing phases in the manufacturing process, the large thermal stresses which arise from elastic and thermal expansion mismatches between the metal and the passivation will reduce r_{cr}, possibly to below the size of manufacturing flaws along the interface, allowing them to grow. Such flaws are likely to be randomly distributed along these interfaces.

In recent years, a number of authors have performed much more detailed calculations of void growth which combine the competing factors of the EM force, the capillary force (which drives the surface diffusion), and the forces which arise from gradients in elastic strain energy density, to model the atomic velocity along the copper topside and at the void surface.[20–31] Each of these factors depends on the local microstructure, leading to anisotropies which may affect the void growth direction and consequently its shape factor in Eq. (3).[27–29] These calculations are able to demonstrate complex morphological changes similar to those of real voids in encapsulated metal lines,[24,25,27–29] including a void splitting mechanism in which, in the presence of a strong anisotropy, a cylindrical void can develop side features which may then detach from the mother void,[27–29] and a description of the migration of voids toward the cathode and a model of void pinning at grain boundaries.[22,31] To understand failure statistics, it is necessary to limit such problems, such as by starting with a single mode, such as via-below failure in M2 lines, and considering such growth complexities at a later stage.

Nucleation of voids, on the other hand, may occur under the high thermal stresses experienced as a result of the manufacturing process or as a result of the random processes of vacancy capture and emission driven by the net EM process. The latter situation can be described by a probability density function (pdf) $f(r,t)$, which determines the number of embryos with radius r at time t.[16] As with many such random processes, this will evolve roughly as the solution to some form of Fokker–Planck equation, in which the fluctuation term describes the random vacancy capture and emission processes.[16,32–34] Some of these embryos may gather enough vacancies to reach r_{cr} and grow subsequently. During the EM process, the tensile stress toward the cathode end increases compared to other parts of the line, reducing the critical radius for void growth close to that end; consequently, voids are expected to nucleate preferentially there. The use of the Fokker–Planck equation effectively models each (subcritical) embryo with a Markov chain (MC), whose state label is the number of vacancies in the embryonic volume.[32–34] Vacancy emission and capture rates are incorporated as transition rates between the states. This MC model can be expected to work quite well if inhomogeneities in the line may be ignored, and if the embryos are small. It should not be expected to work if the embryos are large enough to experience the microstructure of the line, as is the case a growing void.

However they are defined, lifetimes for EM-induced failures of aluminum interconnects generally show very reasonable fits to lognormal statistics,[35–38] although the test structures used have a considerable effect on the details. As a general rule, bamboo and near bamboo Al structures with shunt layers tend to show growth dominated failure while, without the shunt layer, the nucleation time dominates the failure time.[1] In aluminum bamboo structures, voids tend to nucleate close to the cathode and via failure is the dominant cause of line failure. For a number of reasons things are not so simple in the case of copper. The consensus on EM failure in the current generation of copper interconnect is that the interface between the copper metallization and the nitride caplayer remains the fast diffusion path as well as the likely nucleation site,[4–6] but new treatments/liner materials are reducing the significance of this path. Voids, usually forming near the via, can grow to failure or sometimes can de-pin and drift toward the cathode where they can combine with an existing void.[39] This more dynamic picture of void behavior has generated renewed interest in developing a theoretical basis for the distribution of failure times in damascene copper. In the case of short copper pad-stud lines, we have suggested[7] recently that void growth at the via is dominated by a steady-state vacancy current. This steady current of vacancies into a void causes its volume to grow at a steady rate, so that a lognormal distribution arises from a lognormal distribution of steady-state currents, which in turn is determined by the line microstructure due to its effect on local interface diffusivity values. Seen from the point of view of the resistance $R(t)$, this vacancy current may affect the line in a seemingly more stochastic manner depending on the manner in which each vacancy is absorbed into the void. The purpose of the present paper is to consider MC methods to describe this more stochastic variation in resistance values.

A recent article by Tan *et al.*[40] uses such a model for the degradation of the line resistance and finds the failure time distribution to be a gamma mixture, with each element in the mixture corresponding to a different failure mode. It is claimed that the gamma distribution is more general than the lognormal, which they maintain is a special case. This latter statement is somewhat imprecise. It has some validity in the sense that both distributions, along with the Weibull and inverse gamma (Pearson V), are special cases of the generalized gamma distribution, which contains the Weibull factor $\exp(-\alpha t^\beta)$.[41,42] However, the generalized gamma does not arise from the Markov model of Ref. 40. As the standard gamma and lognormal are both two-parameter distributions, it cannot be true that the lognormal is a special case of the gamma distribution unless the two are identical (strictly, this

assumes that any mapping of parameters $(\mu, \sigma) \mapsto (n, \theta)$ from lognormal $LN(\mu, \sigma)$ to gamma $\Gamma(n, \theta)$ has an inverse, but it would be hard to imagine anything else); neither can the lognormal be any less general as a result. In Ref. 40, the lognormal distribution is discounted as it does not describe the Kolmogorov wear failure mechanism when some of the wearing is not gradual; however, the gamma distribution would have to be discounted by exactly the same argument. As indicated in Ref. 7, even gradual wear does not lead to a lognormal distribution of the failure time, but rather of (some notion of) the damage. Perhaps an appropriate conclusion is that Kolmogorov wear is not a suitable model for describing EM failure times.

The parameters in the Gamma mixture in Ref. 40 are obtained through Bayesian inference techniques and the maximum likelihood principle. This is performed using the Akaike information content (AIC) measure[43] to assess the number of adjustable parameters and the expectation-maximization algorithm (EMA) to determine their value.[44] To be both powerful and robust, statistical inference techniques such as these require accurate priors (the probabilistic models of the processes generating the data). Virtually all the statistical analysis of EM data assumes these models to be described by a lognormal distribution for each failure mode rather than a gamma distribution. A change from lognormal (or a lognormal mixture) to a gamma mixture is significant because it alters the reliability estimates obtained following deceleration and extrapolation of EM life-test results back to use conditions. Such a change is not out of the question though as, being fairly close siblings, gamma distributions can resemble lognormal distributions and the gamma cumulative distribution function (CDF) can appear reasonably straight when plotted on lognormal probability paper. As a result both have been used to describe grain size distributions, e.g., Ref. 45.

The purpose of the current work is threefold. First we consider the closeness of gamma distributions to lognormal data in a manner which allows the large quantity of lognormal plots to be represented as gamma distributions in a simple way. Second we review the article of Tan *et al.*,[40] as there are a number of inconsistencies in their theoretical development [for example, the gamma mixture result corresponding to Eqs. (15) and (16) in Ref. 40 does not follow from their Eqs. (9)–(12)], which relate to the difference between a random variable X being represented by a weighted sum of other random variables X_i (weights arbitrary) and a mixture distribution in which the pdf $f_X(x)$ is a weighted sum of pdfs $f_{Xi}(x)$ (weights > 0 which sum to unity). Mainly though, we seek to develop a critical analysis of the general application of Markov models to the EM problem. This application seems fairly reasonable at first, particularly in view of the fact that the nucleation process has been described by a Fokker–Planck equation,[32–34] which treats the embryos (subcritical voids) as a state-based model, with state index (n) equal to the number of vacancies in the embryo. However, it does appear that given the value of the resistance at time t [R(t)=r say], the probability distribution of resistance values at time t+dt is likely to depend as much on the type of void growing (slitlike or transverse, edge displacement or

thinning, via-above, via-below, etc.), and whether A_{void} is sufficiently close to the line cross-section that some of the current will be shunted through the liner, as it does on the value of r. In other words there does appear to be sufficient memory of the previous history of the line resistance to be concerned that the model of the interconnect as a system of states labeled only by R(t) may not possess the Markov property. Perhaps a hidden Markov model would work better, in which the state-space describes possible types of void growth behavior within the interconnect (slitlike, transverse, jumps associated with shunting, etc.) and switches between them according to some stochastic process.

II. GAMMA VERSUS LOGNORMAL

The lognormal distribution or a lognormal mixture has been used to describe EM times for around 40 yr. If t_f is distributed lognormally then $\log(t_f)$ is normally distributed and may be described by parameters μ, the mean of $\log(t_f)$, and σ_{TTF} its standard deviation. Almost all EM failure data is plotted on lognormal paper and consequently tends to come quoted with it a median value t_{50} equal to $\exp(\mu)$ and a value of σ_{TTF}. The pdf for a lognormally distributed t_f is given by $[LN(\mu, \sigma_{TTF})]$

$$f_{LN}(t_f) = \frac{1}{\sigma_{TTF} t_f \sqrt{2\pi}} \exp\left\{ -\frac{[\log(t_f) - \mu]^2}{2\sigma_{TTF}^2} \right\}. \quad (5)$$

By contrast the gamma distribution $\Gamma(n, \theta)$ is defined as

$$f_{GAM}(t_f) = \frac{1}{\theta \Gamma(n)} \left(\frac{t_f}{\theta} \right)^{n-1} \exp\left(-\frac{t_f}{\theta} \right), \quad (6)$$

where n is known as the shape parameter (as it determines the profile of the distribution) and θ is the scale parameter (as it essentially scales the t_f axis).[43,44] $\Gamma(n)$ is the gamma function [such that, for integer n, $\Gamma(n)=(n-1)!$]. For a Gamma distributed t_f, the mean $\bar{t}_f$ is $n\theta$ and the variance is $n\theta^2 = (\bar{t}_f)^2/n$, whereas for lognormally distributed t_f, the mean value $\bar{t}_f$ is $\exp(\mu + \sigma_{TTF}^2/2)$ and the variance is $(\exp(\sigma_{TTF}^2) - 1)(\bar{t}_f)^2$. Consequently, a gamma distribution and a lognormal distribution with the same mean will also have the same standard deviation if $\exp(\sigma_{TTF}^2) - 1 = n^{-1}$ or $\sigma_{TTF}^2 n \sim 1$ if σ_{TTF} is small.

The gamma distribution has been used as an alternative to the lognormal on a number of occasions; particularly, as far as EM is concerned, in relation to the distribution of grain (and generally particle) sizes.[45–47] One means of comparing the two distributions would be to plot the CDF of the gamma distribution $F_{GAM}(t)$ on a lognormal plot [i.e., a plot of $erf^{-1}[2F_{GAM}(t) - 1]$ against $\log(t)$] and look for the best straight line. Perhaps a less biased comparison, performed by Fátima Vas and Fortes,[48] uses the fact that both distributions have two parameters; for the gamma these are the mean $\bar{t}_f$ and the shape parameter n, while for the lognormal these are the mean $\bar{t}_f$ (as the gamma median has no closed form) and the lognormal deviation σ_{TTF}. Assuming that the means $\bar{t}_f$ of the two distributions are equal it is possible to obtain a relationship between σ_{TTF} and n, which minimizes the least-squares best fit of the two pdfs, defined by the minimum of

FIG. 2. A schematic of a MC model, with N states allowing nearest neighbor (m=1) and next nearest neighbor (m=2) jumps.
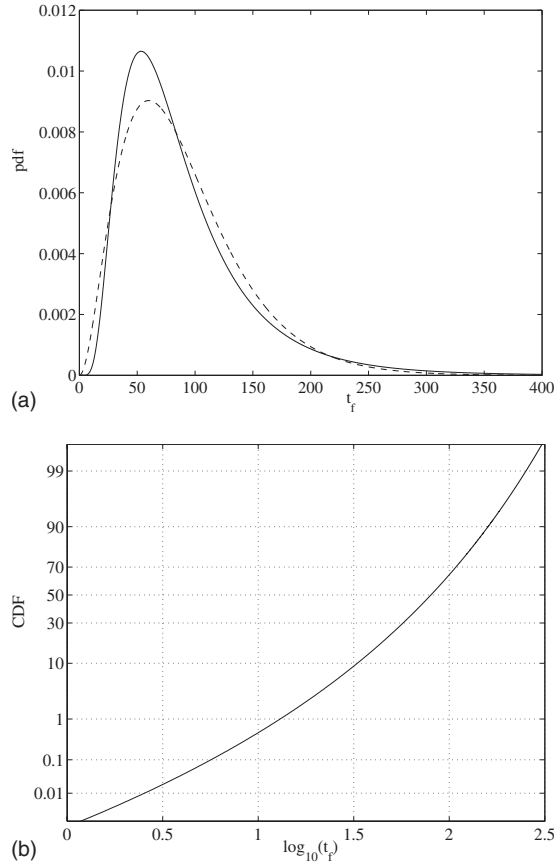
FIG. 1. (a) A comparison of a lognormal plot with $\sigma=0.59$ with a gamma distribution $\Gamma(3,30)$, where both means are equal to 90. (b) shows the CDF of $\Gamma(3,30)$ on lognormal paper.

$$\Sigma = \overline{t_f} \int_0^\infty [f_{LN}(t) - f_{GAM}(t)]^2 dt. \qquad (7)$$

For all vales of σ_{TTF}, it is found[48] that the best fit (the best choice of n for a given value of σ_{TTF}) corresponds to

$$\sigma_{TTF}^2 n \sim 1, \qquad (8)$$

as might also be expected from the discussion above. Thus if data on a lognormal plot are found to yield a lognormal deviation of $\sigma_{TTF} \sim 0.6$ (not untypical of aluminum failure times[49]) its best fit with a gamma distribution will correspond to $n \sim 3$, while data yielding $\sigma_{TTF} \sim 0.2$ (not untypical of a large number of copper samples[50–53]) is best fitted by a gamma with $n \sim 25$. Values of Σ_{min} are less than 0.02 for n > 2 [on a scale where the integrals of f(t) are unity as usual]. The product $\sigma_{TTF}^2 n = 1.09$ for n=2 improving toward 1 from above as n increases. Figure 1(a) compares the pdfs of LN(4.15, 0.59) (solid) and $\Gamma(n=3, \theta=30)$ (dashed), both of which have a mean of 90. It is clear in Fig. 1(b) that the CDF of $\Gamma(n=3, \theta=30)$ plotted on lognormal paper shows up as a reasonably good straight line.

One observation in this regard is that, if $x \sim \Gamma(n, \theta)$, then its generating function E[exp(sx)] is given by g(s)=$(1 -s/\theta)^{-n}$. As a result the generating function for a fixed general gamma mixture is a *sum* of such terms. By a careful choice of the number of terms in the general mixture and the associated weights (some of which may need to be negative)
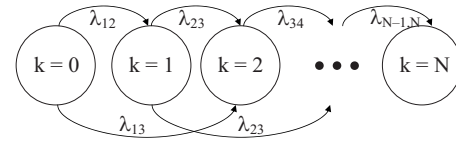
and gamma parameters, the resulting Padé-type structure can become arbitrarily close to any distribution defined on t>0 (gamma mixtures are weakly dense on t>0) (Refs. 54–56). This does not necessarily make it the best prior for the EMA algorithm. One would also have to be careful when asserting that each of these gamma elements corresponds to a different failure mode.[40]

It is likely that, in many samples, the data will be insufficiently good to resolve the gamma versus lognormal question either way. A particular difficulty arises in the case of copper, for which it is commonly found that $\sigma_{TTF} \sim 0.1-0.2$, (e.g., 50–53) when the skewness of the lognormal distribution is quite small and it closely resembles a normal distribution[57] except in the all-important tail region needed for extrapolation. Likewise, in the limit of large n (corresponding to small σ_{TTF}), the gamma distribution $\Gamma(\theta,n)$ also resembles a normal distribution.[58] Consequently all three take on the appearance of a rough straight line on a lognormal plot. As a result, a careful consideration of the physical processes is required in an attempt to settle the issue. A candidate description of the failure process, suggested by Tan *et al.*,[40] treats the line as a multistate system whose dynamics is modeled by a MC.

III. MC MODEL

We consider the MC model as suggested in Ref. 40, an example of which is shown in Fig. 2. The system is specified in terms of a set of states k with resistance R_k. The state k =0 corresponds to a resistance of R_0 and represents, for example, the line just at the point of void nucleation, or perhaps the initial line resistance. The final state k=N corresponds to a failed state $R_N=R_{cr}$, typically chosen as a 10% increase, $R_N=1\cdot1R_0$. In this system, transitions are only allowed in the forward direction (the resistance is assumed to be nondecreasing), with small resistance increases between the neighboring states allowed as well as larger jump transitions. Denote by λ_{ij} the transition rate between state k=i and state k =j$>$i and let $p_k(t)$ be the probability that, at time t, the line resistance is R_k. Although the system has O(N^2/2) parameters there is a large amount of redundancy and only 2N–1 parameters are needed to specify the CDF. Markov methods though are very general, capable of producing an approximation arbitrarily close to any failure time distribution.[54]

Contrary to Tan *et al.*,[40] we find that gamma distributions and true gamma mixtures (without any negative weights in the mixture) are very rare. When they do arise the choice of the R_k values, and thus of N, is crucial, as N is often simply related to the gamma shape parameter n. The analysis of the previous section shows that a significant number of experiments require the value of n for copper (σ_{TTF}

$\sim 0.2, n \sim 25$) to be significantly larger than that for aluminum ($\sigma_{TTF} \sim 0.6, n \sim 3$), and thus $N_{Cu} \gg N_{Al}$. Why there might be many more states for copper than for aluminum is an important question for MMs.

MCs are more familiar in, for example, queuing theory, where the statistical distributions of queue length arise from random processes occurring to a single queue of known structure. By contrast here, in a certain sense, we are looking at a single known process (EM) occurring to a random distribution of (unknown) line microstructures, i.e., the variability is as much between the lines as in the arrivals, so one expects different statistical priors for each line. Consequently, it is clear that there is no simple mapping of the EM condition onto the *states* R_k. These issues, of course, matter less if, as claimed in Ref. 40, the MC always produces the same type of failure distribution.

Naturally the chain parameters λ_{ij} depend crucially on the choice of states R_k. The nature of these states is not discussed in Ref. 40 where a gamma mixture distribution is claimed, whatever the choice. Without any detailed knowledge of the line microstructure, only some fairly simple assumptions may be made, and for lack of any guiding principle on this, we choose the states R_k to be evenly spaced between R_0 and $R_N = 1 \cdot 1 R_0$ (the failure state), thus $R_k / R_0 = 1 + k/10N$. For now we leave N as a parameter taking typical values to be: N small; $N = 10$; $N = 100$; and the large N limit; as these cover the aluminum and copper values $N \sim n \sim 3$ and $N \sim n \sim 25$ reasonably well.

It may be expected that gradual resistance increases shall always be allowed, so that we assume $\lambda_{k,k+1} > 0$, and in general we shall also allow some larger jumps from state k to state $k+m$ at lower rates $\lambda_{k,k+m}$. These jumps may be due to a dynamic change in growth direction, which may cause current crowding and an increased resistance. The resulting change in shape may also increase the capture efficiency of the void and hence the appropriate transition rates may increase when the void volume or R(t) reaches a certain size. Alternatively the jumps may be due to current shunting through the liner or to additional microvoids arriving at the cathode, generating a step increase in the void volume and hence in R(t). Figure 2 shows a chain of length N, with gradual changes between neighboring states (a jump of 1), together with an allowed jump of 2 states (m=2).

From a generalized version of Fig. 2, for which (in principle) any forward transition is allowed, the balance equation for state k=0 is

$$p_0(t + dt) = p_0(t) - \lambda_{01} dt p_0(t) - \lambda_{02} dt p_0(t) - \cdots$$
$$- \lambda_{0N} dt p_0(t), \tag{9}$$

so that, as $dt \to 0$, we obtain the rate equation

$$\frac{dp_0(t)}{dt} = -\left(\sum_{k=1}^{N} \lambda_{0k} \right) p_0(t). \tag{10}$$

Likewise for the other states,

$$\frac{dp_j(t)}{dt} = -\left(\sum_{k=j+1}^{N} \lambda_{jk} \right) p_j(t) + \sum_{k=0}^{j-1} \lambda_{kj} p_k(t), \tag{11}$$

For $1 \leq j \leq N-1$; and

$$\frac{dp_N(t)}{dt} = \sum_{k=0}^{N-1} \lambda_{kN} p_k(t) \tag{12}$$

for state j=N. Conservation of probability is assured, $\sum_{k=0}^{N} p_k(t) = 1$, through the addition of Eqs. (10)–(12). For the reduced set $0 \leq j \leq N-1$, Eqs. (10) and (11) may be written as the matrix equation $(d/dt)\underline{p} = Q\underline{p}$ or

$$\frac{d}{dt}\begin{bmatrix} p_0(t) \\ p_1(t) \\ p_2(t) \\ \vdots \\ p_{N-2}(t) \\ p_{N-1}(t) \end{bmatrix} = \begin{pmatrix} -\sum_{k=1}^{N} \lambda_{0k} & 0 & 0 & \cdots & 0 & 0 \\ \lambda_{01} & -\sum_{k=2}^{N} \lambda_{1k} & 0 & \cdots & 0 & 0 \\ \lambda_{02} & \lambda_{12} & -\sum_{k=3}^{N} \lambda_{2k} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 & 0 \\ \lambda_{0N-2} & \lambda_{1N-2} & \lambda_{2,N-2} & \cdots & -\sum_{k=N-1}^{N} \lambda_{N-2k} & 0 \\ \lambda_{0N-1} & \lambda_{1N-1} & \lambda_{2,N-1} & \cdots & \lambda_{N-2,N-1} & -\lambda_{N-1,N} \end{pmatrix} \begin{bmatrix} p_0(t) \\ p_1(t) \\ p_2(t) \\ \vdots \\ p_{N-2}(t) \\ p_{N-1}(t) \end{bmatrix}. \tag{13}$$

Hence $\underline{p}(t) = \exp(Qt)\underline{p}(t=0)$. Let $-w^T$ denote the vector of column sums of the $(N \times N)$ matrix Q, i.e., $w^T = \{\lambda_{0N}, \lambda_{1N}, \ldots, \lambda_{N-1,N}\} = -\mathbf{1}^T Q$, where $\mathbf{1}^T = [1, 1, 1, \ldots, 1]$. The diagonal elements relate to the transitions out of a given state, while the off-diagonal elements relate to the transitions into other states. Q is lower triangular due to the occurrence of transitions only in the direction of increasing R(t).
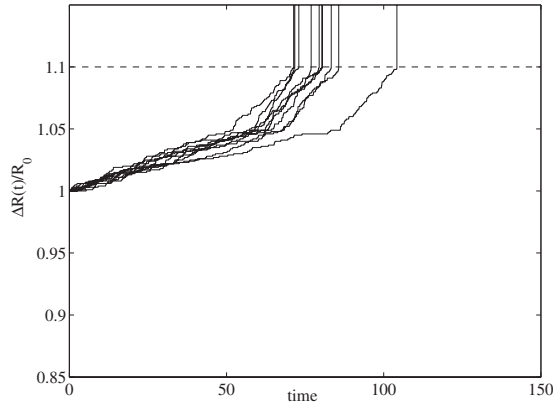
We consider the following problems raised by Ref. 40:

FIG. 3. Fractional resistance increase $\Delta R(t)/R_0$ curves for ten simulated interconnects described by a chain of length N=100, which switched from a nearest neighbor transition rate of λ_1 (with $\lambda_1 dt = 2.5 \times 10^{-4}$) at k=50 to $\lambda_2 = 4\lambda_1$. In the simulation we take the time step between random events to be dt=1/3000.

FIG. 4. Fractional resistance increase $\Delta R(t)/R_0$ curves for three simulated interconnects described by a chain of length N=100, which switched from a nearest neighbor transition rate of $\lambda dt = 2.5 \times 10^{-4}$ with simultaneously the possibility of a jump of 25 states at a rate λ/N.

Problem (i). Transitions only occur between a state k and its immediate neighbor k+1. The transition rate is λ_1 up to some state k=K, and then changes to λ_2 (where λ_2 is significantly different from λ_1). This aims to model a system with an initial gradual degradation, which changes to a more catastrophic rate when the resistance reaches R_K, as suggested by Fig. 1 of Ref. 10. We show that the failure distribution in this case is the convolution of two Erlang distributions (which are integer gamma distributions) with different scale parameters. The result may be extended to any number of rate changes by convolving more Erlang distributions. The final distribution is complicated by the difference in scales of the two Erlang distributions. The convolution of M Erlang variates may be expressed exactly as a sum of gamma variates, although the sum involves O(N) terms and is a generalized mixture in the sense that some of the weights are negative.[54–56] In this case for N=100, with a transition between neighboring states of λ_1 for k<50 increasing to $4\lambda_1$ at that point, typical resistance plots are shown in Fig. 3.

Problem (ii). The overall transition rate $\Lambda_k = \Sigma_{j=k}^{N}\lambda_{kj}$ out of a given state k is different from the others and thus the diagonal elements of Q are all different. In this case it is easy to show that the failure distribution may be written as a sum of exponential functions (a phase-type distribution). Note that the failure distribution is a sum of exponential functions rather than, as assumed in Ref. 40, the failure time being a sum of exponentially distributed variates. This latter case does occur under some conditions, but is not a general result as, with jumps across states, the system need not necessarily proceed through each state in turn.

Problem (iii). The overall probability of a transition out of a state, in a given time, is the same for all states, so that $\Lambda = \Lambda_k = \Sigma_{j=k}^{N}\lambda_{kj}$ is the same for all k. This constraint aside, the λ_{jk}, are arbitrary. Within this problem we consider the special case of a gradual transition between states k and k+1 occurring simultaneously with a small probability jump from state k to state k+m. This jump is assumed to be the failed state, if the jump would take the system resistance beyond $R_N = 1 \cdot 1R_0$. This class of problem has a fairly simple solution in all cases. In the example case, if the jump through
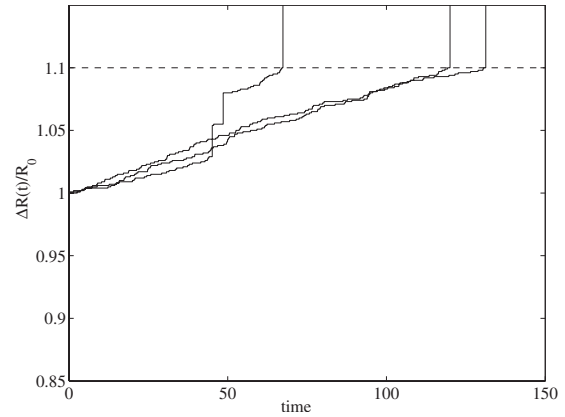
m states occurs with a low probability $\alpha(\ll\lambda)$ and there are N states in the chain, then the lifetime is distributed as a mixture of gamma terms of the form $\Gamma[N-km,(\lambda+\alpha)^{-1}]$, $k=0,1,2,\ldots,[N/m]$. However this is somewhat contrived, once again bringing in to sharp focus the need for some justification of the choice of states, the value of N and the associated transition rates λ_{ij}. With N=100, together with a jump transition of 25 states (allowed from any state) at a rate $\alpha=\lambda/N$, typical resistance plots are shown in Fig. 4.

We have chosen these problems as problem (ii) relates to the analysis in Ref. 40, while problem (i) relates more to the discussion there. Problem (iii) was chosen as it is the only case for which a gamma mixture in the sense of Ref. 40 could be obtained.

IV. FAILURE DISTRIBUTIONS

Whatever the problem, it is necessary to simplify the calculation of exp(Qt) as its evaluation may be required at many t values; Q may be large; or many trial systems Q may be investigated as possible fits to experimental data. We note that, as the matrix Q is lower triangular, its eigenvalues are simply the values along the main diagonal. If these values are different, as in problem (ii), (which occurs if the total transition rate out of each state is different) then Q can be diagonalized according to $D = S^{-1}QS$, where S is the matrix of eigenvectors of Q and D is the matrix of diagonal elements of Q. The failure distribution is then (Appendix)

$$f(t) = \sum_{k=0}^{N-1} (\underline{w}S)_k (S^{-1}\underline{e}_0)_{k0}\exp(-\Lambda_k t), \qquad (14)$$

which is a phase-type distribution.[54–56] The EMA algorithm, used by Tan et al.,[40] is one of the best algorithms to approximate data sets by continuous distributions and in a particular case the phase-type may be best approximated by a gamma distribution or a gamma mixture. Of course, it may be best approximated by a lognormal. A note of caution needs to be injected though as it is known that the EMA does not always capture the details of the distribution tail correctly,[60] which is vital for correct EM extrapolation.

Essentially this analysis is identical to that of Tan *et al.*,[40] as the failure time distribution is now a simple sum of exponential functions. It is not, however, a sum of variables each of which is exponentially distributed and thus convolution, as in Ref. 40, is inappropriate. The result in Eq. (16) of Ref. 40 that

$$f_S(t) = \sum_{k=0}^{N-1} \left[\left(\prod_{j=1}^{N-1} A_j^{(k)} \right) \frac{\Lambda_k^j t^{j-1} \exp(-\Lambda_k t)}{(j-1)!} \right], \quad (15)$$

where $A_j^{(k)}$ is the weight of the exponential term $\exp(-\Lambda_k t)$ in state $p_j(t)$, also suffers from the problem that if $A_j^{(k)}$ is zero for any j, then $\exp(-\Lambda_k t)$ will be missing from the sum. Note from Eq. (7) that $p_0(t)=\exp(-\sum_{j=1}^N \lambda_{0j} t)=\exp(-\Lambda_0 t)$, so that $A_0^{(i)}=0$ for all $i \neq 0$, which would prevent all Λ_k (k $\neq 0$) from appearing in Eq. (15). Clearly this cannot be correct.

A. Problem (i)

Difficulties arise with the development above whenever two or more (ν say) of the diagonal elements of Q are equal (to λ say) as the eigenvalues are degenerate, the matrix is defective and it is then not possible to diagonalize Q. The result is that the $p_k(t)$ are not simple sums of exponentials (for $k < N$) and instead contain the higher order terms $t^\nu \exp(-\lambda t)/\nu!$. Such a situation arises if we allow gradual erosion only ($\lambda_{k,k+1}=\lambda$ and all other $\lambda_{jk}=0$). Then

$$\frac{d}{dt} \begin{bmatrix} p_0(t) \\ p_1(t) \\ p_2(t) \\ \vdots \\ p_{N-2}(t) \\ p_{N-1}(t) \end{bmatrix} = \begin{pmatrix} -\lambda & 0 & 0 & \cdots & 0 & 0 \\ \lambda & -\lambda & 0 & \cdots & 0 & 0 \\ 0 & \lambda & -\lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & -\lambda & 0 \\ 0 & 0 & 0 & \cdots & \lambda & -\lambda \end{pmatrix}$$

$$\times \begin{bmatrix} p_0(t) \\ p_1(t) \\ p_2(t) \\ \vdots \\ p_{N-2}(t) \\ p_{N-1}(t) \end{bmatrix}. \quad (16)$$

Eq. (16) may be simply integrated iteratively as, from Eq. (10), $p_0(t)=\exp(-\lambda t)$ and setting $f_k(t)=\exp(\lambda t)p_k(t)$, so $f_0(t)=1$, we have from Eq. (11)

$$f_k(t) = \int_0^t \lambda f_{k-1}(u) du. \quad (17)$$

Consequently $f_k(t)=(\lambda t)^k/k!$ and $p_k(t)=(\lambda t)^k \exp(-\lambda t)/k!$. The reliability is then

$$r(t) = \exp(-\lambda t) \sum_{k=0}^{N-1} \frac{(\lambda t)^k}{k!}; \quad (18)$$

the failure distribution is then gamma (or more precisely Erlang)

$$f_S(t) = \frac{\lambda^N t^{N-1}}{(N-1)!} \exp(-\lambda t) \sim \Gamma(N, \lambda^{-1}) \quad (19)$$

with MTTF$=N/\lambda$. One immediate difficulty from Eq. (19) is that for a large number of copper interconnects, σ is small (~ 0.2), corresponding to $N \sim 25$, while for aluminum interconnect $\sigma \sim 0.6$ say, corresponding to $N \sim 3$. Why there should be 25 resistance stages for copper and three for aluminum presents a difficulty for the MC method in the present form.

Note that although the gamma distribution arises only in this specific case, in which the transition rates are equal and does not arise in the general case of arbitrary rates. Its origin is in a convolution, as in this simple case the system must pass through each state in turn, consequently we may write t_f as the sum of the times from state k=0 to k=1, plus that from k=1 to k=2, etc., i.e., $t_f = t_{0,1}+t_{1,2}+\ldots+t_{N-1,N}$. Each of the times is exponential and (importantly) iid, as the transition rates are equal, so that the distribution of t_f is gamma as above. If the rates are all different then the convolution only produces a sum of exponentials in place of Eq. (18). If V different rates Λ_k occur $(k=1,\ldots,V)$, each with multiplicity ν_k, so that $\sum_{k=1}^V \nu_k=N$, then the failure time may be reordered in sets. The sum of the ν_k times corresponding to rate Λ_k will give rise to $\Gamma(\nu_k, \Lambda_k^{-1})$ and t_f will be a V-fold convolution of such gamma distributions, all with different scale factors. A particular case (V=2) of this is problem (i) when a period of significant degradation follows an initial period of gradual degradation, as in Ref. 40. The failure time distribution is then the convolution $\Gamma(K, \Lambda_1^{-1}) \otimes \Gamma(N-K, \Lambda_2^{-1})$. As the scale parameters Λ_k are assumed to be significantly different, this is not itself a gamma distribution, nor indeed a true mixture. A more general problem which can give rise to some gamma mixture solutions follows.

B. A gamma mixture

Now consider the case of a more general transition matrix. As in the previous section we assume that the total transition rate out of each state is the same. However, the states into which the transitions are made are arbitrary subject to this constraint, i.e., we assume that the sum $\Lambda = \Lambda_k = \sum_{j=k}^N \lambda_{kj}$ is the same for all k, although subject to satisfying this sum rule, the λ_{jk} are arbitrary. It may well be possible to choose the states R_k so that this is true. In this case we may write

$$Q = -\Lambda I + L, \quad (20)$$

where L is strictly lower triangular and consequently nilpotent with $L^N=0$. Essentially this latter condition is true as L represents forward transitions so that, since the chain is unidirectional, after N forward transitions all systems will end up at k=N irrespective of their start configuration, and consequently will have left the set $\{k|0 \leq k \leq N-1\}$ covered by the matrix Q. As the matrices on the right hand side of Eq. (20) commute we may write

$$\exp(Qt) = \exp(-\Lambda t)\exp(Lt) = \exp(-\Lambda t)\sum_{k=0}^{N-1}\frac{L^k t^k}{k!} \qquad (21)$$

using the nilpotency of L. The reliability function r(t) is then simply

$$r(t) = \exp(-\Lambda t)\sum_{k=0}^{N-1}\frac{t^k}{k!}\underline{1}_N^{\ T}L^k\underline{e}_0 \qquad (22)$$

of which Eq. (18) is just a simple case. The failure distribution is

$$f(t) = -\frac{dr(t)}{dt} = \frac{\underline{1}_N^{\ T}L^{N-1}\underline{e}_0}{\Lambda^{N-1}}\Gamma(N,\Lambda^{-1}) + \sum_{k=0}^{N-2}\frac{\underline{w}L^k\underline{e}_0}{\Lambda^{k+1}}\Gamma(k+1,\Lambda^{-1}), \qquad (23)$$

which clearly is a general, rather than a true, mixture as some mixture coefficients may be negative.

As a simple example we consider the case that transitions between states k and k+1 take place at a rate λ and jump transitions to state k+m take place at a rate α. All states are assumed equivalent in this regard. The lower triangular matrix Q then has $-\Lambda = -(\lambda + \alpha)$ along the main diagonal, λ along the first subdiagonal and α along the mth subdiagonal. Equation (12) ensures that a jump past the failed state k=N is simply counted as a jump into k=N. In this case the values of the coefficients $c_k = \underline{w}L^k\underline{e}_0/\Lambda^{k+1}$ follow a fairly simple pattern. For $N-jm \le k \le N-(j-1)m$, $c_k \sim \alpha^j$ except at changeover points $k = N-j(m-1)$, where c_k has local peaks. The failure distribution keeping only those peaks is roughly

$$f(t) \approx \frac{\lambda^{N-1}}{\Lambda^{N-1}}\Gamma(N,\Lambda^{-1}) + \sum_{k=1}^{[N/m]}p^k(1-p)^{N-km}\begin{bmatrix}N-k(m-1)\\k\end{bmatrix}\Gamma[N-k(m-1),\Lambda^{-1}], \qquad (24)$$

(where $p = \alpha/\Lambda$) as could be expected. This is a true gamma mixture.

Here we have taken k=0 as the initial state with resistance R_0; state k as representing a line with resistance $R_k = R_0(1+k/1000)$ and k=N=100 as the failed state with resistance $R_{100} = 1 \cdot 1R_0$, and jumps of m=25 states. If the jumps were excluded, taking time units such that $\lambda dt = 0.025$ and $\alpha = \lambda/100$ the MTTF would be expected to be $N/\lambda = 4 \times 10^3$ units. Individual results for three lines with jumps included are shown in Fig. 4 while the complete failure time distribution, Eq. (17), is shown as the solid curve in Fig. 5. Note that again if N, N–m, N–2m, etc., are large, then the gamma distributions are roughly normal. The normal curves would be expected to be centered roughly on $100/\lambda, 75/\lambda, 50/\lambda, \ldots$ with variance $100/\lambda^2, 75/\lambda^2, 50/\lambda^2 \ldots$. Fig. 5 also shows the distribution fitted to a normal mixture (dashed curve) given by $f(t) \sim 0.67N(98.5/\lambda, 98.5/\lambda^2) + 0.3N(50/\lambda, 50/\lambda^2) + 0.03N(25/\lambda, 25/\lambda^2)$.

For aluminum with $\sigma \sim 0.6$, and thus $N \sim 3$, the only means of justifying the gamma mixture approach used in Ref. 40 is to remove the gradual degradation from the chain
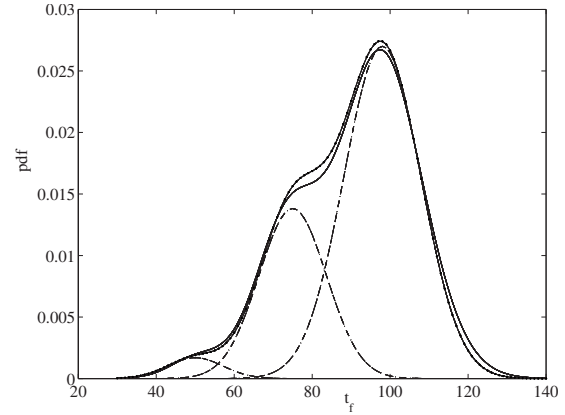

FIG. 5. The failure time distribution for N=100, a nearest neighbor transition rate of $\lambda dt = 1$, and a jump of 25 states with transition rate $\alpha dt = 5 \times 10^{-3}$. The dashed curve represents a fit using a mixture of normal distribution with means and variances of 98 (rather than 100), 75, and 50, with weights, respectively, of 0.67, 0.3, and 0.03. The dotted-dashed curves show the individual normal distributions.

so that N is kept much smaller than for copper, however artificial this may appear. In this way we assume that a (more or less fixed) time t_0 is required to create a void of sufficient size to cause failure. This time may be reduced by a small number of severe events which are modeled by the multistate chain. All the results above apply with small N. In the general case in which the transition rates are different between states, the failure distribution will be a sum of exponential functions. Where there are a smaller number of discrete transition rates Λ_k with multiplicity ν_k the result will be a convolution of gamma distributions. The assumptions required to generate a gamma mixture are (i) the gradual degradation is described by the chain which only describes severe events; (ii) the transition rates between neighboring states in the chain $\lambda_{k,k+1}$ are degenerate; and (iii) these transitions occur simultaneously with larger jump transitions to nonneighboring states occurring with small probability. These assumptions are too specific to justify the use of a gamma mixture without a careful factoring of the physical processes into the identification of the states R_k, their number N, and the relevant matrix of transition rates Q. Nonetheless the case of N=8, m=5 fits reasonably well to a single gamma distribution of n=3, Fig. 6 also shown is the nearest lognormal plot.

In the general case of arbitrary triangular (N×N) matrix Q, a certain amount can be said regarding any phase-type distribution, e.g., Refs. 54, 55, and 58. The best that one can say in relation to the accuracy the gamma mixture Ref. 40, is that the CDF of a phase-type distribution can be written as a true mixture of the CDFs of the basic paths in the chain.[54] Each path is a convolution of exponential or Erlang distributions (Erlang if some transition rates are repeated on the path) and each of these may be written as general mixtures of Erlang distributions of order N, although again the number of terms is large and some of mixture coefficients are negative.[59] By increasing further the number of Erlang distributions through dummy poles in the Laplace transform of the overall CDF, it is likely the mixture coefficients will eventually become positive.[54,55] This is largely a restatement
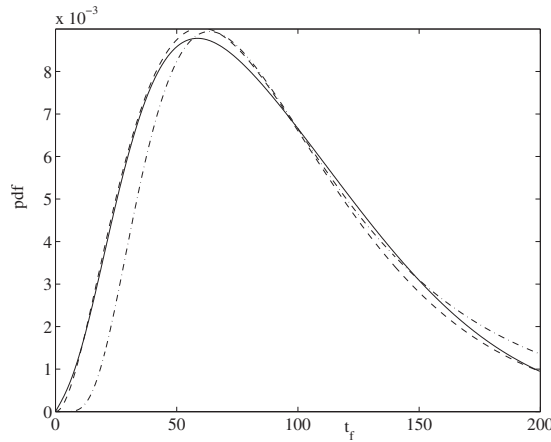
FIG. 6. The failure time distribution (full curve) for a chain with N=8, a nearest neighbor transition rate of $\lambda dt=0.01$ and a jump of five states with transition rate $\alpha dt=0.002$. The dotted-dashed curve is a lognormal with $\sigma=0.59$ and $t_{50}=90$, while the dashed curve is a gamma distribution $\Gamma(3,90)$.
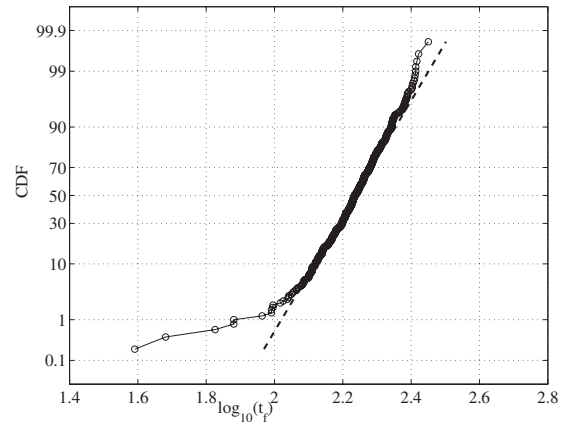

FIG. 7. A HMM with a gradual void growth in state 1, which increases the FRC at a constant rate of $\lambda dt=5\times10^{-5}$, with small jumps in FRC of 0.0175% in state 2 and near fatal increases of 7% in state 3. The transition matrix is taken arbitrarily as $Q=[0.85\ 0.15\ 0.00;0.3\ 0.6999\ 0.0001; 0.41\ 0.583\ 0.007]$.

of the fact that Erlang distributions are weakly dense in the space of pdfs, and is not really in the spirit of what is intended in Ref. 40.

C. A hidden Markov model (HMM)

There have been a number of attempts to link the vacancy build-up and void behavior to resistance plots, such as that by Doan *et al.*[10] on Al lines. In aluminum interconnect, void nucleation often plays a key role (particularly in the absence of a shunt layer) and can occupy a large portion of the failure time. Summarizing the results presented in Ref. 10 for Al lines, during the initial period vacancies gather until a void nucleates. The fractional change in resistance (FCR) increases relatively slowly (indeed it sometimes decreases but Doan *et al.*[10] conclude that this is not an EM effect). On nucleation the FCR increases gently as first slit-like void growth occurs along the line and then more strongly as transverse growth occurs. Following this phase a large jump is seen in the resistance plot as the current is forced into shunt layers (usually Ti/TiN or Ta/TaN). This suggests that the behaviors of an aluminum line are (i) void nucleating, (ii) slitlike void growth, (iii) transverse void growth, and (iv) a jump change. A HMM (Ref. 61) describing the system with a chain whose states are these *behaviors* would at least have its roots more firmly in the physical EM process. As a consequence we discuss briefly how such a model might be made to work.

It has been shown that, in the case of a pad-stud short copper interconnect, with one end blocked $J_{atom}(L)=0$ and stress relaxation at the other, $\sigma(x=0,t)=0$, the void growth occurs at a fairly steady rate which depends on the line microstructure and which appears to lead to a failure time distribution, which is close to lognormal. In the case of copper, the failure time was found to be given by[7]

$$t_f = \frac{\Delta V_{cr}}{J_{SS}\Omega} - \delta + t_{nucl}, \qquad (25)$$

where J_{SS} is the steady-state EM flux, δ is a correction factor which accounts for the fact that much of the vacancy

build-up before nucleation is quickly incorporated into the void volume, and t_{nucl} is the time for void nucleation. In the system considered δ and t_{nucl} are individually small compared to t_f as is common in many copper-based interconnect systems. They are also of similar magnitude so that cancellation occurs with $t_{nucl}-\delta\ll t_f$.[7] It was also found that if the diffusion activation energies are normally distributed then, by the permanence property of the lognormal when summed, J_{SS} is lognormal and consequently so is t_f. In this case no resistance jumps were allowed.

In a HMM the value of J_{SS} may be chosen from a lognormal distribution for each interconnect in the test batch. Likewise a nucleation current may be chosen for aluminum systems perhaps from a similar distribution. Jump changes in resistance could be allowed by changing the behavior-state. However the transition rates between states and the resistance jump mechanism and probabilities would require further study. The failure time CDF for a simple HMM is shown in Fig. 7 It is interesting in that the fit to lognormal data is good and similar to that seen in many EM experiments. We consider a very simple version of a HMM for a copper interconnect after void nucleation. In state k=1 (slitlike void growth), the resistance is assumed to increase at a constant rate $(dR/dt)dt=0.00005$, so that in the absence of any other behavior the failure time would be fixed with no variance. In state k=2 a small jump in resistance $\delta R=0.175\%$ occurs to model the faster, transverse growth behavior, while in state k=3 a near fatal $40\delta R=7\%$ jump occurs representing current shunting. Here the 3×3 Q matrix of transition probabilities between states is taken (without any justification) to be $Q=[0.85\ 0.15\ 0.00;0.3\ 0.6999\ 0.0001; 0.41\ 0.583\ 0.007]$. The variation in failure time leading to this lognormal behavior here is produced entirely by the probabilistic mechanism for resistance jumps. Failure times are determined here by how soon transverse growth occurs following nucleation and how soon after that current shunting occurs. The simulation works by starting the line in state k=1 in which the resistance is growing at a constant rate (although in an HMM this could be made stochastic also), it remains in that state for the time interval dt with probability

$Q_{1,1}=0.85$ and makes a transition to state $k=2$ with probability $Q_{1,2}=0.15$. Once in state $k=2$ the resistance jumps by δR, it makes a transition to state $k=1$ and a constant growth rate with probability $Q_{2,1}=0.3$, remains in state $k=2$ with probability $Q_{2,2}=0.6999$, in which further small jumps occur and makes a transition to state $k=3$, where it makes a possibly fatal jump of $40\delta R$ with probability $Q_{2,3}=0.0001$. The behavior in state $k=3$ is similar. 1000 lines were simulated in this manner, using the built in random number generator of MATLAB. The resulting CDF is plotted as a solid curve on lognormal paper in Fig. 7.

The HMM is perhaps more in the spirit of Ref. 62 as it allows for the integration of void growth at the cathode with arrival of microvoids. Their arrival amounts to something resembling a queuing system, and it may be that failure can be defined by a critical queue length, but again such a model requires careful consideration.

V. SUMMARY

When using Bayesian inference techniques, such as that the AIC information content and maximum likelihood estimation (MLE), it is important to have correct priors for the system being modeled. Historically the failure time distribution for a given EM failure mode has been assumed to be lognormal, and hence multimodal failure is assumed to be described by a lognormal mixture. Recently Tan *et al.*[40] reported that if the degradation process can be described by a multistate MC, the failure distribution will always be a gamma mixture and that such distributions should act as a prior for MLE methods. The current paper has considered such a system in more detail and concludes that, despite being able to generate resistance curves R(t), which are similar to typical curves seen in practice (in particular Refs. 10 and 36) overall we find that there is no evidence that the model generates a failure time distribution in the form of a true gamma mixture as suggested in Ref. 40. In general the model produces distributions which are mostly sums of exponentials (phase-type), convolutions of gamma distributions (which are general rather than true mixtures, with many terms), or close to normal. It is possible to choose the state number N (the length of the chain), states and transition rates to lead to an approximate gamma mixture, but that the choice merely reflects the inherent flexibility of MCs and appears somewhat arbitrary here. In particular the number of states is closely related to the lognormal standard deviation which can be very different for different copper technologies. As a result, we conclude that MC models are unable to justify the widespread use of a gamma mixture as a prior in failure distribution fitting and that, in the form suggested by Tan *et al.*,[40] the MC model is probably inappropriate to EM failure.

A HMM in which the states describe the current behavior of the line (e.g., slitlike void growth, sudden shape change, etc.), rather than the current resistance R(t), may be a more appropriate application of such models. However the nature of such states needs much greater investigation before such a model could be workable.

APPENDIX: FAILURE DISTRIBUTION FOR PROBLEM (ii)

If all the transition rates out of the states k are different then Q may be diagonalized. Then the solution to Eq. (13) is simply

$$\underline{p}(t) = \exp(Qt)\underline{p}(0) = S\exp(Dt)S^{-1}\underline{p}(0). \qquad (A1)$$

The reliability r(t) of the system is

$$r(t) = 1 - p_N(t) = \sum_{k=0}^{N-1} p_k(t) = \underline{1}_N^T \underline{p}(t), \qquad (A2)$$

where $\underline{1}_N$ is the vector of length N whose entries are all unity. Assuming that the system starts in state $k=0$, the initial vector $p(0)=\underline{e}_0$, where $\underline{e}_0$ is the unit vector $[1,0,0,\dots]^T$. Thus

$$r(t) = \underline{1}_N^T S \exp(Dt) S^{-1} \underline{e}_0. \qquad (A3)$$

The failure distribution is

$$f(t) = -\underline{1}_N^T SD \exp(Dt) S^{-1} \underline{e}_0 = \underline{w}^T S \exp(Dt) S^{-1} \underline{e}_0 \qquad (A4)$$

as $SD=QS$ and $\underline{1}^T Q=-\underline{w}^T$. This may be written as

$$f(t) = \sum_{k=0}^{N-1} (\underline{w}S)_k (S^{-1}\underline{e}_0)_{k0} \exp\left(-\sum_{j=k+1}^{N-1} \lambda_{kj} t\right). \qquad (A5)$$

The mean time to failure is, as usual,

$$\int_0^\infty r(t)dt = \underline{1}_N^T SD^{-1} S^{-1} \underline{e}_0 = \underline{1}_N^T Q^{-1} \underline{e}_0. \qquad (A6)$$

Here Q (or D) may be inverted as none of its eigenvalues are zero. There remains only to find the matrix S whose columns are the eigenvectors $\underline{u}^{(k)}=[u_0^{(k)},u_1^{(k)},\cdots,u_{N-1}^{(k)}]^T$ corresponding to eigenvalues Λ_k of Q. For the kth eigenvalue, the matrix Q may be block partitioned thus

$$Q = \begin{bmatrix} Q_{1,k} & 0 & 0 \\ \underline{q}_{1,k} & \Lambda_k & 0 \\ Q_{2,k} & \underline{q}_{2,k} & Q_{3,k} \end{bmatrix}. \qquad (A7)$$

$Q_{1,k}$ is the $(k-1)\times(k-1)$ lower triangular matrix at the top left of Q, $q_{1,k}$ is the $1\times(k-1)$ vector at the far left of row k of Q, and $q_{2,k}$ is the $(N-k)\times 1$ vector at the bottom of column k of Q. $Q_{3,k}$ is the $(N-k)\times(N-k)$ lower triangular matrix in the bottom right of Q and $Q_{2,k}$ is the $(N-k)\times(k-1)$ matrix at the bottom left of Q. Assuming that all the Λ_k are different, Λ_k will not be an eigenvalue of $Q_{3,k}$ and hence $(\Lambda_k I - Q_{3,k})$ is invertible. Noting the result that

$$Q_{3k}(\Lambda_k I - Q_{3k})^{-1}\underline{q}_{2k} = \Lambda_k(\Lambda_k I - Q_{3k})^{-1}\underline{q}_{2k} - \underline{q}_{2k} \qquad (A8)$$

it is clear that the eigenvector corresponding to Λ_k is $[0,\dots,0,1,v_k]^T$, where 1 appears in the kth position and v_k is an $(N-k)\times 1$ vector which satisfies

$$(\Lambda_k I - Q_{3,k})\underline{v}_k = \underline{q}_{2k}. \qquad (A9)$$

The solution to Eq. (A9) is straightforward by forward substitution and consequently the matrix of eigenvectors S may be obtained easily. It is clear that S will also be lower trian-

gular. This may be inverted to give S^{-1} by recursively using, in an obvious notation,

$$S^{-1} = S_N^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ v_{0,1} & 1 & 0 & 0 \\ \vdots & v_{1,1} & \ddots & \vdots \\ v_{0N-1} & \cdots & v_{N-1,1} & 1 \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} 1 & 0 \\ \underline{v}_0 & S_{N-1} \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ -S_{N-1}^{-1}\underline{v}_0 & -S_{N-1}^{-1} \end{pmatrix}. \quad (A10)$$

The normalization of the eigenvectors presents no problem here.

[1] J. R. Lloyd, J. Phys. D: Appl. Phys. **32**, R109 (1999).
[2] Z. S. Choi, C. L. Gan, F. Wei, C. V. Thompson, J. H. Lee, K. L. Pey, and W. K. Choi, in Materials, Technology, and Reliability for Advanced Interconnects and Low-K Dielectrics-2004, edited by R. J. Carter, C. S. Hau-Riege, G. M. Kloster, T.-M. Lu, and S. E. Schulz (Mater. Res. Soc. Symp. Proc. No. 812, Warrendale, PA, 2004), Paper No. F7.6.
[3] C.-K. Hu and B. Luther, Mater. Chem. Phys. **41**, 1 (1995).
[4] E. Liniger, L. Gignac, C.-K. Hu, and S. Kaldor, J. Appl. Phys. **92**, 1803 (2002).
[5] A. V. Vairagar, S. G. Mhaisalkar, and A. Krishnamoorthy, Microelectron. Reliab. **44**, 747 (2004).
[6] E. T. Ogawa, K. D. Lee, V. A. Blaschke, and P. S. Ho, IEEE Trans. Reliab. **51**, 403 (2002).
[7] V. M. Dwyer, J. Appl. Phys. **104**, 053708 (2008).
[8] A. S. Oates and S. C. Lee, Microelectron. Reliab. **46**, 1581 (2006).
[9] J. R. Lloyd, "Electromigration for Designers," http://www.edadesignline.com/.
[10] J. C. Doan, J. C. Bravman, P. A. Flinn, and T. N. Marieb, Microelectron. Reliab. **40**, 981 (2000).
[11] D. Tio Castro, R. J. O. M. Hoofman, J. Michelon, and D. J. Gravesteijn, J. Appl. Phys. **102**, 123515 (2007).
[12] C.-K. Hu, L. Gignac, and R. Rosenberg, Microelectron. Reliab. **46**, 213 (2006).
[13] A. S. Oates and M. H. Lin, Proceedings of the IEEE International Reliability Physics Symposium, Phoenix, AZ, 2008 (unpublished), Vol. 46, p. 385.
[14] A. V. Vairagara, S. G. Mhaisalkar, A. Krishnamoorthy, K. N. Tu, A. M. Gusak, M. A. Meyer, and E. Zschech, Appl. Phys. Lett. **85**, 2503 (2004).
[15] M. Hauschildt, M. Gall, S. Thrasher, P. Justison, R. Hernandez, H. Kawasaki, and P. S. Ho, Appl. Phys. Lett. **101**, 043523 (2007).
[16] L. Zhang, S. Ou, J. Huang, K. N. Tu, S. Gee, and L. Nguyen, Appl. Phys. Lett. **88**, 012106 (2006).
[17] K. N. Tu, C. C. Yeh, C. Y. Liu, and C. Chen, Appl. Phys. Lett. **76**, 988 (2000).
[18] E. C. C. Yeh, W. J. Choi, K. N. Tu, P. Elenius, and H. Balkan, Appl. Phys. Lett. **80**, 580 (2002).
[19] J. W. Christian, The Theory of Transformations in Metals and Alloys, 3rd ed. (Pergamon, Oxford, 2002), Part I, Chaps. 10 and 11.
[20] V. Sukharev, E. Zschech, and W. D. Nix, J. Appl. Phys. **102**, 053505 (2007).
[21] A. F. Bower and S. Shankar, Modell. Simul. Mater. Sci. Eng. **15**, 923 (2007).
[22] M. R. Gungor and D. Maroudas, J. Appl. Phys. **101**, 063513 (2007).
[23] C. Basaran, S. Li, and M. F. Abdulhamid, J. Appl. Phys. **103**, 123520 (2008).
[24] T. O. Ogurtani and O. A. Akyildiz, Int. J. Solids Struct. **45**, 921 (2008).
[25] A. Averbuch, M. Israeli, M. Nathan, and I. Ravve, J. Comput. Phys. **188**, 640 (2003).
[26] T. O. Ogurtani and E. E. Oren, J. Appl. Phys. **96**, 7246 (2004).
[27] O. Kraft and E. Arzt, Acta Mater. **45**, 1599 (1997).
[28] M. Schimschak and J. Krug, J. Appl. Phys. **87**, 695 (2000).
[29] M. Mahadevan and R. M. Bradley, Phys. Rev. B **59**, 11037 (1999).
[30] M. Nathan, A. Averbuch, and M. Israeli, Thin Solid Films **466**, 347 (2004).
[31] M. R. Gungor and D. Maroudas, J. Appl. Phys. **85**, 2233 (1999).
[32] M. P. Surh, J. B. Sturgeon, and W. G. Wolfer, J. Nucl. Mater. **325**, 44 (2004).
[33] H. Trinkaus and M. H. Yoo, Philos. Mag. A **55**, 269 (1987).
[34] A. A. Semenov and C. H. Woo, Phys. Rev. B **66**, 024118 (2002).
[35] J. H. Abeles, W. K. Chan, E. Colas, and A. Kastalsky, Appl. Phys. Lett. **54**, 2177 (1989).
[36] R. Rosenberg, D. C. Edelstein, C.-K. Hu, and K. Rodbell, Annu. Rev. Mater. Sci. **30**, 229 (2000).
[37] M. Ding, G. Wang, B. Chao, P. S. Ho, P. Su, and T. Uehling, J. Appl. Phys. **99**, 094906 (2006).
[38] M. H. Lin, Y. L. Lin, K. P. Chang, K. C. Su, and T. Wang, Jpn. J. Appl. Phys., Part 1 **45**, 700 (2006).
[39] E. Zschech and V. Sukharev, Microelectron. Eng. **82**, 629 (2005).
[40] C. M. Tan, N. Raghavan, and A. Roy, J. Appl. Phys. **102**, 103703 (2007).
[41] D. B. Kececioglu, Reliability Engineering Handbook (Prentice Hall, Englewood Cliffs, 1991), Vol. 1.
[42] J. F. Lawless, Statistical Models and Methods for Lifetime Data (Wiley, New York, 1982).
[43] H. Akaike, Ann. Inst. Stat. Math. **30**, 9 (1978).
[44] A. P. Dempster, N. M. Laird, and D. B. Rubin, J. R. Stat. Soc. Ser. B (Methodol.) **39**, 1 (1977).
[45] V. M. Dwyer, J. Phys. D: Appl. Phys. **37**, 422 (2004).
[46] V. Luzin, Textures Microstruct. **31**, 177 (1999).
[47] D. Weaire, J. P. Kermode, and J. Wejchert, Philos. Mag. B **53**, L101 (1986).
[48] M. Fátima Vaz and M. A. Fortes, Scr. Metall. **22**, 35 (1988).
[49] J. Cho and C. V. Thompson, Appl. Phys. Lett. **54**, 2577 (1989).
[50] L. Arnaud, G. Tartavel, T. Berger, D. Mariolle, Y. Gobil, and I. Touet, Microelectron. Reliab. **40**, 77 (2000).
[51] M. Hauschildt, M. Gall, S. Thrasher, P. Justison, L. Michaelson, R. Hernandez, H. Kawasaki, and P. S. Ho, Appl. Phys. Lett. **88**, 211907 (2006).
[52] M. H. Lin, K. P. Chang, K. C. Su, and T. Wang, Microelectron. Reliab. **47**, 2100 (2007).
[53] C. Ryu, A. L. S. Loke, T. Nogami, and S. S. Wong, Proceedings of the IEEE International Reliability Physics Symposium, Denver, CO, 1997 (unpublished), Vol. 35, p. 210.
[54] A. Cumani, Microelectron. Reliab. **22**, 583 (1982).
[55] C. A. O'Cinneide, Commun. Stat. Stoch. Models **15**, 731 (1999).
[56] B. Ko, A. C. Y. Ng, and N. Am, Scand. Actuar. J. **11**, 170 (2007).
[57] J. A. Koziol, J. Theor. Biol. **180**, 81 (1996).
[58] J. V. Grice and L. Bain, J. Am. Stat. Assoc. **75**, 929 (1980).
[59] J. Mi, W. Shi, and Y. Y. Zhou, Stat. Probab. Lett. **78**, 2378 (2008).
[60] A. Hovath and M. Telek, in Advances in Algorithmic Methods for Stochastic Models, edited by G. Latouche and P. Taylor (Notable Publications, New Jersey, 2000), p. 191.
[61] S. M. Ross, Introduction to Probability Models (Academic, New York, 2006), p. 256.
[62] C. W. Chang, C. V. Thompson, C. L. Gan, K. L. Pey, W. K. Choi, and Y. K. Lim, Appl. Phys. Lett. **90**, 193505 (2007).