PHYSICAL BOUNDARIES OF PERFORMANCE: THE INTERCONNECTION PERSPECTIVE

S. K. Tewksbury
Dept. of Electrical and Computer Engineering
West Virginia University
Morgantown, WV 26506

Abstract

Several interconnection issues relating to faults and reliability are reviewed here. Whereas the occurrence of opens in interconnections or shorts between interconnections is well understood within conventional models of digital systems, the faults originating from the analog characteristics of signals propagating across interconnection lines (particularly long lines) is less often discussed. However, such functional faults are likely to become increasingly important, not only due to the higher frequency operation of VLSI circuits but also due to the development of advanced packaging schemes using thin film technologies and multi-chip modules (MCMs). Such MCMs are characterized by line lengths much longer than typically encountered within VLSI circuits. The “digital” signal being transmitted across a long VLSI or MCM interconnection line is represented here as an “analog” signal which must be restored to a legitimate digital signal level at the specified times imposed by flip-flops. Incorrect restoration of the “digital” signal at the far end is treated as a fault.

Introduction

This paper, parts of which are adapted from [1], considers a variety of interconnection characteristics which can induce faults or failures in digital integrated circuits. This emphasis on interconnects contrasts with the traditional emphasis on faults and failures induced by active devices but mirrors a general increasing focus on system performance limits imposed by interconnects.

The term “faulty interconnect” is often taken to mean an open in an interconnect or a short between adjacent interconnections. However, longer interconnects can significantly affect the signal waveforms received by a digital circuit function, as illustrated in Figure 1b. In such cases, the interconnection is more correctly viewed as an analog circuit function (or, in the case of a set of interconnection nets, an analog net function) than as a passive channel passing “digital” signals undisturbed (aside from a possible delay) as in Figure 1a. The ideal “digital” signal appears at the input to the line driver and may be distorted when viewed at the output of the driver (e.g. due to signal reflections from an unterminated far end). At the far end of these longer interconnections, the active digital circuitry receiving the analog signal from the interconnection must restore the
received signal to an unambiguously "digital" signal. Failures in achieving a properly restored signal at the expected time within the digital logic immediately following (and perhaps including) the far end receiver then constitutes an important, potential interconnection fault class. The resulting faults can be permanent, randomly occurring, or perhaps dependent on the data on the given line and adjacent interconnection lines. This broader class of interconnection faults is considered here, with the emphasis accordingly on those interconnections for which the analog nature of signal transmission is a significant issue. The relevant interconnections are physically longer than those generally encountered internal to highly localized digital functions included within the overall IC circuit function, the length necessary for analog behavior to be important often decreasing as the speed (e.g. clock rate) of the system increases.

![Figure 1: Electrical interconnections as analog circuits. (a) Short interconnection within logic function acting as "digital" line. (b) Long interconnection in IC or MCM acting as an "analog" line.](image)

**Interconnection: An Analog Circuit**

**Generic Waveform Degradations**

The full specification of possible degradations is difficult since the specific degradations will depend on the nature of defects, on the signal data rate (and expected bandwidth), on the driver and receiver circuitry, on the presence or absence of terminations, on the nature of the net itself when a single driver output fans out to several receivers (perhaps passing through various vias and layers of interconnection along the path to a specific receiver).
Figure 2 shows examples of various waveform degradations (signal delay, decreased rise-times, ringing, noise, and reflections) which may be induced by interconnections on a nominal "digital" signal from the line driver. These degradations are largely negligible for short interconnections and low frequencies but become more severe as the line length or signal bandwidth increases. Such waveform degradations may lead to a variety of faults in the operation of the overall digital circuit function. Waveform degradations do not themselves indicate faulty circuit operation since a properly designed circuit will function correctly due to correct restoration of the digital signals. However, variations in the details of the specific signal being restored (either relative to other signals arriving on the same line or relative to the average shape of signals within an IC product) can be a source of faults.

Simultaneous Switching Noise

Schaper [2] and others [3,4,5] have emphasized $\Delta I$ noise (also called simultaneously switching noise), an increasingly important problem. Figure 3 illustrates the basic constraint. $L_p$ is the parasitic inductance of the power (or ground) connection between the pc-board and the IC chip's power (ground) pads. Consider an $N_b$-bit parallel data output port, with a capacitive load $C_b$ per output line. To charge that capacitance in a time $t_n$ from voltage $V = 0$ to $V = V_b$ requires that the current be switched from 0 to $I_b = 0.5 N_b C_b V_b / t_n$. Since the output drive current (for this 0 to 1 output signal transition) is provided through the power connection between the circuit board and the chip, a voltage transient $\Delta V$ appears on the power voltage line, where

$$\Delta V \approx L_p \times \frac{dI_b}{dt} = \frac{N_b I_b C_b V_b}{2 t_n^2}.$$

For $N_b = 20$, $C_b = 1 \mu F$, and 1 nsec rise-times, the power (and ground) lead inductance between circuit board and the IC pads must be less than about 10 nH to obtain $\Delta V/V \leq 10\%$. Canright [6], using as an example a package with 8 power pins, 8 ground pins, 81 output signal pins and 151 input signal pins, gives an inductance of 22 nH and resistance of about 1$\Omega$ for an IC I/O pin.
To avoid the accumulation of effects of risetime degradations and other signal waveform degradations, the signals in a digital system can be systematically restored to digital signals using flip-flops, establishing a time at which the restored signal can be unambiguously interpreted as a "digital" signal. Regarding the flip-flop input signal as an analog signal, the flip-flop serves the function of a sample-and-hold circuit. A variety of clocking schemes can be used to pass digital signals through a pipeline of retiming flip-flops. The example below assumes a common clocking time for each flip-flop.

Figure 4 illustrates such reclocking of the analog signal at the far end of a long interconnection. The first clock pulse generates the near end signal while the second clock pulse "times" the restoration of the digital signal at the far end. In Figure 4a, the clock period $T_{\text{clock}}$ considerably exceeds the sum of the signal propagation delay $T_{\text{delay}}$ and the far end signal's settling time $T_{\text{settle}}$. This rather conservative timing would lead to a system operating well below its maximum clock rate.

In Figure 4b, the clock rate has been increased (while the signal propagating across the interconnection is unchanged) with the result that the margin of time following the settling of the far end signal has decreased. By operating near the region of significant ringing in the transient response, the circuit has been effectively sensitized to variations in signal delay, in the clock pulse timing, etc. In Figure 4c, the clock rate has been increased further, with the signal being "sampled" by the flip-flop during its transient period. Although the restoration of the digital signal may be correct, the operation has sensitized the circuit to detailed changes in the transient waveform. As data rates
achievable within localized digital circuit functions increase, the sensitivity of the circuit performance to faults due to the analog characteristics of the far end signal on longer interconnections increases.

Clock Skew

Figure 4 shows a sharp and precisely timed clock pulse. However, maintaining a sharply defined and clean clock pulse across the long length of the clock network is difficult in high speed circuits, just as preserving the shape of data signals on such long lines would be difficult as discussed above. The widely discussed problem of clock skew is considered here.

The clock skew can be defined generally as the deviation of the clock transition time at a local register (or flip-flop) from its desired transition time. If the desired clocking instants may vary throughout the circuit, then the clock skew is the local deviation of the clock transitions from these desired local desired clocking instants. Alternatively, if the desired clocking instant is the same throughout the circuit, then the clock skew, though still defined locally, is the local deviation of the clock transitions relative to a global clocking instant. Deviations from the desired clock timing may be constant offsets in time and/or may be variations in the clocking edge among successive clocking pulses. Constant offsets can arise from several factors [7] including (i) differences in line length from clock source to the clocked circuit, (ii) differences in delays through any active devices (e.g. buffers) in the clock network, (iii) differences in line parameters, including resistivity, dielectric constant, via resistances and including variations in line capacitance due to the generally random surface topography over which the clock distribution lines are run, and (iv) differences in threshold voltages in different clocked circuits. Random
variations in clocking edges can result from the analog propagation of clock signals on the network (combined with crosstalk and other effects which impact the detailed clock waveform of successive clock pulses).

Line resistance and capacitance obviously impact clock distribution. However, the net input capacitance of the several clock input transistors of a data register may exceed the capacitance of the clock line [8]. This suggests placement of a clock driver close to the register, the driver outputs seeing a small resistance when driving the substantial capacitance of the registers. As circuits become more irregular, fabrication process variations contribute to variations in the clock timing. Shoji [9] has recommended some convenient rules such that although process variations may change the capacitance along a given path, the relative values of line capacitances on different branches of the clock network remain unchanged.

Data signals, even though originating at a common point and time, may lead to digital signals whose timing varies at various successive points within a combinational logic circuit. Management of clock distribution requires not only control of the clock phase but also an understanding of distances over which variations in data timing can be ignored. Anceau [10] introduced the concept of isochronic region, defined as regions within which interconnection delays could be ignored (e.g. the skew are less than 10% of the clock period). A similar concept (i.e. equipotential region) was suggested in [11]. These regions represent distances over which the phase variations of clock pulses due to interconnect delays are largely negligible. Over longer distances, one can consider whether synchronous clocking approaches need be maintained. For example, an IC including a core microprocessor and several asynchronous peripheral circuit functions could be operated with independent clocks within the various functions. Just as one can define distances beyond which phase shifts are significant and need to be carefully controlled, one can also consider distances beyond which it may (depending on the digital function) be preferable to use asynchronous communications, avoiding the need for careful control of clock skew throughout a complex system. This issue has been addressed recently by Messerschmitt [12].

Metastability

In 1973, Chaney and Molnar [13] suggested that the metastable point in the characteristic of a bistable device would create errors in clocked bistable devices receiving asynchronous inputs. Several studies [14,15,16,17,18,19,20,21] have since established that error-free input of data into a synchronous system from a non-synchronized source is not possible, even though all circuitry may be free of faults.

Figure 5a illustrates the potential energy $V(x)$ vs distance $x$ for a bistable physical system. This familiar physical model has a close analogy to the electrical characteristics of a bistable device. The force $F$ on a particle moving in this potential is $F = -\partial V / \partial x$. At the points $x_0$, $x_1$ and $x_m$ in Figure 5a, $\partial V / \partial x = 0$ and a particle initially at rest at one of those points remains at rest. However, unlike the stable points $x_0$ and $x_1$ from which small displacements create restoring forces pulling the particle back to its original stable point, small displacements away from $x_m$ create forces which push the particle further away from $x_m$. The lack of restoring forces motivates the terminology metastable point $x_m$ in contrast to the stable points $x_0$ and $x_1$. Since the force driving the particle...
away from the metastable point toward a stable point is proportional to \( \partial V(x)/\partial x \), which becomes vanishingly small as the displacement from \( x_m \) decreases, there can be considerable delays for small initial displacements \( \epsilon = x - x_m \). Furthermore, if some random noise is added to the system, the particle will, through random displacements, eventually move to a stable point, regardless of a particle’s initial state.

In the case of a real electronic latch or flip-flop, the stable points correspond to the normal digital logic voltage levels of the circuit, i.e. \( V(0) \Rightarrow x_0 \) and \( V(1) \Rightarrow x_1 \). The metastable point corresponds to some intermediate voltage level, here represented as \( V_m \Rightarrow x_m \). For voltage levels \( V(0) \) and \( V(1) \), the circuitry is operating near saturation. However, near \( V_m \), the circuitry is operating in the linear range.

Figure 5b shows the metastable response of a flip-flop’s output signal. During the flip-
flop's aperture time, the flip-flop's clock "samples" and stores the input signal level at an 
internal node of the flip-flop. If the stored signal level is at a valid digital logic level, the 
flip-flop's output relaxes quickly to a correct digital output voltage level. However, if the 
stored voltage level is near the metastable voltage level, then a voltage level intermediate 
between valid logic "1" and "0" levels appears at the output, persisting for some period 
of time, until the gain and feedback of the flip-flop drives the output from the metastable 
level to a valid logic level.

Models for metastability-induced synchronization faults have been reviewed by Kleeman 
and Cantoni [14]. The transient response of a flip-flop in the metastable state can be 
evaluated using the first order flip-flop model (i.e. a linear model of the flip-flop in the 
neighborhood of metastability) of Veendrick [21]. With such simplified models of the 
circuit behavior, the probability that the transition time to a valid logic level exceeds 
allowed time margins can be estimated, providing the expected error rate. The digital 
logic which follows the flip-flop imposes an interpretation of even an invalid voltage level 
as a logical "1" or "0" (it knows no other interpretations), the fault being generated 
by this "interpretation". In fact, the general belief that metastability failures can not 
be avoided (by detection and correction with binary logic circuits) follows from the 
propagation of improper voltage levels through subsequent combinational logic circuitry, 
corrupting any decisions that might be made.

Interconnection Classification

The extent to which a line exhibits various types of signal degradations such as shown in 
Figure 2 depends on several factors including signal bandwidth, interconnection length, 
electrical properties of the interconnection etc. The different generic classes of thin film 
interconnections are briefly reviewed here.

Interconnection lines considered here are generally a conducting strip separated by a 
dielectric layer from a ground plane (though more complex structures appear in multi-
layer metallization schemes, a complication not specifically considered here but clearly 
important). Figure 6a shows the simple model of the distributed LRC transmission line 
used for lower frequencies. \( R_l \) is the series resistance per unit length, \( L_L \) is the series 
inductance per unit length, \( C_I \) is the shunt capacitance per unit length and \( G_I \) is the 
leakage conductance per unit length (here assumed negligible). This "circuit" model 
directly emphasizes the analog nature of signal propagation on interconnections. The 
current and voltage equations, for a Fourier component of the signal, are

\[
\frac{dI}{dx} = -Y^*V, \quad \frac{dV}{dx} = -Z^*I
\]

with \( Y^* = jwC_I \) the shunt admittance of the line and \( Z^* = R_l + jwL_l \) the series impedance 
of the line. Another important parameter associated with the line is the impedance 
\( Z = \sqrt{L_L/C_I} \) seen by the driver. Solutions of (1) have the general form \( \exp(\pm ax) \) with 
a, the propagation constant given by

\[
a = (Z^*Y^*)^{1/2} = \alpha + j\beta.
\]

Here, \( \alpha \) is the attenuation factor and \( \beta \) is related to the wavelength \( \lambda \) associated with
signal propagation along the line according to
\[ \lambda = \frac{2\pi}{\beta} \]  
(3)

The signal propagates along the line with a phase velocity \( v_p = \omega/\beta \) and energy propagates at the group velocity \( v_g = \partial \omega/\partial \beta \). If \( \omega/\beta \) is a constant, then \( v_p = v_g \) (and both are independent of frequency) and the signal propagates without dispersion. If \( v_p \) depends appreciably on signal frequency \( \omega \), the line is dispersive and different frequency components exhibit different propagation velocities.

**Figure 6:** Common electrical interconnection models.

Keyes [22] has provided a convenient summary of the conditions under which an interconnection line can be represented by various simple models. Three parameters, with dimensions of length, are used. These are (1) \( L \) the line's characteristic wavelength defined above and (3) \( Y \), the ratio of the characteristic impedance \( Z \) to the line's resistance per unit length, i.e.

\[ Y = \frac{Z}{R_0} \]  
(4)

Qualitatively, \( Y \) is the line length at which its series resistance equals the line's characteristic impedance. The following general rules were suggested in [22].

**Short Lines**

If the line length greatly exceeds the wavelength \( L \ll \lambda \), the line is electrically short [23,24], i.e. not long enough for the transmission line properties to be established, and can be modeled using lumped circuit models such as the T-network shown in Figure 6b [24]. The overall delay is

\[ t_d = \left\{ 2\pi(t^2_i + t^2_e) \right\}^{1/2} \]
where $t_r = r_d(C_1L + c_i) + 5R_1C_1L^2$ is the risetime and $t_p = L/\sqrt{L C_1} = L/\sqrt{\pi f}$ is the propagation delay. Here, $r_d$ is the driver resistance and $c_i$ is the input capacitance of the MOS receiver. For very low resistance lines (e.g. the metal interconnections considered here) $t_r = r_d(C_1L + c_i)$. Typically, $t_p \ll t_r$ giving $t_d \approx 2.5 t_r$. Since $r_d$ is related to the driver current and the voltage drop across the driver (i.e. $r_d \approx V_d/I_d$), the rise time is approximately $t_r \approx (C_1L + c_i) V_d/I_d$. The principal signal degradation here is the slower risetime.

**Transmission Lines**

If the line length is much longer than the wavelength ($L \gg \lambda$), then electromagnetic effects can be expected to be important. In fact, if $L \gg \lambda$ and $Y > \lambda/2\pi$, the characteristic line impedance dominates the line's series resistance and the electromagnetic effects dominate. Under these conditions, the transmission line models are useful. The line's response is dominated by its inductance and capacitance, as suggested in Figure 6c. With $R_0 \ll \omega L_0$,

$$\beta = \sqrt{\omega^2 L_0 C_1}$$

and

$$v_p = \frac{1}{\sqrt{\epsilon_0 C_1}}$$

(5)

For frequencies $\omega$ low enough that the line inductance and capacitance are frequency independent, the line is non-dispersive and approximates a standard idealized transmission line. For an ideal transmission line, a pulse is propagated unchanged aside from a delay $r_d = L/v_p$. At very high frequencies, $L_0$ and $C_1$ become frequency dependent, the line becomes dispersive, and the ideal transmission line is a poor approximation. Different Fourier components of the digital pulse signal propagate with different frequencies and attenuations, with the result that the far end signal reconstructed from the propagated Fourier components differs from the initial pulse. The resulting signal degradations increase as the line length increases. In a real system, the interconnects are not simple microstrip transmission lines but instead generally are complex interconnect nets with imperfect terminations, varying structures across surface topography, and close placement relative to other interconnection lines. High frequency modeling of microstrip transmission lines under such "real-world" conditions is a complex topic [25,26].

**Distributed RC Lines**

As above, if the line length is long compared to the wavelength ($L \gg \lambda$), electromagnetic effects may dominate. However, if $L \gg \lambda$ and $Y < \lambda/2\pi$, the series resistance of the line dominates its characteristic impedance and a distributed RC model such as shown in Figure 6d is appropriate. For such distributed RC lines,

$$\beta = (\omega R_0 C_1)^{1/2}$$

and

$$v_p = \frac{\omega}{\beta} = \frac{\omega}{\sqrt{\epsilon_0 C_1}}$$

(6)

The phase velocity (and attenuation) depends on frequency (i.e. the line is dispersive). Because of the waveform degradations, the line delay $\tau$ cannot be obtained simply as $\tau = L/v_p$. There has been considerable study of such resistive lines over the past few years and several useful engineering approximations have been obtained. Sakurai [27] obtains, for the total delay (including both the signal rise time and the propagation delay),

$$t_d = 1.09 R_0 C_1 L^2 + 2.25 r_d + 2.25 R_0 C_1 L.$$  

(7)
The intrinsic delay of a MOSFET is contained in the second term (i.e., \( r_dC_l \)). If the line resistance \( R_L \) is small relative to the driver resistance, then the primary impact of the line is to add the \( r_dC_l \) delay per unit length associated with the line's capacitance. However, for sufficiently long lines, the line resistance \( R_L \) will become the dominant resistance term and the intrinsic line delay will dominate. In this latter case the delay increases quadratically with line length.

Lossy transmission lines (i.e., transmission lines with significant resistance per unit length) can have advantages over perfect transmission lines. In particular, unattenuated reflections in perfect transmission lines causing pulse distortions must be absorbed in line terminations, introducing significant static power dissipation.

**Multi-Chip Module Interconnections**

Multi-chip modules are rapidly being developed to provide high performance interconnections for fast circuit technologies. A convenient rule of thumb for metal line cross sections is obtained by considering the dependence of the high frequency signal attenuation on line resistance \( R \). In particular [28], the high frequency content of a traveling step wave discontinuity at distance \( l \) along a slightly lossy line is attenuated, relative to low frequency components, by a factor

\[
\mathcal{A} = \exp\left(\frac{-R_l \cdot l}{2Z_0}\right). \tag{8}
\]

Here, \( R_l \) is the line resistance per unit length and \( Z_0 \) is the characteristic impedance of the transmission line. Attenuation of high frequency components of a step function leads to an increasing risetime as the line length increases. To avoid degraded risetimes, the line resistance \( R_l \) must be sufficiently small that \( R_l \ll 2Z_0 \).

![Figure 7: Interconnection delay/risetime vs (a) line resistance and (b) length [28].](image)

The effect of metal line resistance on signal risetime is illustrated in Figure 7, adapted from [28]. Figure 7a shows the signal (initially a step signal) after traveling a distance 3cm along lines of varying resistance per unit length. The signal exhibits a delay of about 0.4 nsec, corresponding to electromagnetic propagation delay (speed of light divided by \( \sqrt{\epsilon_0} \)). Following this initial delay, the signal displays an abrupt initial step followed by an \( RC \) charging delay, due to line resistance and capacitance, to the final signal level.
For the lower resistance line (4Ω/cm), Figure 7b shows the signal at various distances along the line. The degradation in signal performance at increased distance is obvious, with lower resistance lines needed for the longer line lengths.

Microwave Performance Issues

Kwon and Pease [29,30] have reported numerical simulations of the performance of multi-level, thin film interconnections assuming a multi-level metalization geometry having the following characteristics. (i) The first signal level is a set of north-south, gold interconnection lines (10μm wide 5μm thick) with a 35μm line pitch. (ii) The second signal level is a set of east-west, gold interconnection lines, with 10μm line width and 10μm line thickness. Shielded ground lines run between lines on this second layer. (iii) First and second metal layers are separated by a 10μm thick dielectric, a 5μm power plane, and a second 10μm dielectric. Stubs extend from this power plane between first layer metal lines, providing some shielding against crosstalk for first level signal lines. This geometry is chosen to support very high data rates across interconnections.

The numerical analysis in [29] for a 10 cm line using this multi-level structure suggests a signal attenuation by a factor less than 2, low delay time (0.65 nsec), and crosstalk less than 10%.

Impedance discontinuities at bends in lines, at vias and at fanout points will impact the signal waveform. A numerical analysis is used in [30] to evaluate the impedances associated with such discontinuities. Figure 8a illustrates a right-angled bend in the upper signal layer of the multilayer metalization model. Also included in Figure 8a are the equivalent circuit and the circuit parameters (capacitance, inductance and resistance) of the bend. Reflections at the bend are found to be negligible for frequencies extending beyond 100 GHz. Figure 8b shows similar characteristics of a via between first and second level signal metalization. The 10μm x 10μm cross section via is estimated to have a resistance of about 15mΩ. Reflections of about 20% at frequencies below 20 MHz, decreasing to less than 6% above 1 GHz were found, though the reflection was attributed to the impedance mismatch between first and second level metal lines, rather than to the via itself.

Physical Defects

The summary below of physical defects provides several examples commonly used to illustrate the familiar open and short defects in interconnect nets. However, several defects have the potential of substantially changing the electrical characteristics of interconnections, particularly for very high frequency components of signals. The resulting variations in the electrical properties of the interconnections can extend into significant variations in the signal waveform at the far end of a long interconnection. In this sense, the physical defects can give rise not only to the classical open and short defects but also to the faults caused by incorrect signal waveforms. Planar defects within a metal interconnection layer are similar to the defects seen on photolithography masks [31] and are illustrated in Figure 9a. Figure 9b shows two examples of vertical defects within a given metalization layer. Connections (through vias) between layers of interconnection present significant problems as linewidths decrease without corresponding decreases in
the dielectric thickness [32,33]. Two examples of via defect mechanisms are shown in Figure 9c.

A pinspot (Figure 9a) is a spot of metal appearing in a region otherwise devoid of metal. Aside from changing the topography of overlying layers, such isolated pinspots are not serious defects unless filling a significant portion of the gap between closely spaced metal lines or unless having a height significantly greater than the normal metal interconnection.

Several defects provide sites for electromigration failure and may introduce significant changes in the electrical characteristics of an interconnection, even though the line may appear functional during DC testing. These include the following.

- A protrusion (Figure 9a) is additional metal extending from the edge of a metal interconnection.
An intrusion (Figure 9a) is missing metal from the edge of a line. It can be prone to electromigration failure, not only because of the geometric constriction but also because of the higher current density flowing through the remaining metal. Higher currents enhance electromigration failure rates and can also lead to catastrophic failure (e.g. "burnout") of the line early in service. Intrusions can also develop during service, again, for example, due to electromigration.

A pinhole (Figure 9a) is a region of missing metalization within the metal interconnection. Again, electromigration failure can be enhanced, both due to the geometric constraints on either side of the pinhole and due to the higher current densities in the remaining metalization.

Step coverage problems (Figure 9b) are severe, given the vertical sidewalls produced by dry etching procedures and the increasing aspect ratio of features in smaller geometry device technologies. A wide variety of techniques for planarization of surfaces or for sloped sidewall metalization have been developed to reduce step coverage problems. Surface topography becomes a more serious problem as additional metal layers are formed, due to the sidewalls of metal interconnections and due to vias between metal levels. Planarization is generally not perfect, being limited by surface tension effects [34]. Achieving planar surfaces is important not only to avoid step coverage problems but also to allow high resolution photolithography, where depth of field is very shallow, as discussed earlier.

Resistive contacts (Figure 9c) between metalization layers can greatly impact the line resistance seen by a line driver. Insulating films on metal surfaces (e.g. the native oxide film on aluminum) must be eliminated before a subsequent metal is overlaid to ensure a good ohmic contact between successively deposited/grown metal.

Figure 9: (a) In-plane interconnection defects (based on mask defect classification [31]). (b) Vertical metal defects. (c) Via defects.

Other defects generally lead to a classical open or short fault, as in the case of the following.
- A break (Figure 9a) in the interconnection causes an open line and failure of the interconnection.

- A bridge (Figure 9a) is unwanted metalization connecting adjacent interconnection lines. The shorting of two interconnection lines can have several fault effects on the circuit.

- A missing feature is a gross defect, corresponding to unwanted removal of significant amounts of metal, effectively causing the interconnection to disappear.

- Hillocks (Figure 9b) arise from the difference between the thermal expansion coefficient of aluminum and of the dielectric and can produce shorts between adjacent metalization layers.

- An open via (Figure 9c) introduces opens on connections between metalization layers. As the aspect ratio (depth-to-width) of the via increases, metal can cover the top of the opening (during deposition) before sufficient metal has been deposited into the via to provide a reliable, low resistance contact. The problem of filling steep sidewall vias due to increasing aspect ratio is worsened by the surface topography associated with multi-layer device structures. Variations in via depth through the overlaying planarizing dielectric to first level metal range from 300 nm to 1300 nm in the example in [32].

Not shown in Figure 9 are defects originating in the dielectric layers. SiO₂, Si₃N₄ and polyimide are the three principal dielectrics used as insulating layers. Defects in dielectric films include physical faults such as pinholes, voids, cracks "rocks" and poor step coverage. Such physical defects can lead to shorts between interconnection levels, faults in overlying films and degraded breakdown performance. Chemical impurities can also impact device performance, causing changes in device threshold voltages, influencing the breakdown failure mechanisms, etc.

The increasing complexity of metalization for interconnections in VLSI circuits suggests that the yield/reliability of interconnections vs the yield/reliability of devices must be reassessed as the technologies evolve. Compared to earlier technologies, where device yield/failure may have been significantly more important than interconnections, future technologies may result in interconnections which are a major yield loss and reliability issue for advanced ICs and for MCMs.

**Failure due to Metal Electromigration**

Electromigration is the macroscopic (and normally slow) flow of metal or semiconductor ions. The ions are "pushed" by collisions with electrons during current flow (i.e. the "electron wind") towards the positively biased end of the interconnection line with an ion velocity \( v_j \), \( j \) the local current density (Amps/cm²). Electromigration in typical IC aluminum metalization is dominated by the electron wind mechanism. This gradual flow of ions towards the positively biased end leads to depletion (voids) or buildup (widenings and hillocks) at discontinuities in the mass flow. After some period of operation, electromigration leads to failures of the interconnection.
Schreiber [35] has described the principal electromigration mechanisms in aluminum interconnections. The ionic current \( J_1 \) is related to the ion velocity \( v_1 \) by

\[
J_1 = N_1 v_1
\]  

(9)

with \( N_1 \) the ion density. The thermally activated ion velocity is described by [35,36]

\[
v_1 = \left( j - j_{th} \right) \frac{p e Z^*}{k T} D_0 \exp \left( -E_a / k T \right)
\]  

(10)

for \( j > j_{th} \). The ion current vanishes, in this model, for currents below the threshold current \( j < j_{th} \). In (10), \( p \) is the metal resistivity, \( eZ^* \) is the effective ion charge, \( j_{th} \) is the threshold current and the diffusion constant is \( D_0 \exp \left( -E_a / k T \right) \). The threshold current is inversely proportional to the line length, as discussed in [35,39,40,41]. This inverse relationship between \( j_{th} \) and \( l_{line} \) is a result of a second mechanism for atomic flux, i.e. "backflow" diffusion to reduce gradients. This suggests that the threshold current \( j_{th} \) is important for short lines but becomes a negligible correction for long lines.

The mean time to failure (MTTF) has been discussed extensively [42,43,44,45,46,47]. There are significant variations in measured values for electromigration MTBF parameters [47,42,48,49,50]. In general, gold and tungsten interconnections are regarded as having far superior resistance to electromigration failure than aluminum.

Yield Simulation in VLSI CAD Tools

Yield is becoming an increasingly severe issue in submicron VLSI and in WSI. As a result, there has been a significant increase in studies of yield, emphasizing extension of CAD VLSI design tools to achieve less defect-sensitive layouts in critical circuitry. Maly [56] has discussed not only simple particulate-induced defects and point defects but also lithography alignment errors as contributors to yield degradation, from this perspective of optimizing the layout to achieve high circuit yield. A specific CAD tool, RYE, has been developed at Carnegie-Mellon [57] to evaluate yield of specific IC chips, using analytic techniques within a hierarchical representation of the circuit. The hierarchical approach locally evaluates specific physical defect mechanisms, translating those defect mechanisms to a less specific defect description at the higher levels of circuit description. Other work at Carnegie-Mellon University on CAD yield analysis tools is described in [58,59,60].

References

Tutorial 3 243


