

Electromigration is the motion of atoms in a conductor due to an applied electric field. The forces acting on these ionized atoms are (a) the direct electrostatic force acting on the charge and (b) the force exerted by free carrier collisions. In metals, this force on ions exerted by collisions with the "electron wind" of free carriers is called the "wind force" and is usually the dominant component of the total force. Since the two forces are in opposite directions, the greater of the two determines in which direction the ions will flow (1,2).

By far the greatest technological and economic significance of electromigration is its deleterious effect on the reliability of metal interconnects found in integrated circuits (ICs). Generally, it is the wind force that causes the motion of atoms within the metallic interconnects from the "upwind" (cathode) terminal to the "downwind" (anode) terminal, leading eventually to both (a) open circuits from a depletion of material and (b) short circuits between conducting lines, due to the formation of metal hillocks or whiskers which can bridge the dielectric between conductors. Penetration of passivation layers by hillocks or whiskers can also expose metal and lead to failure by corrosion. Figure 1 shows an electron micrograph of a void and whisker on an aluminum line.

The first observations of electromigration occurred in molten alloys, where the large ionic diffusivity D enhances the magnitude of the effect. The original studies were conducted by Gerardin in 1860 (3), and over the next hundred years most studies continued to be performed in liquid metals or at high temperatures. The earlier studies focused on electromigration as a separation technique. Electromigration did not become commercially important until the development of ICs.

Electrical connections between IC devices consist of segments oriented parallel and perpendicular to the wafer surface (Fig. 2). The horizontal sections are referred to as runners, stripes, or lines; and the vertical sections, referred to as posts, studs, or plugs, are formed within openings in the dielectric called vias (4). Interconnects in ICs consist of some combination of layers of polycrystalline metals, most commonly Al alloys and refractory metals. Multilevel interconnects are separated by insulating interlevel dielectrics (ILDs). It is predicted (5) that as many as nine interconnect levels,

Figure 1. Electron micrograph of aluminum line exhibiting electromigration effects. The whisker to the left and the voids to the right are important reliability problems in metal interconnect technology. (Photograph: R. Frankovic and G. H. Bernstein.)

linked by vias, will be employed in ICs by the year 2012. The layers which comprise these levels are thin metallic films intimately laminated on each other. For Al alloys, each level is composed of some three to five layers of different metallic components. The evolution from Al alloys to pure Cu will be discussed at the end of this article.

Nearly every characteristic of metal lines has a measureable effect on electromigration lifetime. Included among these characteristics are the alloy compositions, layer thicknesses, line dimensions and shapes, crystallographic orientation of the grains, method of deposition, annealing and heat treatments, the material, thickness, and other details of the passivation layer (overcoat), and terminations at both interconnect-level vias and ohmic contacts. In addition, external factors such as temperature profiles, current densities, and the nature of the time-dependent waveform [i.e., direct current (dc), pulsed, or alternating current (ac)] in operation affect the lifetime. As a further complication, the definition of failure depends on the particular application. It may be a completely open circuit, it may involve some percentage of increase in resistance, or it may be a short circuit between lines. For these reasons, nearly all designs relating to reliability of interconnects due to electromigration require phenomenological studies, as opposed to theoretical modeling. Although many theoretical studies have been performed over the past 30 or so years to understand details of electromigration-induced behavior, a comprehensive theory to predict interconnect reliability based on first principles has not been achieved. Therefore, the approach used by IC designers is to establish simple constraints based on extensive tests on real interconnects—for example, 5×10^5 A/cm² at 55°C (6). Relaxing such design rules would in some cases result in immediate gains in IC performance and/or profit margin, but any change in a specified design rule or metallization scheme is a matter of real concern.

Since ICs must be capable of operating continuously for periods of years or decades, it is impractical to test the life-



Figure 2. Schematic cross section of multilevel metallization scheme. Ohmic contacts form electrical connection between metal levels and the substrate. The polysilicon/metal dielectric (PMD) separates the substrate and polysilicon material from metal level 1. A dielectric material, designated DM1, separates metal level 1 from metal level 2. Other dielectric layers are named similarly. Vias filled with refractory plugs connect metal levels. A section of a metal level terminating near a via, referred to as a reservoir, provides replacement material as voiding occurs at the downstream end of a plug. A dielectric passivation protects the IC everywhere except at openings for external electrical connection.

times of metal lines by operating them at use conditions and waiting for a statistically significant sample of failure times. Therefore, accelerated measurement techniques are employed in which interconnects are stressed at elevated temperatures, typically 175° C to 275° C, and current densities, typically 10^{6} A/cm² to 10^{7} A/cm². This combination of stresses frequently lowers the time to failure (TTF) from years to weeks or days. Owing to the complexity of possible failure mechanisms, extrapolation from the stress conditions to lifetimes during use is highly approximate, but one general rule describing the median time to failure (MTTF), due to Black (7), has been accepted as the standard extrapolation:

$$MTTF = CAj^{-n} \exp(E_a/kT)$$
(1)

where C is an experimentally determined constant which depends on the process, A is the cross sectional area of the interconnect, j is the current density, T is the absolute temperature, n is a positive constant, k is the Boltzmann constant, and E_a is an activation energy. This equation was proposed by Black on the basis of empirical results. Due to the importance of having a reliable rule for extrapolating to use condi-

tions, great effort has gone into determining the parameters. The utility of Black's equation depends on determining a physical basis for its components so that the activation energy and therefore the dominant failure mechanism can be accurately determined. Black originally proposed a value of 2 for n, but subsequent studies often found better fits using other values of n, typically ranging between 1 and 3, with values as high as 15 used.

The exponent n in Black's equation can also be investigated in pulsed-current experiments. If it is assumed that electromigration acts to cause degradation instantaneously only during current flow, then at a fixed temperature, Black's equation predicts an MTTF proportional to the time average of j^{-n} . In particular, in pulsed dc at a fixed time-average current, with a duty cycle r (current flowing a fraction r of the time), it predicts an MTTF proportional to r^{-n} . In practice, this behavior is observed at frequencies no higher than about 1 kHz (8). At higher frequencies and in ac current, MTTFs can be considerably longer than predicted by this extrapolation from Black's equation. This deviation is generally taken as an indication that small amounts of electromigration damage caused by brief current stressing can either heal spontaneously in zero current or be reversed by current of opposite sign.

It is widely accepted that failure times are adequately described by a lognormal distribution (9,10). A lognormal distribution of failures is a normal, or Gaussian, distribution in the logarithm of time to failure. That electromigration failures follow a lognormal distribution has been attributed to normal distributions of activation energies (11) and oven temperatures (9). How well a failure distribution fits a lognormal form is usually determined by creating a "Weibull plot," that is, by plotting the logarithm of stress time versus cumulative failures, where the axis for failures is nonlinearly scaled so that a lognormal distribution of the data appears as a straight line. A sample Weibull plot is shown in Fig. 3. Since measurements are often very time-consuming, the median time to fail-



Figure 3. Example of electromigration-induced lifetime data displayed as a Weibull plot exhibiting lognormal behavior. The median time to failure at which 50% of samples failed is denoted as t_{50} . The shape parameter σ is determined from the slope. (Data of R. Frankovic and G. H. Bernstein.)

ure (i.e., that time for which half of the test samples have failed) is more easily obtained than is the mean time to failure, which requires that all samples fail, including the long-lived ones, before a mean can be determined. The variable name t_{50} emphasizes that MTTF should stand for *median* rather than *mean* TTF. The deviation in the time to failure (DTTF), usually denoted by σ , is a measure of the width of the lognormal distribution. For a given value of MTTF, a greater σ implies a greater frequency of early failures and is undesirable for ICs. An IC lasts only as long as its shortest-lived component, so achieving a small value of σ , and hence an acceptable frequency of early failures, can be more critical to the overall reliability of an IC chip than lowering the MTTF.

The temperature dependence exhibited by Black's equation is called activated, or Arrhenius, behavior. The value of the activation energy, $E_{\rm a}$, often indicates which failure mechanism dominates, allowing interconnect engineers to modify their processes to increase reliability. Possible failure mechanisms include atomic diffusion along grain boundaries, within the atomic lattice of the grains, and along heterointerfaces and free surfaces. Thermal expansion mismatches between the metallic runners and the wafer substrate can create large stresses in the lines and also lead to failure. This phenomenon, referred to as "stress migration," or "stress voiding," is intimately related to electromigration. In general, E_{a} can be determined from the slope of an Arrhenius plot of the MTTF (a plot of logarithm of the failure time versus 1/T). Various diffusion mechanisms give rise to a wide range in $E_{\rm a}$ values. As an example, E_a for grain boundary diffusion of Al is approximately 0.55 eV (12), whereas E_a for bulk, or lattice, diffusion (through the grains) in Al is about 1.5 eV (13). This difference in activation energy typically accounts for the dominant role of grain boundary diffusion in wide, polycrystalline metal interconnects. As an illustration of this phenomenon, d'Heurle and Ames (14) grew single-crystal aluminum lines and found them to have lifetimes many orders of magnitude greater than comparable polycrystalline lines tested under like conditions.

Electromigration through a homogeneous region causes material to flow without creating voids or hillocks. However, in the presence of any inhomogeneity, such local changes in material occur. If a region either accumulates or is depleted of atoms—that is, if the flux divergence is nonzero—then a macroscopic defect in the form of a hillock or void, respectively, will arise. Conversely, if atoms flow through a region (zero flux divergence), then neither hillocks nor voids will form.

The flux of atoms due to the total electromigration force is given by

$$\boldsymbol{J} = CDZ^* e\rho \boldsymbol{j}/kT \tag{2}$$

where C is the concentration of atoms, D is the atomic diffusivity, Z^* is their effective valence due to both the direct and wind forces, e is the magnitude of the electron charge, ρ is the resistivity of the metal line, and **j** is the current density.

METALLIC GRAIN EFFECTS

Variation in any of the material parameters in Eq. (2) along the line leads to flux divergence. It was recognized very early in the study of IC-related electromigration failure that an important source of flux divergence is "triple junctions" (or "triple points") at which three grains meet. Figure 4 shows flux divergence at triple junctions. Since the particle current along the grain boundaries into the triple junction does not equal the particle current out, nonzero flux divergence occurs, and nucleation of voids or hillocks can begin there. Because of this effect, much effort has been put into coaxing grains to grow much larger than the linewidth during metal deposition and annealing, to minimize the number of triple junctions (15). The resulting "bamboo structure" minimizes diffusion along grain boundaries. MTTF is increased because all the remaining mechanisms for electromigration are less effective. The next easiest mechanism for electromigration is through diffusion in the bulk, for which the activation energy is much larger. Grains that span the width of the line ("blocking grains") impede diffusion along grain boundaries, and the lower density of triple junctions decreases related failures.



Figure 4. Schematic diagram of grain boundary orientations at triple junctions: (a) General orientation of grains, (b) one grain boundary perpendicular to electron wind, (c) symmetric triple junction leading to void formation, and (d) leading to hillock formation.

The number of downwind versus upwind grain boundaries at a triple junction does not alone determine whether any depletion or accumulation will occur. The ionic current along a grain boundary is proportional to the component of electron wind along that boundary (16). A particular consequence of this fact is that grain boundaries perpendicular to the electronic current direction do not contribute to the net atomic flux divergence. This accounts for the relative longevity of bamboo-structured interconnects.

If rays drawn along the grain boundaries from a triple junction are oriented at angles θ_1 , θ_2 , and θ_3 away from the electron wind direction [see Fig. 4(a)] then the atomic current out of the triple junction is proportional to $\cos \theta_1 + \cos \theta_2 + \cos \theta_3$. In Fig. 4(b), depletion (accumulation) occurs at the triple junction if $\cos \theta_1 + \cos \theta_2$ is greater than (less than) zero. By a trigonometric identity, symmetrically arranged grain boundaries ($2\pi/3$ angle separations) lead to zero net accumulation or depletion. In a symmetric configuration such as illustrated in Fig. 4(c,d), with one grain boundary along the wind direction, whether the triple junction is accumulating or depleting depends on the opening angle. In Fig. 4(c), with α less than $2\pi/3$, depletion occurs; in Fig. 4(d), with $\alpha > 2\pi/3$, accumulation occurs.

Pure aluminum is highly susceptible to electromigrationinduced failure, but it has been found that the addition of impurities can dramatically decrease this susceptibility. The most common impurity used for this purpose is copper, in concentrations ranging from a typical value of 0.5 atomic percent (at. %) up to 4 at. %. Small amounts of silicon, typically 1 at. % in the lowest metallic layer, are sometimes added to the mixture in order to saturate the Al and prevent Si diffusion into the metal at ohmic contacts. Si makes a negligible improvement in electromigration reliability (17). The effects of adding Cu to Al are not permanent. Cu diffuses along grain boundaries, just as Al does, leaving depleted material behind. $CuAl_2$ precipitates can dissolve at the cathode (18), thus replenishing Cu until it is exhausted. Once depleted of Cu, the material becomes susceptible to electromigration failure (1,14). X-ray fluorescence mapping of AlCu lines has shown that electromigration failure is more likely to occur in regions of Cu depletion (19). The net result, though, is vastly improved lifetimes over those of pure Al.

THE ROLE OF STRESS

An important related aspect of electromigration is the role of mechanical stress on the diffusion of ions. Mechanical stress exerts a third force on the metal atoms, distinct from those of the direct and wind forces. Mechanical stress can be induced by the fabrication process or by the forces of electromigration themselves. Migration due to stresses arises as an important issue in two ways discussed below: (a) the existence of a shelf life due to stress-induced degradation, known as "stress migration," and (b) a strong deviation from Black's equation due to the interplay of stress and wind forces in relatively short interconnects (20).

Many IC fabrication steps are performed at elevated temperatures, including deposition of interlevel dielectrics. As the semiconductor wafer cools from, say, 400°C, the higher thermal coefficient of expansion (TCE) of the metal relative to that of the substrate results in severe tensile straining of the interconnects (21). Because the metal is pinned on all sides by some dielectric material (either the ILD or surface passivation), simple shrinkage, which would require slip at the metal-dielectric interface, does not take place (22). Instead, these stresses are relieved by migration of metal atoms, resulting over time in voiding which, like that caused by electromigration, tends to be associated with grain boundaries. Voids due to stress migration tend to form distinctive thin cracks or triangular notches at the edges of lines. Stress is greatest at sharp features such as steps, contacts, or corners (23).

This phenomenon was not reported until 1984 (24,25), since migration is more severe for the narrower lines and generally thicker dielectrics found in multilevel interconnects (26). It was found that circuits showed damage or even failed after storage with no current passing through the interconnects. This raised extremely serious concerns for the IC industry.

Just as mechanical stress can lead to migration, migration caused by other effects can lead to mechanical stress. In particular, electromigration generates a stress gradient between the compressed anode end of the line and the cathode (27,28). This stress gradient causes a backflow in the atomic flux which opposes the flux due to the wind force. For sufficiently small current density or short lines, equilibrium is established between the wind force and gradient of mechanical stress. For sufficiently large current density or long lines, the stress at the anode exceeds the critical stress required for plastic deformation of the material, and hillocks can form. On the cathode end, atoms can be swept toward the anode by a combination of the wind force and mechanical stresses. The stress at the anode end is the stress gradient times the conductor length. For a given current there exists a line length, called the Blech length or critical length, below which the stress at the anode is below the critical stress of the conductor material, the net flux of atoms is zero, and MTTF is dramatically large. The relationship between current density and this line length is $jl_c = \kappa$, where l_c is the critical length and κ is a constant (21).

Encapsulation by hard passivation layers and interlevel dielectrics has two competing effects on the interconnect lifetimes. First, it constrains the volume of the conductor so that the critical stress is increased. Accumulation of atoms at the anode creates a compressive stress, and tensile stress at the cathode is associated with accumulation of vacancies. This stress gradient opposes further electromigration of atoms and increases the lifetimes. The effect occurs without passivation as long as the Blech length is not exceeded, but is enhanced in the presence of a hard passivation. On the other hand, a passivation, if under compressive stress, will increase the tensile stress in the metal lines, and migration of atoms toward regions of higher tensile stress at regions of stress gradient will relieve the stress by forming voids. Such voids can then serve to decrease the overall electromigration lifetimes. Therefore, the ideal passivation layer is strong enough to constrain the volume of the material without yielding to compressive stresses and cracking, but is not under compressive stress relative to the line.

Stress migration is more difficult to induce than is electromigration, so its study is correspondingly difficult. In addition, it exhibits complicated non-Arrhenius behavior. This arises from a combination of an Arrhenius dependence of the

mobility and a linear dependence of the driving force. At low temperatures, mobility of the atoms is lower, although the driving force, which is proportional to the difference in temperature between the deposition of the interlevel dielectric and that of the test condition or storage, is higher (29). At higher temperatures, the force is lower but the mobility is higher. It turns out that void formation is maximized at about 175° C to 200°C for commonly used IC materials.

MICROSCOPIC EFFECTS

A variety of experiments have established that in a constant field, ions in a homogeneous conductor electromigrate at a constant drift velocity. This was first determined by direct measurement; for example, penetration of electromigrating gold atoms into a Cu matrix was measured by chemically etching through the copper and chemically determining the Au concentration. Seith and Wever initiated the use of marker motion experiments [see (30)], which measure electromigration by tracking the motion of an indentation on the surface of a wire. These experiments show a similar linear behavior. More recently, the inverse dependence of the Blech length on current confirms the linear dependence of electron wind force on current density.

The linear dependence of the electron wind on the current is generally understood as a consequence of ordinary atomic diffusion, characterized by the diffusivity D. In a metal, the free carriers arise from ionization of the bulk of the atoms (i.e., metallic bond formation), which have a charge $Z_{ion}e$. For ions moving in response to electric field alone, Einstein's relation predicts a mobility $\mu_{ion} = Z_{ion}eD/kT$, so that in an electric field \boldsymbol{E} the atoms drift at an average velocity

$$\boldsymbol{v}_{\rm d} = \mu_{\rm ion} \boldsymbol{E} \tag{3}$$

A way to understand Eq. (3) is to say that a force $\mathbf{F}_D = \mathbf{Z}_{ion} \mathbf{e} \mathbf{E}$ acting on an ion is balanced by microscopic friction forces when the average velocity is $D\mathbf{F}_D/kT$.

Ordinary electrical resistance arises from the collision of electrons with defects and lattice vibrations (phonons); and in these collisions, momentum is transferred to the lattice. This momentum transfer gives rise to the electron wind force acting on the ions. Although the electron wind force differs from the direct electrostatic force, it is resisted by the same microscopic forces that determine the diffusivity and the drift velocity arising from the direct force. Thus, the drift velocity arising from the electron wind force F_{ew} can also be written DF_{ew}/kT . The rate of momentum transfer from the electrons is proportional to the current density \mathbf{j} , and this is related to the electric field by the microscopic form of Ohm's law:

$$\boldsymbol{E} = \rho \boldsymbol{j} \tag{4}$$

where ρ is electrical resistivity. Hence, the component of drift induced by the electron wind has the same dependence on diffusivity and **E** as the direct force, and can be combined in a form similar to Eq. (3), so one can write the drift velocity as

$$\boldsymbol{v}_{\rm d} = \frac{Z^* e D}{kT} \boldsymbol{E} \tag{5}$$

arising from the combined effects of direct and electromigration forces. The parameter Z^* [introduced by Skaupy (31)] contains a contribution $Z^* - Z_{\rm ion}$ which encapsulates the dependence of electron wind force on the electric field that gives rise to current. The factor Z^* , on the order of 10, usually has the same sign as the majority carriers (electrons or holes) in the conductor, indicating the dominance of electron wind over direct force. Wever and Seith (32) demonstrated this dramatically in the Al–Cu system: The β phase, in which conduction is by electrons, exhibits ion transport toward the anode; in the γ phase, which is a hole conductor, ions move toward the cathode.

Moving with drift velocity given by Eq. (5), a particle density C corresponds to an ion flow that is described by the particle current density or ion flux J, a vector field equal to the drift velocity times the particle density, introduced in Eq. (2).

According to the continuity equation, the time rate of change of particle density equals minus the divergence of the ionic current. From Eq. (2), we obtain

$$-\operatorname{div} \boldsymbol{J} = -\operatorname{grad} \left(\frac{C\rho Z^* e D}{kT} \right) \cdot \boldsymbol{j} - \frac{C\rho Z^* e D}{kT} \operatorname{div} \boldsymbol{j}$$
(6)

where we have used Ohm's law [Eq. (4)] to eliminate E.

Under dc conditions, the continuity equation for electrons (and holes) implies that div j = 0, so the second term on the right-hand side of Eq. (6) vanishes.

It is important to note that a mathematical divergence div J or div j, which indicates particle density change, is not the same as the "divergence" (in the sense of increasing separation) of current streamlines, which indicates direction of current flow. To take the example of electronic current in a wire, at an increase in wire diameter, the current spreads outward from the center. Following a tube of current through this width change, one finds an area change that is exactly compensated by a change in current density, so no electron density change over time occurs in any fixed region of space.

Continuous current streamlines correspond to a situation of zero divergence. When atomic current streamlines are represented graphically, the nonzero div J corresponding to time rate of change in density is represented by the termination of streamlines, just as nonzero charge density corresponds to the beginning and end of electric field lines.

Under normal conditions, then, changes in mass arise from the nonvanishing (first) term on the right-hand side of Eq. (6). That is, depletion and accumulation require some variation in C, ρ , Z^* , D, or T along the current direction \mathbf{j} . Such variations can arise from a number of different factors. Abrupt changes in material parameters are associated with microscopic grain structure, as discussed above, and macroscopic transitions between conductor materials in a device. Also, variations in concentrations C, along with associated variations in resistivity ρ , can occur within a single phase of material. Temperature T, because it can be increased by Joule heating, is affected by any factor that affects electronic current.

In a transition between different materials, one may generally expect material parameters such as ρ , Z^* , and D to change abruptly. A common instance of particle accumulation and depletion associated with material interfaces occurs at vias in integrated circuits. Within a metallization layer, the interconnects, which have high electromigration mobility μ , terminate at plugs of a refractory material such as tungsten or a salicide. Although the refractory materials typically have larger resistivities ρ , this change is more than compensated by sharply lower ionic diffusivities D, so they have lower electromigration mobilities μ . For the usual case of Z^* negative in the (aluminum, gold or copper) interconnect, particle accumulation occurs if there is a decrease in diffusivity (a plug) in the direction of the anode. At such an interface, electrons flow into the plug, and electromigrating ions, blocked from entering the plug, accumulate. Conversely, voids can form where electrons flow from a tungsten plug into aluminum.

Void and hillock formation (atomic flux divergences) at discontinuities associated with discontinuities such as grain boundaries and bulk material interfaces are known to be important mechanisms in electromigration failure. However, these mechanisms have common features that are problematical.

The first problem is that Eq. (6) predicts a flux divergence proportional to the electronic current, implying that n in Black's equation [Eq. (1)] should be unity: The MTTF is the time it takes to deplete enough material to cause failure, and the rate of material loss is proportional to electronic current. This is inconsistent with typical observed values of n around two. The prediction of linear (n = 1) behavior holds so long as atomic flux divergence arises from fixed spatial variation of the material parameters appearing in Eq. (6). If the electron wind acts over a long time, of course, those material parameters may change as atomic material electromigrates. Nevertheless, scaling arguments demonstrate that if material parameters change linearly due to electromigration itself, a unit value of exponent n still results.

Another inconsistency concerns generally observed deviations from Black's law predictions for the MTTF under highfrequency ac and pulsed current stress: At high frequencies, the longer-than-expected lifetimes imply a short-term healing effect when current is off.

The above-described problems of Eq. (6) disappear if material parameters can be modified by some mechanism other than electromigration. Diffusion has been proposed as such a mechanism. A diffusant that affects the resistivity or diffusivity can lead to nonunit values of n. Furthermore, electric currents generate diffusant gradients that dissipate by diffusion when current is turned off, and this provides a mechanism for "healing" that can explain high-frequency lifetime effects under pulsed and ac current.

It is easy to understand that intentional copper impurity in aluminum conductors can serve as a diffusant species. Even within a single-component conductor, however, mobile vacancies can play the role of a diffusant. The mathematical treatment of diffusing, electromigrating species (atoms, ions, vacancies) described by a concentration C requires consideration of an additional term in Eq. (2). Adding the appropriate diffusion term to the nonvanishing term on the right-handside of Eq. (2), one obtains

$$\boldsymbol{J} = \frac{C\rho Z^* e D}{kT} \, \boldsymbol{j} - D \, \text{grad} \, C \tag{7}$$

In one dimension *x*, Eq. (7) has the one-parameter family of stationary (J = 0) solutions $C(x) = C(0) \exp(+\kappa x)$, where $\kappa = \rho Z^* ej/kT$. (If κ is not a constant, κx is replaced by the integral

of κ from the origin to x.) This stationary solution can be used to determine Z^* (31).

Shatzkes and Lloyd have used a model incorporating vacancy diffusion and void nucleation time to provide a theoretical explanation of a value of 2 for n in Black's equation (the value Black introduced on an empirical basis) (33). Although n = 1 and n = 2 have been demonstrated as consequences of specific mechanisms acting in isolation, the values of n found experimentally represent an interplay of mechanisms (8,34,35).

A factor that affects the measured value of n is the heat generated by current flowing in the metal, called Joule heating. This heat generation is quadratic in the electronic current density, and it increases the effective value of n. Any local concentration of current, by a bend or material inhomogeneity, can lead to local heating. This has a direct effect on Eq. (6), through T, and a typically larger indirect effect through the temperature dependence of the diffusivity D, which normally is an activated function of T. The mechanism for breakdown at a local "hot spot" is that the gradients associated with a local increase in diffusivity lead to (a) upwind depletion and voiding and (b) downwind accumulation or hillock growth. This has been proposed as the mechanism for the formation of narrow "slit voids" (36), which are able to grow across grain boundaries.

STATISTICAL CONSIDERATIONS

The lifetime, or more explicitly the time-to-failure (TTF), of a microelectronic device depends on microscopic details, such as crystalline grain structure, that cannot be controlled precisely in fabrication. This implies that the lifetimes of interconnects or devices are best described statistically. The lifetime distribution arises from a probability density p(t) that is the failure rate at time t.

The failure rate p(t) exhibits different regimes of behavior at different times. At short times, p(t) is large and decreases with time. This represents the "early failures" often associated with fabrication errors such as poor contacts (37). After this short burn-in period, there may be a latency period during which the failure rate is low. This period may be associated, for example, with the emptying of reservoirs designed to delay electromigration-induced failure, as illustrated in Fig. 2. Depending on a number of conditions, the failure rate may eventually enter a final period during which the failure rate again increases. When this occurs, the failure rate is described as having a wash-basin form.

Since real devices can fail by a number of different mechanisms, the full behavior of p(t) over all time cannot be modeled by a simple function for any even moderately complicated device. However, in simple interconnects, carefully fabricated to minimize extraneous early failures, failure occurs by a single electromigration mechanism or family of related mechanisms, and fits by two- and three-parameter formulas have been successful.

One very successful two-parameter fit, used for the overwhelming majority of statistical data, is the lognormal distribution. When the logarithm of a quantity is normally distributed, the quantity itself is said to be distributed according to the lognormal distribution. A lognormal distribution for the

lifetime *t* is thus given by

$$p(t) = \frac{1}{\sigma t \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} (\log t - \log t_{50})^2\right) \tag{8}$$

where t_{50} is the median time to failure (MTTF) and also the median time before failure (MTBF). The median for this distribution is equal to the *geometric* mean, but not to the *arithmetic* mean or average lifetime, which is $\langle t \rangle = t_{50}e^{\sigma^2/2}$. The parameter σ , the standard deviation of log t, is also called the shape factor or the deviation in the time to failure (DTTF). In contrast, the standard deviation in t is $t_{50} \sqrt{e^{2\sigma^2} - e^{\sigma^2}}$. For small values of σ , this approaches $t_{50} \sigma$, so σ itself approximates the coefficient of variation.

The integral of the failure rate p up to time t—that is, the fraction F(t) of the initial set that fails before time t—is given by

$$F(t) = \int_0^t p(t') dt' \tag{9}$$

and is called the cumulative failure probability or cumulative distribution function (cdf). For the particular case of a lognormal failure distribution, cumulative failure probability is given by

$$F(t) = \frac{1}{2}(1 + \operatorname{erf}(z))$$
(10)

where we have written z for $\log(t_{50}/t)/\sqrt{2}\sigma$, and $\operatorname{erf}(z)$ is the error function.

Statistical data from lifetime experiments is generally displayed on lognormal probability paper. This is paper ruled according to nonlinear scales in such a way that a curve of cumulative failure probability versus time will be a straight line if it arises from a lognormal distribution. The resulting graph, called a Weibull plot, provides a test for the validity of the distribution, as well as a straightforward means of estimating graphically the parameters t_{50} and σ . As shown in Fig. 3, the value of t_{50} can be read directly off the plot as the value of t at 50% cumulative failure, and σ can be determined as $0.5 \log(t_{84}/t_{16})$.

At small times t (large negative values of z), F in Eq. (10) approaches $\exp(-z^2)/2z\sqrt{\pi}$ asymptotically. Because z depends only logarithmically on t_{50} , but inversely on σ , early time failures may depend more sensitively on DTTF than on MTTF. For example, if design specifications concern average failure rates for the first year, then a change that doubles the MTTF from 10 years to 20 years will be erased by an increase of the DTTF by only 30%. This kind of compensation is an issue in bamboo-structured lines, which show increased t_{50} partly negated by increased σ .

In addition to the ordinary failure rate p and its integral, the cumulative failure probability F, it is useful to define a quantity called the intensity (38), the hazard rate, or the instantaneous failure rate (39,40) by

$$\mu(t) = \frac{p(t)}{1 - F(t)} \tag{11}$$

In fact, both p(t) and $\mu(t)$ measure an instantaneous failure rate at time *t*. The ordinary rate *p*, however, is referred to the whole population: p(t) dt gives the fraction of the *original* set

that fails during a time interval (t, t + dt). In contrast, $\mu(t)$ dt gives the fraction of those remaining at time t that fails during the same interval (t, t + dt). This difference makes μ insensitive to early-failure mechanisms: If, in some initial population, some fraction P are exposed to damage that causes early failures, then after these initial failures p(t) is reduced by a factor (1 - P)—not because long-time reliability is improved, but because fewer devices are left to fail. In contrast, the rate $\mu(t)$ is unaffected at long times, since this measures only the failure rate of the survivors. If μ approaches a constant after a certain time, then Mill's ratio, $\lambda = 1/\mu$, is the mean life expectancy of the remaining unfailed devices. Because of its insensitivity to early failure mechanisms, μ is the basis for determining asymptotic forms of failure distribution.

The primary justification for using lognormal lifetime distributions is that they fit measurements, but an interpretation can be made in terms of Black's equation [Eq. (1)]. On both physical and general statistical grounds (central limit theorem) it is reasonable to expect an at least approximately normal distribution of activation energies E_a . Since, according to Eq. (1), E_a varies logarithmically as the MTTF, this leads to a lognormal distribution of E_a , and the variance in TTF is often expressed as an equivalent variance in E_a . A normal distribution of temperature fluctuations in Black's equation also gives rise to a lognormal distribution of TTFs, so long as the coefficient of variation is small.

Lognormal distributions can also be justified in terms of microscopic correlates of electromigration failure. The size distributions of grain sizes are accurately lognormal, as determined both in painstakingly measured physical systems and in models of grains based on Voronoi and Voronoi-like cells constructed around random cell-center position distributions (41). Strong linear correlations have been found between (a) the median grain radius r_{50} and the lognormal shape parameter σ_r of metal grain size distributions and (b) the median σ_r and shape parameter σ of lifetime distributions for interconnects made from the same material (42). Stress voids, characterized by both void volume and penetration length of voids into lines, have also been found to be lognormally distributed (43).

An alternative approach to lifetime distributions starts from the assumption that a long wire consists of segments that are essentially independent. Since the wire fails as soon as any single segment fails, the distribution of failure times can be formulated as a problem in the distribution of extreme values: For a wire regarded as a chain of N identical independent segments, the wire's failure time is the minimum (extreme) value of N lifetimes taken from a common distribution for segment lifetimes.

This analysis in terms of extreme values immediately imposes general constraints on the allowable form of the lifetime distribution. In particular, it restricts the possible applicability of the lognormal distribution. If, for example, the interconnects of length L have lifetimes accurately described by a lognormal distribution, then under the assumption of independent segments, interconnects of length 2L cannot be fit by a lognormal distribution using any values of t_{50} and σ (10). Despite this shortcoming in principle, the lognormal distribution is widely used as a good approximation in statistical reliability analyses.

Just as the central limit theorem predicts the emergence of normal statistics under general conditions from sums of non-normally distributed variates, so there exist "asymptotic" theorems that predict the emergence of certain classes of extreme (i.e., whole-wire) distributions from a broad range of original (wire segment) lifetime distributions. An appropriate distribution for interconnect lifetimes is the double exponential distribution, one of a few distributions derived by Weibull. (Care should be taken not to confuse these Weibull distributions with the nonlinear Weibull plots, described earlier. Weibull plots can be defined for any distribution, including lognormal and Weibull distributions.) The cumulative failure distribution for this Weibull distribution is

$$F(t) = 1 - \exp[-(t/t_{37})^{1/g}]$$
(12)

where g is a shape parameter greater than unity, and t_{37} is the time when on average a fraction $1/e \approx 0.368$ of samples should have failed.

Comparative studies done to determine whether Eq. (12) or Eq. (10) is a better fit to observed data have found that for the parameter values and ranges of observed time that occur in practice, lognormal and Weibull distributions are very similar and hence achieve comparable accuracy in fits to data.

Lifetime distributions arising from electromigration failure are very broad, even for circuits or components with identical layout, fabricated simultaneously on the same wafer. The lifetime distributions are broad in two different ways: (1) Quantitatively, they are broad in the sense that the width of the lifetime distribution, as characterized by the standard deviation, is comparable to the mean lifetime. In terms of the lognormal distribution, this appears as DTTF values on the order of 0.5 or more. (2) Qualitatively, the distributions are broad in the sense that the long-time tail of the distribution falls off more slowly than any exponential. This leads to a Mill's ratio λ that always increases with time, implying that the *remaining* life expectancy of *surviving* devices increases with time.

Because of the long duration required even for accelerated life tests and because of the requirement of a temperaturestabilized environment and stable high current density, statistical studies have generally been limited to a few hundred samples and have found small (<1%) deviations from lognormal form (9,10). In one larger study of single-layer films (37), the failure distribution was found to be bimodal (see Ref. 39), corresponding to a sum of two accurately lognormal distributions. The early failures arose from the smaller component (integrated failure probability 3% to 5% of the whole) which had much larger shape parameter σ than the main population (3 versus 0.6).

A qualitatively different situation appears to hold for multilayer interconnects. A three-parameter fit that has been found (44) to describe failure in this case is an extension of the lognormal distribution, made by substituting

$$t/t_{50} \to (t - t_0)/(t_{50} - t_0) \tag{13}$$

in the usual expression Eq. (8) for the lognormal distribution. Here t_0 is the minimum time necessary for a failure to occur. This parameter, determined empirically, represents an incubation time associated with initial depletion of aluminum (45).

MANAGING ELECTROMIGRATION

Pure Al interconnects exhibit electromigration that is intolerably high for commercial ICs, so various schemes have been devised to increase the reliability of modern interconnects to acceptable levels. Electromigration can occur either within the runner, at an ohmic contact to the Si substrate, or at a via between multilevel interconnects. Each situation presents its own problems and requires unique solutions.

For runners, Al alloys with Cu are used, as discussed above, to slow the diffusion of Al along grain boundaries, leading to an overall diffusivity closer to that of bulk material and also leading to much higher lifetimes. Coupled with this, multilayer, or laminated, schemes are employed with layers in many combinations. These include cladding over- and underlayers, as well as a diffusion barrier between AlCu layers. These layers are formed from combinations of Ti, W, TiN, TiW, and other refractories. Despite their higher electrical resistivity, refractory metals and alloys are useful because of their much higher resistance to electromigration compared with Al alloys. Thus, if a void forms in the Al alloy, the break is shunted by a higher-resistance, but short, bridge [Fig. 5(a)] which adds a small amount of resistance to the total resistance of the line. In the event of many voids forming on the line, a large increase in resistance [i.e., greater than 20% (46)] may cause failure of the IC, but lifetimes are increased considerably. One variation of this is a sandwich of AlCu/(W or Ti)/AlCu, as shown in Fig. 5(b). The refractory layer separates the AlCu so that voids cannot propagate across both AlCu layers, and the resistance at a void on one-half of the



Figure 5. Barrier layers in the form of (a) overlayers and underlayers and (b) middle layers help to increase MTTF substantially. In (a), refractory metal layers (Ti, W, TiN, etc.) shunt a void with a small amount of electrical resistance added to the line. In (b), a refractory middle layer prevents void propagation across the entire thickness of the line. The probability of voids forming near each other is low, so the overall increase in resistance is kept to a minimum.

line is tolerably small. The probability of two separate voids forming contiguously is much smaller than that of a single void propagating across a line with no refractory core.

Low resistance, nonrectifying connection to Si contacts must be made for proper circuit operation. During annealing above 450°C or so, Si diffuses into the Al at grain boundaries, and Al diffuses back into the Si substrate to fill the voids, resulting in Al spikes (47). These spikes can be deep enough to cause significant leakage current or short circuits at PN junctions. To alleviate this problem, Si is added to the first level metal to saturate it and prevent further uptake of Si. One drawback is that Si can precipitate out of the supersaturated solution, leading to deleterious flux divergences and premature wearout of the lines. Also, Si can electromigrate into the Al. One way that this problem is addressed is to deposit a layer of Ti/TiN before the AlCuSi, where Ti acts as an adhesion layer and TiN acts as a barrier layer to interdiffusion. It should be noted that by taking into account materials, thicknesses, layer combinations, and so on, a great variety of metallization schemes is possible. Those described here are representative examples.

An additional and very serious electromigration problem is shared by ohmic contacts and plugs due to flux divergence at their interfaces to the metal runners, and they can be the source of significantly shorter lifetimes than those of the runners themselves. As discussed above, flux divergence is required for electromigration failure. If a large reservoir of metal is provided, then material can flow along a line with less detrimental effect. Vias must be formed during IC fabrication to link the various levels of interconnects. Ideally, they are filled with the same alloy as the runners, so that various levels serve as reservoirs, and no large flux divergences occur. Older via technology utilized sputtering of the same metal (Al alloys) as the runners to fill wide vias. Reliability at the vias tended to be limited by step coverage at the edges, which was a cause of flux divergence. However, flux divergence at the interface between levels was not a problem with this technology.

Modern ICs employ vias that are deep, to decrease interlevel capacitance, and narrow, so as to decrease the total area taken up by via structures. Unfortunately, it is difficult to deposit Al into deep, narrow via holes without forming voids, hence creating other reliability problems. Therefore, techniques such as chemical vapor deposition (CVD) are used to create plugs made from, say, Ti/W that link the various levels. In this case, the different Al alloy levels do not serve as reservoirs for each other, and flux divergences between the plugs and runners can lead to electromigration failure. The failure mechanism at vias (as well as at ohmic contacts) occurs in stages (45). The electron wind force causes Cu on the downwind side of the plug (or ohmic contact) to flow downwind through the runner. The plug acts as a blocking boundary, so no Cu can flow in to replace that which flows away. After an incubation period, Cu has depleted from a length equal to the Blech length in the material. Once this has occurred, Al is free to migrate downwind, forming voids at the interface between the plug and the runner, raising the total resistance to a failure level. Reservoirs placed near studs, such as those shown in Fig. 2, can replace voided material for a length of time, greatly increasing the overall reliability of the via structures.

Although the combination of AlCuSi alloys with refractory barrier layers and interlevel studs has been used extensively



Figure 6. The dual damascene process, typically performed with Cu metallization. Two separate lithographic and etch steps form a trench and via of differing widths, as shown. Metal is deposited by CVD, filling the trench and covering the surface with undesired material. A chemical mechanical polish step removes the metal down to the surface of the interlevel dielectric.

in the past, the use of pure Cu interconnects has emerged as a replacement technology. Six-level, all-Cu technologies, originally developed by both IBM (48) and Motorola (49), boast vastly improved reliability properties, as well as lower resistivity (1.7 $\mu\Omega$ -cm for bulk Cu versus about 3.1 $\mu\Omega$ -cm for AlCu). Typically, Cu interconnects are formed by the dualdamascene process (Fig. 6) in which trenches are formed in a suitable dielectric (usually SiO₂) using two levels of lithography and etching, forming both the lines and the vias, followed by (a) CVD of a diffusion barrier such as TiN, (b) CVD or plating of Cu completely filling the trenches, and (c) etching back by chemical-mechanical polishing to the surface of the dielectric. Besides increasing the speed by significantly lowering resistance-capacitance ("RC") delay in the lines, Cu is a much more reliable interconnect material. Electromigration in Cu interconnects is at least one (48) and possibly two or more (49) orders of magnitude smaller than that in Ti/AlCu/ Ti interconnects, and stress migration is virtually undetectable.

BIBLIOGRAPHY

- H. B. Huntington, Electromigration in metals, in A. S. Nowick and J. J. Burton (eds.), *Diffusion in Solids: Recent Developments*, New York: Academic Press, 1975, pp. 303–349.
- P. S. Ho and T. Kwok, Electromigration in metals, *Rep. Prog. Phys.*, **52**: 301–348, 1989.
- M. Gerardin, De l'action de la pile sur les sels de potasse et de soude et sur les alliages soumis à la fusion ignée, *Compt. Rend.*, 53: 727-730, 1860.
- 4. S. Wolf, Silicon Processing for the VLSI Era, 2—Process Integration, Sunset Beach, CA: Lattice Press, 1990.
- 5. Semiconductor Industry Association SIA, *The National Technology Roadmap for Semiconductors*, Austin, TX: Sematech, 1997.

- P. B. Ghate and J. C. Blair, Electromigration testing of Ti:W/Al and Ti:W/Al-Cu film conductors, *Thin Solid Films*, 55: 113–123, 1978.
- J. R. Black, Electromigration failure modes in aluminum metallization for semiconductor devices, *Proc. IEEE*, 57: 1587–1594, 1969.
- J. Tao et al., Modeling and characterization of electromigration failures under bidirectional current stress, *IEEE Trans. Electron Dev.*, 43: 800–808, 1996.
- J. R. Lloyd, On the log-normal distribution of electromigration lifetimes, J. Appl. Phys., 50: 5062–5064, 1979.
- J. M. Towner, Are electromigration failures lognormally distributed?, Proc. 28th IEEE Int. Rel. Phys. Symp., 1990, pp. 100-105.
- J. A. Schwarz, Distribution of activation energies for electromigration damage in thin-film aluminum interconnects, J. Appl. Phys., 61: 798-800, 1987.
- 12. A. D. Brailsford and H. B. Aaron, Growth of grain-boundary precipitates, J. Appl. Phys., 40: 1702–1710, 1969.
- T. S. Lundy and J. F. Murdock, Diffusion of Al²⁶ and Mn⁵⁴ in aluminum, J. Appl. Phys., 33: 1671–1673, 1962.
- F. d'Heurle and I. Ames, Electromigration in single-crystal aluminum lines, Appl. Phys. Lett., 16: 80-81, 1970.
- S. Vaidya et al., Linewidth dependence of electromigration in evaporated Al-0.5% Cu, Appl. Phys. Lett., 36: 464-466, 1980.
- R. Rosenberg and L. Berenbaum, Resistance monitoring and effects of nonadhesion during electromigration in aluminum films, *Appl. Phys. Lett.*, 12: 201–204, 1968.
- I. Ames, F. M. d'Heurle, and R. E. Horstmann, Reduction of electromigration in aluminum films by copper doping, *IBM J. Res. Dev.*, 14: 461–463, 1970.
- T. M. Shaw et al., The microstructural stability of Al(Cu) lines during electromigration, *Appl. Phys. Lett.*, 67: 2296–2298, 1995.
- B. N. Agarwala, G. Digiacomo, and R. R. Joseph, Electromigration damage in aluminum-copper films, *Thin Solid Films*, 34: 165, 1976.
- I. A. Blech and E. S. Meiran, Electromigration in thin aluminum films on titanium nitride, J. Appl. Phys., 47: 1203-1208, 1976.
- J. T. Yue et al., Stress induced voids in aluminum interconnects during IC processing, *Proc. 23rd IEEE Int. Rel. Phys. Symp.*, Orlando, FL, 1985, pp. 126–137.
- S. M. Hu, Stress-driven void growth in narrow interconnection lines, Appl. Phys. Lett., 59: 2685–2687, 1991.
- 23. F. G. Yost and F. E. Campbell, Stress-voiding of narrow conductor lines, *IEEE Circuits Dev. Mag.*, 40-44, May, 1990.
- 24. J. Klema, R. Pyle, and E. Domangue, Reliability implications of nitrogen contamination during deposition of sputtered aluminum/silicon metal films, *Proc. 22nd IEEE Int. Rel. Phys. Symp.*, Las Vegas, NV, 1984, pp. 1–5.
- J. Curry et al., New failure mechanisms in sputtered aluminum– silicon films, Proc. 22nd IEEE Int. Rel. Phys. Symp., Las Vegas, NV, 1984, pp. 6–8.
- I.-S. Yeo et al., Characteristics of thermal stresses in Al(Cu) fine lines. II. Passivated line structures, J. Appl. Phys., 78: 953–961, 1995.
- J. R. Lloyd and J. J. Clement, Electromigration damage due to copper depletion in Al/Cu alloy conductors, *Appl. Phys. Lett.*, 69: 2486-2488, 1996.
- M. A. Korhonen et al., Stress evolution due to electromigration in confined metal lines, J. Appl. Phys., 73: 3790-3799, 1993.
- L. Kisselgof et al., Thermally induced stresses and electromigration failure, Proc. SPIE, Submicrometer Metallization: The Challenges, Opportunities, and Limitations, 1805: 1992, 154–163.
- 30. W. Seith, Diffusion in Metallen, Berlin: Springer-Verlag, 1955.
- F. Skaupy, Die Elektrizitätsleitung in Metallen, Verh. Deutsch. Physik. Ges., 16: 156-167, 1914.

- 32. H. Wever and W. Seith, Neue Ergebnisse bei der Electrolyse fester metallischer Phasen, Z. Electrochem., **59**: 942–946, 1955.
- 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., 59: 3890–3893, 1986.
- K. Hinode et al., Dependence of electromigration lifetime on the square of current density, *Proc. 31st IEEE Int. Rel. Physics Symp.*, Atlanta, GA, 1993, pp. 317–326.
- A. S. Oates, Current density dependence of electromigration failure of submicron width, multilayer Al alloy conductors, *Appl. Phys. Lett.*, 66: 1475-1477, 1995.
- J. E. Sanchez et al., Slit morphology of electromigration-induced open-circuit failures in fine-line conductors, J. Appl. Phys., 72: 3201–3203, 1990.
- H. H. Hoang et al., Electromigration early-failure distribution, J. Appl. Phys., 65: 1044–1047, 1988.
- E. J. Gumbel, *Statistics of Extremes*, New York: Columbia University Press, 20, 1958.
- 39. S. K. Kurtz, S. Levinson, and D. Shi, Infant mortality, freaks, and wear-out: Application of modern semiconductor reliability methods to ceramic multilayer capacitors, J. Amer. Ceram. Soc., 72: 2223–2233, 1989.
- 40. D. Kececioglu, *Reliability Engineering Handbook*, Englewood Cliffs, NJ: Prentice-Hall, 1991.
- S. Kumar et al., Properties of a three-dimensional Poisson– Voronoi tessellation, a Monte Carlo study, J. Stat. Phys., 67: 523, 1992.
- W. Baerg and K. Wu, Using metal grain size distributions to predict electromigration performance, *Solid State Technol.*, **34** (3): 35–37, 1991.
- S. Kordic, J. Engel, and R. A. M. Wolters, Size and volume distributions of thermally induced stress voids in AlCu, *Appl. Phys. Lett.*, 68: 1060–1062, 1996.
- R. G. Filippi, G. A. Biery, and R. A. Wachnik, Paradoxical predictions and a minimum failure time in electromigration, *Appl. Phys. Lett.*, **66**: 1897–1899, 1995.
- C.-K. Hu, M. B. Small, and P. S. Ho, Electromigration in Al(Cu) two-level structures: Effect of Cu and kinetics of damage formation, J. Appl. Phys., 74: 969–978, 1993.
- C. A. Martin et al., Electromigration performance of CVD-W/Alalloy multilayered metallization, *Proc. 28th IEEE Int. Rel. Physics* Symp., New Orleans, LA, 1990, p. 31.
- S. K. Ghandi, VLSI Fabrication Principles, New York: Wiley, 1983.
- D. Edelstein et al., Full copper wiring in a sub-0.25 μm CMOS technology with copper metallization, *Proc. IEEE IEDM*, Washington, DC, 1997, pp. 773-776.
- S. Venkatesan et al., Performance 1.8V, 0.20 μm CMOS technology with copper metallization, *Proc. IEEE IEDM*, Washington, DC, 1997, pp. 769-772.

GARY H. BERNSTEIN ALFRED M. KRIMAN University of Notre Dame

ELECTROMYOGRAPHY 523