Backend Low-k TDDB Chip Reliability Simulator

Muhammad Bashir, Dae Hyun Kim, Krit Athikulwongse, Sung Kyu Lim and Linda Milor
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA USA
mbashir@gatech.edu

***Abstract*—Backend low-k time-dependent dielectric breakdown degrades reliability of circuits with Copper metallization. We present test data and link it to a methodology to evaluate chip lifetime due to low-k time-dependent dielectric breakdown. Other failure mechanisms can be integrated into our methodology. We analyze several layouts using our methodology and present the results to show that the methodology can enable the designer to consider easy design modifications and their impact on lifetime, separate from the design rules.**

Keywords-Copper interconnects; TDDB; dielectric breakdown; chip lifetime; reliability

I. INTRODUCTION

Copper (Cu) is the choice material for metallization in today's very-large-scale integrated (VLSI) circuits. When Cu metallization is used with materials termed low-k dielectrics, which have a dielectric constant (k) lower than that of Silicon Dioxide (SiO_2), the combination is known as Cu/Low-k interconnect. Cu/Low-k interconnects reduce interconnect delays and coupling capacitances. However, these performance advantages are accompanied by drawbacks critical to reliable chip operation and lifetime.

The primary reason of using low-k dielectrics is the reduction in parasitic capacitance in comparison with SiO_2 dielectrics. However, lower-k dielectrics are formed by increasing porosity, at the possible cost of reducing reliability. Moreover, each technology node reduces the interconnect dimensions without always reducing the supply voltage by the same proportion. This results in the backend dielectric in the newer nodes being subjected to higher electric fields than the previous nodes.

Figure 1(a) shows a cross-section of a conventional Cu/Low-k interconnect stack. The dielectric between the interconnect lines is the one that is most vulnerable to breakdown since it is stressed by the highest electric fields and uses material with the lowest dielectric constant. Hence, test structures used to evaluate backend dielectric breakdown are single-layer test structures that stress the dielectric between the lines in the same layer.

The standard approach to assess backend dielectric reliability is by using process data. The typical test structure is a comb structure, as shown in Figure 1(b) (top view). In testing a comb structure, a voltage difference is applied between the two combs. The current between the combs is monitored to determine the time-to-failure (*TF*).

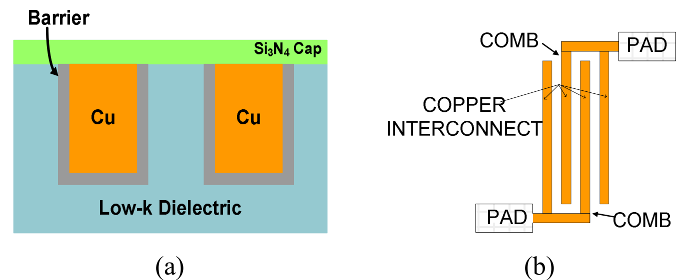


Figure 1. (a) Cross section of a typical Cu/Low-k interconnect (b) top view of a comb test structure.

Test structures are stressed at high voltages and high temperatures to accelerate dielectric breakdown and device aging. Appropriate adjustments, and extrapolations, are made to the test results to scale them to actual operation conditions. In addition, corrections are also needed to account for the difference between the vulnerable area of the chip and the test structure. There is no literature on the method to find the vulnerable area for a chip for backend dielectric breakdown, except for the statement that the vulnerable area is "the total length of such [minimum spaced] lines within a product" [1].

Chip reliability analysis requires techniques to extend the results gathered from small test structures and circuits to large complex chips. Such an endeavor must also be accompanied by solutions to manage and use the deluge of data that comes with analyzing large layouts. The physics describing IC failure mechanisms both in the front-end and in the backend has matured as a result of years of refinement to existing theories. However, the extension of these models to large and complex circuits has not proven to be straightforward and is complex.

The purpose of this paper is to present a methodology to assess chip lifetimes based on low-k TDDB chip lifetimes, by developing the link between data collected from test structures and the chip. We demonstrate the feasibility of our methodology by presenting results from a simulator based on the proposed methodology. Our methodology includes all layers of a chip, which can have different vulnerable areas.

The ultimate purpose of our work is to introduce backend dielectric reliability in design. The onus of meeting this end falls on the designers, if the reliability concerns and the accurate estimation of chip lifetimes can be conveyed to the designer in a designer-friendly manner. Designers ensure reliability, often inadvertently, by strictly adhering to design rules that assume worst case scenarios. Design rules often do not adapt themselves to the complexity of different circuits and

This work has been sponsored by SRC.

operating conditions. Instead, design rules are kept general enough to encompass a large number of circuits. A design rule that may be too restrictive for one design can be completely unrealistic for another design.

As chip complexity increases, designers become less aware of the actual physical functioning of their chip. Similarly, extending the models of the physics of failure to a chip requires the consideration of a myriad of factors. The task can be simplified if test results are used incrementally, as proposed in this work.

We start by commenting on the efforts to simulate chip level reliability, followed by laying out the guidelines for our methodology, and consequently the simulator. In section IV, we summarize our test structures and test results. In the following section, we outline our methodology. In section VI, we detail the circuits we have used to run our simulations. In section VII, we study the results from our simulator and present the insights gathered from our results. Section VIII concludes the paper.

II. Reliability Simulations and Simulators

The most important reliability concerns for interconnects are electromigration, stress-induced voiding, and TDDB of the backend dielectric [1]. To date, reliability simulators for interconnects have only been developed for electromigration and stress evolution [2], [3]. Our purpose is to consider an additional wear-out mechanism: backend low-k TDDB that is currently not included in reliability simulators.

A. *Modeling system lifetime*

It should be noted that circuits wear-out for a variety of reasons, both related to devices and interconnect. All of these wear-out mechanisms happen simultaneously. It is common to describe reliability mechanisms with a Weibull distribution

$$P(t) = 1 - \exp\left(-\left(\frac{t}{\eta}\right)^{\beta}\right), \quad (1)$$

having two parameters: the characteristic lifetime (η) and the shape parameter (β) [4]. The characteristic lifetime is the time-to-failure at the 63% probability point, when 63% of the population has failed, and the shape parameter describes the dispersion of the failure rate population. Typically, the shape parameter is close to one. If we have a collection of n independent wear-out mechanisms modeled with Weibull distributions, having parameters, $\eta_i, i = 1,\ldots,n$, and $\beta_i, i = 1,\ldots,n$, then the characteristic lifetime of the system, η, i.e. the time when 63% of the population has failed from any mechanism, is the solution of [5]

$$1 = \sum_{i=1}^{n}\left(\frac{\eta}{\eta_i}\right)^{\beta_i}. \quad (2)$$

If the characteristic lifetime of one mechanism is significantly smaller than others, this mechanism will dominate the failure rate. However, in general, it is prudent to consider *all* major sources of wear-out. Hence, as we scale the dimensions of interconnect and lower the dielectric constant, one should no longer neglect the potential reliability failures due to backend low-k dielectric.

B. *TDDB Models*

We note that models that describe backend TDDB, although they may have been initially developed for device TDDB, are of the general form [6], [7], [8], [9]

$$\ln TF = A - \gamma E^{m}, \quad (3)$$

where A is a constant that depends on the material properties of the dielectric, γ is a field acceleration factor, m is one for the E model [6],[7] and $m = 1/2$ for the $\sqrt{E}$ model [8], and TF is the time-to-failure. Note that although these models can be generally represented in this form, they are based upon very different physical mechanisms. This representation is only used for modeling.

Time-to-failure is both a function of the electric field and temperature. Equation (3) provides the correction that takes into account the difference in the electric field between use conditions and during accelerated stress test. The temperature dependence is modeled with an Arrhenius relationship [8]

$$\ln TF = B - \frac{C}{T}, \quad (4)$$

where B and C are constants. Equation (4) provides a correction between chip operating conditions and accelerated stress conditions. There is a concern that stressing at high temperatures can activate failure modes that are not present during use conditions. Hence, stressing at high electric fields is preferred in comparison with testing at high temperatures. Our tests were conducted at $150^{o}C$.

III. Ground Rules

This work forms an interface among reliability physicists, semiconductor foundry engineers, and designers by suitably partitioning their combined effort, thereby keeping each unburdened with the details of the other's efforts.

Test structures have identified the well known sensitivities of backend dielectric breakdown to area and distance between the lines (electric field in the dielectric) and the less well known sensitivity to metal linewidth. Through our methodology [5], we link test structure data to the entire product die (chips), by extracting the corresponding "vulnerable areas" for a chip, and we characterize the failure rate for the chip by combining the failure rates due to all "vulnerable areas".

The simulator focuses on chip wear-out due to backend dielectric breakdown. The results can be combined with other wear-out mechanisms via (2).

The simulator combines the test results from test structures that stress different vulnerable areas to determine the failure rate for a chip. In other words, our methodology links chip layout geometries to those on the test structures. We assume

that these results can be scaled to use conditions with whatever version of (3) and (4) that is deemed to be appropriate.

The focal point of our work is the mapping between the test structures and the chip. We assume that the conductors in the chip are very similar to the conductors in the test structures and that they have similar geometries. Conductor geometry does impact the data that is collected from the test structures and the models developed from these data. The methods that account for conductor geometry will be noted in the following sections.

IV. TEST STRUCTURE DESIGN AND TEST RESULTS

A. Test structures

We have designed test structures to assess the impact of area and linewidth on Cu/low-k TDDB. The details of the test structures, their design and results, are given in [10] and [11]. The test structure set includes comb structures that vary area, linespace, and linewidth.

The test structures were manufactured with an industrial 45nm dual-damascene process and subjected to accelerated stress tests at $150^{o}C$ with electric fields ranging from $0.25MV/cm$ to $1.5MV/cm$. Breakdown was considered as the point of onset of leakage current greater than $100\mu A$. The sample size was 30 and the Weibull failure rate distribution was used to model the failure population.

B. Test results

Test results indicated a strong impact of area. Die-to-die linewidth variation creates curvature in failure rate distributions. This curvature does not impact η. Hence we extract η and determine β by area scaling. Once these parameters have been determined for the unit area, the relationship between characteristic lifetimes for different areas in known.

Note that LER can also be taken into account when extrapolating failure rates. The impact of LER is considered in [11].

Lifetime is also impacted by the linewidth on each side of the dielectric segment. We consider linewidth variation when determining η and β for our test structures [10]. Specifically, linewidth can be taken into account by extracting η as a function of linewidth, since β, which is extracted from the area test structures, is assumed to be constant.

Test results showed a strong impact of linewidth on low-k TDDB, even when the vulnerable area and linespace of the dielectric under stress remained constant. If W_a is the actual linewidth and W_d is the drawn linewidth, then the difference between them is given by $\Delta W = W_a - W_d$. The shift in linespace is $S_a = S_d - \Delta W$, where S_a is the actual linespace and S_d is the drawn linespace. This shift arises because of aspect-ratio-dependent-etching (ARDE) [12]. During etch a protective compound builds up on the sidewalls of wide trenches, preventing lateral etch. This compound does not build up as much in narrow trenches. Hence, narrow trenches suffer from greater lateral etch near the critical CMP interface. Linespaces with larger positive values of ΔW breakdown faster, since $E = V/S_a$. We can use data directly to determine the relationship between the drawn linespace and η through regression.

Our data indicates a difference in linespace as a function of the widths of lines on the right and left as shown in Figure 2.

If SEM data is available, then Equation (3) along with $E = V/S_a$ can be used to calculate A and γ. The ΔW can also be found for any dielectric with any linewidth on the right and the left. The model in Figure 2 gives an estimate of ΔW and S_a. The constants from Equation (3), in turn, provide an estimate of the corresponding characteristic lifetime.

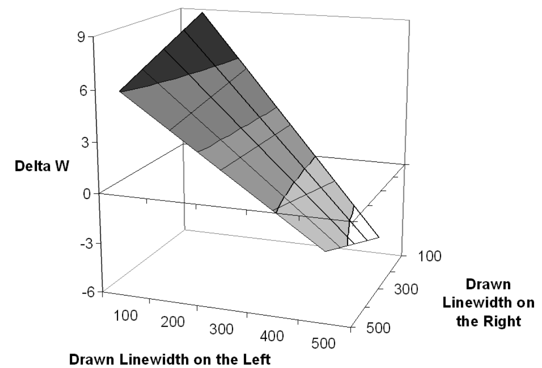


Figure 2. Variation in linespace as a function of the widths of the lines in either side of the dielectric. The data was collected using scanning electron microscopy (SEM).

Figure 3 shows a plot of characteristic lifetime varying with area and linespace, extracted from our test structures.

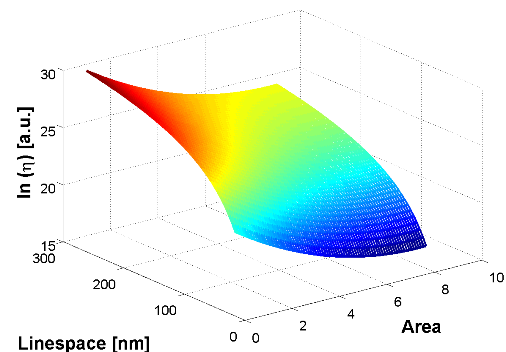


Figure 3. η for the $\sqrt{E}$ model. Area is the ratio of the area of extrapolation to the unit area test strcture.

This plot is obtained by determining the characteristic lifetime for different area ratios, in comparison with the $1X$ test structure, for different linespaces, i.e.,

$$\ln \eta \propto f(S, A), \quad (5)$$

where S is the linespace and A is the area.

C. Scaling test results to use conditions

Electric field and temperature can affect the relationship between test conditions and use conditions. The relationship between test conditions and use conditions is given in Equation (3). However, the test structure is stressed with DC stress while the chip dielectrics undergo AC stress. Nonetheless, it should be noted that the backend dielectric TDDB under AC stress does not show any recovery [13], as observed in bias temperature instability degradation, and lifetime relaxation or healing, as observed in degradation due to electromigration.

In our analysis we assume a signal activity factor of 0.5. Figures 4 shows our results scaled to use conditions for 45nm technology, with a supply voltage of 0.8V under alternating pulsed stress.

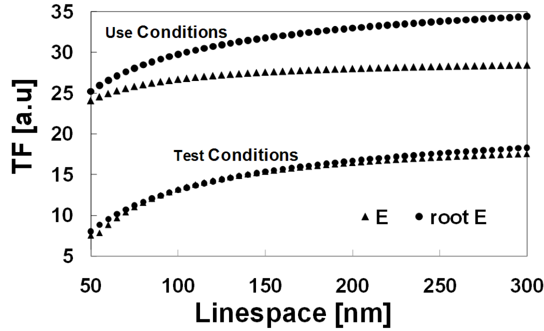


Figure 4. Test results, for structures used to isolate the impact of linewidth, scaled to use conditions.

It should be noted that segments of the circuit may undergo different signal activity factors. In that case, these sectors should be analyzed separately, before combining them with estimates of lifetime of other sectors. Similarly, segments of the circuit may experience different values of average temperature. These sectors should also be analyzed separately, before combination with estimates of lifetime from other sectors.

V. TDDB CHIP LIFETIME

A. Feature extraction

We determine the vulnerable sites, vulnerable area, in a given layout from the layout features. The vulnerable area is a block of dielectric between the two Copper lines separated by linespace S_1 for length L_1 and having an area S_1L_1.

The feature that is extracted from layouts is the vulnerable length between two lines L_i associated with a linespace S_i, which is a function of the widths of the two adjacent lines, $W_{i,L}$ and $W_{i,R}$, determined by the model in Figure 2.

A given layout is analyzed by determining the pairs $\left(S_i(W_L, W_R), L_i\right)$ for each layer for all linespaces surrounded by the linewidths W_L and W_R. When we integrate temperature profiles, we partition the layout with a $5\mu m \times 5\mu m$ grid prior to the extraction of $\left(S_i(W_L, W_R), L_i\right)$.

The details of our methodology can be found in [5]. Here we give its gist in the following subsections.

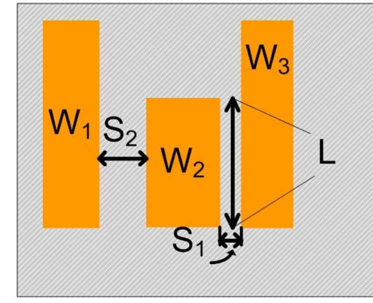


Figure 5. Vulnerable area characterized by the linespace. The rectangles are Cu wires and the shaded area is the low-k dielectric.

B. TF for a layer

Let η_t be the characteristic lifetime associated with vulnerable linespace S_t of length L_t. η_t is determined by stressing a test structure with vulnerable linespace S_t of length L_t. In the layout, if the vulnerable length corresponding to S_t is L_v, then the corresponding characteristic lifetime is

$$\eta_v = \eta_t \left(\frac{L_t}{L_v}\right)^{1/\beta}. \tag{6}$$

Since a chip has many different linespacings, we combine failure rates for all linespacings by computing a defect count for each linespacing in the layout. The total defect count for a layer is the sum of all these defect counts. The characteristic lifetime for a layer (η_l) at the probability point $P = 0.63$ is

$$\eta_l = \sum_n \left(\frac{1}{\eta_n^\beta}\right)^{-1/\beta}. \tag{7}$$

C. Chip lifetime

The defect density of the chip is computed from the defect densities of the individual layers that are in turn computed by the defect densities of all the linespace groups present within a layer. The chip lifetime (η_{chip}) is computed as

$$\eta_{chip} = \sum_l \sum_n \left(\frac{1}{\eta_n^\beta}\right)^{-1/\beta}. \tag{8}$$

Unlike for a single layer, multiple layers of a chip may have different process details. In that case, data would be collected for each layer separately. If β were not common to all layers, then η_{chip} is implicitly defined as

$$1 = \sum_l \sum_n \left(\frac{\eta_{chip}}{\eta_n}\right)^{\beta(l)}. \tag{9}$$

Since we only have data from one layer, we assume that CMP, etching and photolithography impact all the layers in the same way. This assumption is simplistic, and if data from

different layers is available, it can be easily incorporated into Equations (7) - (9).

Reliability is adversely affected by linewidth variation and line edge roughness (LER). It has been shown that large scale linewidth variation impacts β [10], while LER impacts η with little or no impact on β [14]. We consider linewidth variation when determining η and β for our test structures [11]. Any effect of LER is reflected in the characteristic lifetime through η_t in equations (7) and (9) [14], and thus is included in the results.

D. Temperature profile

Including the temperature map in the layout statistics adds another dimension to the problem because now we have to consider the different characteristic lifetimes at different temperatures for every linespace. If we have a collection of m different temperatures for a linespace S_1, then the corresponding characteristic lifetime for the linespace is

$$\eta_{S_1} = \sum_m \left(\frac{1}{\eta_m^\beta} \right)^{-1/\beta}. \qquad (10)$$

Characteristic lifetimes for the layer and the chip can be calculated using (7) and (9).

E. Overview

We extract pairs $\left(S_i(W_L, W_R), L_i\right)$ for each layer to determine the vulnerable length associated with each linespace in a layout and then determine the associated η. We also partition the layout by temperature. For the chip, we determine lifetime from (9) after summing the defects on all the layers.

VI. Details of the Circuits

A. Process

The NCSU 45nm technology library was used for our experiments [15]. This process has ten metal layers and the details of relevant features are given in Table 1.

TABLE I. METAL LAYERS IN NCSU 45NM PDK

Metal Layer	Minimum Linewidth [nm]	Minimum Linespace [nm]
1	65	65
2,3	70	70
4, 5, 6	140	140
7, 8	400	400
9, 10	800	800

B. Circuits

We synthesized the radix-2 pipelined, 256-points and 512-points, Fast Fourier Transform (FFT) HDL source code [16]. The circuit *cf_fft_256_8* has 324k gates and 329k nets. The circuit *cf_fft_512_8* has 708k gates and 712k nets. Both circuits have precision eight. The names of different instantiations of *cf_fft_256_8* and *cf_fft_512_8* start with *f1* and *f2*, respectively. The block diagram of the circuit is shown in Figure 6.

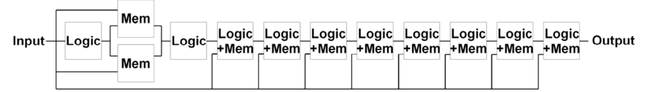


Figure 6. Block diagram of a FFT circuit.

Synopsys Design Compiler is used for synthesis [17]. Cadence SoC Encounter is used for placement, clock-tree synthesis, routing, optimization, and RC extraction [18]. Synospys PrimeTime is used for timing analysis [19]. We use seven different instantiations of *cf_fft_256_8* and three different instantiations of *cf_fft_512_8* .

Our metrics of performance comparison are the number of layers in a circuit and its timing performance. The details of the circuits are given in Table 2. Table 2 shows the timing performance of each circuit, the total wirelength of each circuit, and the percentage of total wirelength in each layer, referred to as wire density.

Circuits *f1_M5*, *f1_M6*, *f1_M7*, and *f1_M8* are used to isolate the impact of the number of layers on reliability. Circuits labeled '*M'* use Metal1 to Metal'*X'* during routing. Using more routing layers tends to result in shorter wirelengths and better timing performance, as shown in Table 2.

Circuits *f1_RT1*, *f1_RT2*, *f1_RT3*, *f2_RT1*, *f2_RT2*, and *f2_RT3* are used to analyze the impact of timing performance on reliability. In RT'*Y*', we optimize timing using buffer insertion and gate sizing. M'*X*' does not do this. A higher value of '*Y*' means more aggressive timing optimization with a higher clock frequency.

All the circuits have been synthesized using the same technology library. Thus the values in Table 1 are consistent across all the instantiated circuits.

C. Vulnerable Area Extraction

We have developed our layout extraction tool using standard object oriented programming languages. The layout extraction flow is shown in Algorithm 1. We explain the flow for horizontal line segments only. Vertical line segments can be handled in a similar way.

We start by reading in line segments for each metal layer from a given layout (ReadLineSegments in Algorithm 1) and by sorting all the horizontal segments in the ascending order of the y-coordinate of the bottom-left corner of the line segments, and the x-coordinate of the bottom-left corner (tie-breaker for equal y-coordinates). A layout may have a large number of line segments. For instance, there are eight million segments in Metal2 of *cf_fft_512_8*. Hence, a fast sorting algorithm is required. Therefore we use Bucket sort in the ReadLineSegments process in Algorithm 1.

After the reading-in process, we compare two adjacent line segments. If their vertical spacing is less than or equal to the pre-determined maximum line spacing, there exists a vulnerable area surrounded by these two line segment. We

TABLE II. WIRELENGTH OF INDIVIDUAL METAL LAYERS, AS WELL AS TIMING PERFORMANCE AND RELIABILITY, OF THE USED DESIGNS. THE TABLE SHOWS THE PERCENTAGE OF TOTAL WIRELENGTHS OF CHIPS PRESENT IN EACH LAYER. WIRELENGTH (WL) AND TIMING PERFORMANCE, CRITICAL PATH DELAY (CPD), ARE ALSO GIVEN.

Layer	Design									
	f1_M5	**f1_M6**	**f1_M7**	**f1_M8**	**f1_RT1**	**f1_RT2**	**f1_RT3**	**f2_RT1**	**f2_RT2**	**f2_RT3**
Metal1	1.4	1.4	1.4	1.4	2.59	0.77	2.20	2.23	2.24	1.19
Metal2	18.8	18	18.1	18.1	22.61	14.10	19.56	20.83	23.37	15.80
Metal3	33.9	33.1	33.1	33.1	38.41	25.80	32.01	35.44	37.92	27.90
Metal4	29.5	25.7	25.7	25.6	24.95	18.89	23.11	24.90	25.3	22.75
Metal5	16.2	16.1	15.4	15.3	9.52	19.92	15.49	13.42	9.04	19.83
Metal6		5.4	5.07	5.03	1.86	15.91	6.99	2.77	2.07	12.53
Metal7			0.9	1	0.05	4.60	0.65	0.38	0.04	
Metal8				0.1		0.01			0.004	
Metal9						0.01				
Metal10						0.0001				
Total Wirelength [m]	8.06	8.05	7.99	7.99	6.28	13.56	7.52	15.06	13.91	21.09
CPD [ns]	3.51	3.51	3.33	3.29	3.16	2.9	2.86	2.909	2.96	2.98

Algorithm 1 : Layout extraction flow.

Input: The maximum line spacing S_{max} and a given layout L
Output: A table of vulnerable areas (VulnerableAreaTable)

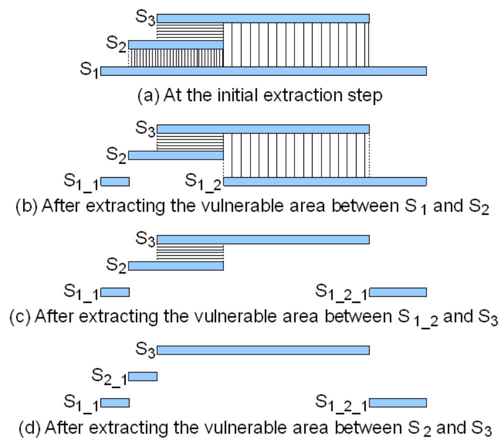
for each metal layer m do
  LineData $(m) \leftarrow$ **ReadLineSegments** (L); // BucketSort
  $c \leftarrow 1;$
  $n \leftarrow 2;$
  while true **do**
    if $c = N_{line}$ **then** // N_{line} : # lines in LineData
      break;
    end
    $L_1 \leftarrow$ LineData (m,c); // $c-$th line
    $L_2 \leftarrow$ LineData (m,n); // $n-$th line
    if Spacing $(L_1, L_2) \le S_{max}$ **then**
      VulnerableAreaTable $(m) \leftarrow$ **VulnerableArea** (L_1, L_2);
      LineData $(m) \leftarrow$ **Cut** (L_1, L_2);
      Adjust (N_{line}, c, n)
    else
      $c \leftarrow c+1;$
      $n \leftarrow n+1;$
    end
  end
end

insert this into our vulnerable area table. After the extraction of a vulnerable area, we apply two post processes (Cut and Adjust in Algorithm 1). In the Cut process, we remove one of the compared line segments from our data structure (LineData), and insert one or two new segment(s) into the data structure. In the Adjust process, we readjust indexes of line segments for the next comparison.

We illustrate our algorithm with the help of an example shown in Figure 7. In this example, there are three vulnerable segments (between S_1 and S_2, between S_1 and S_3, and between S_2 and S_3) as shown in Figure 7(a). We first compare two line segments S_1 and S_2, and find a vulnerable area. After we store this vulnerable area in our vulnerable area table, we apply the Cut process. In this process, we remove S_1 from LineData, cut the overlapped part from S_1, create two segments (S_{1_1} and S_{1_2}), and insert these segments into LineData (sorting is performed during insertion) as shown in Figure 7(b). Since there are four new line segments, we increase the total number of segments by one and start comparing segments with S_{1_1} (Adjust process). S_{1_1} does not overlap with any other line segment, and therefore we proceed to S_{1_2} and find a vulnerable area between S_{1_2} and S_3. We store this vulnerable area in our vulnerable area table, and cut S_{1_2} as seen in Figure 7(c). In Figure 7(d), we find a vulnerable area between S_2 and S_3. We store it and cut S_2.

D. Runtime

The runtime for the simulator is the sum of the time taken to extract features from the layout and a constant time to



Step	VA	W_1	W_2	S	L	T
(b)		0.2	0.2	0.36	1.80	61.5
(c)		0.2	0.2	0.92	3.86	61.6
(d)		0.2	0.2	0.36	0.26	61.4

(e) Vulnerable area (VA) table

Figure 7. Extraction of vulnerable area (a) at the initial step (b) after extracting the vulnerable area between S1 and S2 (c) After extracting the vulnerable area between S1_2 and S3 (d) After extracting the vulnerable area between S2 and S3 (e) Vulnerable area (VA) table: W1 and W2 denote linewidths, S is the linespace, L is the vulnerable length and T is the temperature

evaluate Equations (7) - (9). Complexity of feature extraction and database extraction is $O(n)$, where n is the number of features, since bucket-sort is used. Complexity of extracting statistics from the features is also $O(n)$, because we scan the bucket from the bottom most element, and the maximum number of features within a fixed distance from an element is constant. Lifetime is estimated in constant time.

VII. Effect of Layout on TDDB Reliability

A. Results based on geometry

Figure 8 shows η for Metal1-Metal6 for the circuits used in the study and the chip according to the $\sqrt{E}$ Model.

B. Observations

η for chips are more pessimistic than η for individual layers, because in calculating characteristic lifetime for the chip we combine the vulnerable areas for all the layers. Figure 8 does not indicate a trend for reliability with respect to timing performance. Figure 9 shows the lack of correlation between timing performance and reliability. Our results show that increasing the number of layers affects reliability marginally, while decreasing wirelengths increases reliability, as shown in Figure 9.

1) Number of layers

Using more metal layers generally results in a decrease in routing congestion and reduces the need for long detours to avoid routing congestion. This leads to less coupling capacitances between wires, which results in a decrease in the critical path delay. Since a router can spread out wires in several metal layers, we expect this to improve reliability.

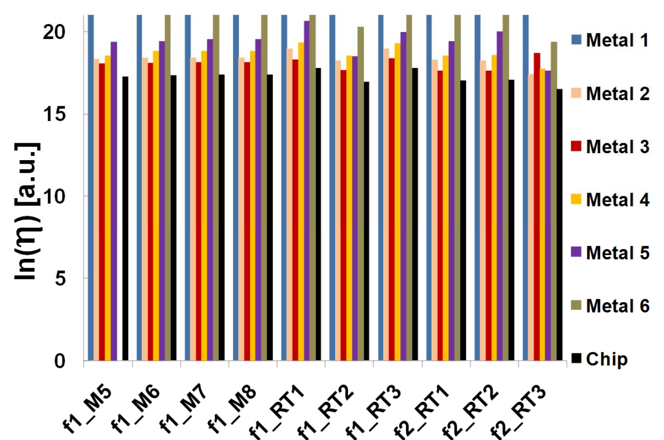


Figure 8. η for each layer and η for the chip for the circuits using the $\sqrt{E}$ Model.

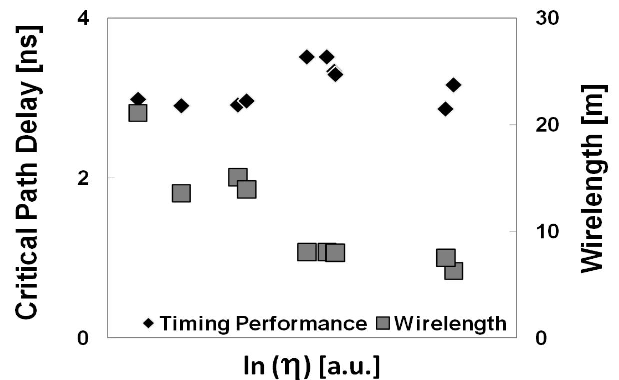


Figure 9. A comparison of reliability, timing performance and wirelength for the circuits under study.

critical path delay. Since a router can spread out wires in several metal layers, we expect this to improve reliability.

As expected, the critical path delay decreases as we increase the number of metal layers. However, as shown in Figure 8, reliability increases only marginally as the number of layers increases, and the layer most critical to lifetime remains the one with the highest wire density. This change of reliability as a function of the number of layers, or lack thereof, can be expected because even though the number of layers increases from five to eight, the percentage of total wirelength in the additional layers is less than 6%. Moreover, a large percentage of the wirelength still remains in a single layer, Metal3. Metal3 has mid-distance interconnects performing vital operations and it is highly unlikely that any optimization would cause major changes in Metal3. An even distribution of wirelengths can lead to an increase in lifetime This will the topic of future research.

2) Critical physical features

Figure 10 shows the characteristic lifetime for each layer of *f1_M5*, η for the critical linespace, and the linespace group with the smallest η according to (6), for the same layer. The critical linespace group is the most frequent linespace in a layout. It is not necessarily the smallest linespace.

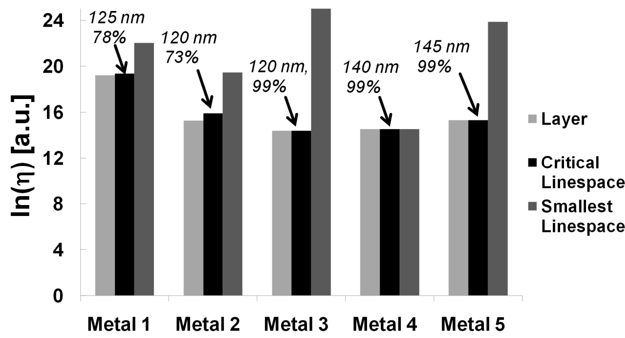


Figure 10. η for each layer of f1_M5, the critical linespace group, and for the smallest linespace group in each layer. The critical linespace, along with its percentage, is also given.

All dielectric area in a layer and in the chip falls in some linespace group, determined by its immediate neighbors. According to (7), η for a layer is dominated by the η of the critical linespace group, i.e. the most frequent linespace group. Figure 10 shows the percentage of dielectric area formed by the critical linespace group. A way to increase η for layers with a critical linespace greater than the smallest linespace is to redistribute linewidths. We could optimize this distribution for reliability

If we were to estimate the characteristic lifetime based on the most frequent (critical) linespacing alone, we only need to determine the area for this single linespace in each layer. Such an approach is simplistic. However, Figure 10 shows that lifetime estimates based on the critical linespace group are reasonably accurate.

Figure 10 also shows the characteristic lifetime, η, for the minimum linespace group for each layer, where the minimum linespace for each layer is given in Table 1. Consider Metal2 in f1_M5, the smallest linespace in Metal2 is 70nm, but the linespace dominating the characteristic lifetime, η, for this particular layer is 120nm for both the E Model and the $\sqrt{E}$ Model. Forty different linespaces are present throughout the layer, with the minimum being 70nm and the maximum being 252.5nm. However, 73% of the dielectric segments in this layer have a linespace of 120nm, with only 0.11% of dielectric segments having a linespace of 70nm between them. Therefore, we cannot just consider the smallest linespace group, as suggested in [1], when the layout is dominated by a linespace group other than the minimum linespace. Such an approach leads to lifetimes that are optimistic by orders of magnitude.

3) Wire density

Figures 11 shows that there is a strong correlation between wire density, the proportion of total wirelength in a layer, and η_l, with layers having the highest wire density dominating η_l in (8). This result is expected because of the nature of the breakdown mechanism. Higher wire coverage is achieved by closely packing the metal lines together, resulting in an increase in E and consequently degrading reliability. Also, as expected, reliability increases with a decrease in wirelength.

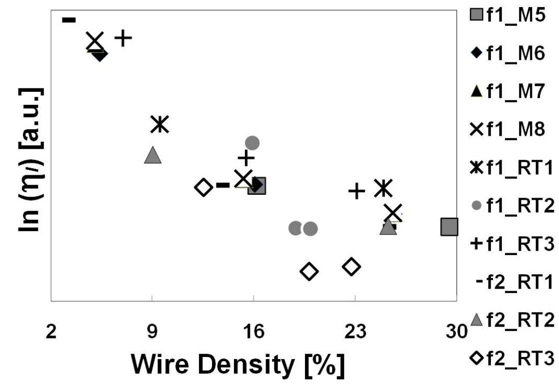


Figure 11. Wire density for Metal4-Metal6 and their characteristic lifetimes according to the $\sqrt{E}$ model.

4) Linewidth and Reliability

Linewidth impacts TF by affecting etching [11] and photolithography. As shown by our previous results, η increases as linewidth increases, for a given linespace, because of the interaction of physical design with etching and photolithography [11]. However, we found that increasing the linespace design rule does not improve reliability because the overall dielectric area increases [20] after re-routing .

5) Timing and Reliability

Timing optimization is achieved through buffer insertion, changing gate locations, and gate sizing. In terms of interconnect, densely routed areas raise coupling capacitance issues, which are addressed by ripping-up and re-routing the nets. Wire sizing is another way to obtain timing performance although we did not use it.

Buffer insertion results in an increase in total wirelength, resulting in a decrease in reliability, as apparent from the results. Gate re-placement needs to be managed carefully from the reliability perspective because re-placing gates in a crowded region can result in higher electric fields. If gate sizing is used for timing optimization, then the goals of timing align with those of reliability. Increasing the gate size increases the degrees of freedom for wire-to-pin connections, and this can be taken advantage of to enhance reliability. Rip-up and re-routing are aimed at reducing wiring congestion and coupling capacitance, factors that are also critical to reliability.

Despite all of these factors we did not observe any relationship between timing and reliability, because timing optimization generally uses heuristic algorithms instead of deterministic algorithms.

The results showed a strong correlation between the coverage in a given layer by the critical linespace group and the TF. For instance, for circuits *f1_RT1*, *f1_RT2*, and *f1_RT3*, 99% of the lines in Metal3 are separated by two linespacing groups, 120nm and 310nm. The lifetime is determined by the 120nm linespace group. Interestingly, for the circuits optimized for timing, in every layer 95% of the lines fall into only three linespace groups, and out of these three, more than 50% of linespaces were from the critical linespace group.

C. Results based on geometry and function

The steady state temperature of a point $p = (x, y, z)$ inside a thermal structure can be obtained by solving the heat equation

$$\nabla \cdot (k(p)\nabla T(p)) + S_h(p) = 0, \qquad (11)$$

where k is the thermal conductivity, T is temperature and S_h is the volumetric heat source. This model can be implemented by meshing the integrated circuit (IC) structure into thermal cells. To perform the thermal analysis, we start with the layout in DEF or GDSII format from Cadence SoC Encounter [18] and then perform static power analysis, for a given circuit frequency (f) and logic cell switching activity (α_i), to determine power dissipation

$$P_i = (\alpha_i C_i V_{DD}^2 f)/2, \qquad (12)$$

where C_i is the loading capacitance of a logic cell, V_{DD} is the supply voltage, and f is the clock frequency of each logic cell. The layout along with the logic cell power dissipation is then used by our analyzer. The analyzer automatically generates the meshed structure for the IC along with the thermal conductivity and the volumetric heat source of each thermal cell. This information is used to perform thermal analysis using ANSYS FLUENT [21]. Figure 12 shows the thermal map, with an activity factor of 0.5, for Metal3 of *f2_RT3*.

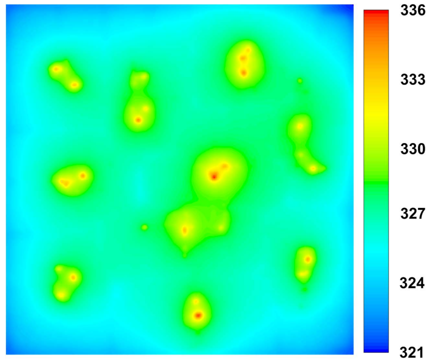


Figure 12. Thermal map of Metal3 of the circuit *f2_RT* for an activity factor of 0.5.

1) Runtime of thermal simulations

The runtime of thermal analysis consists of the runtime for determining the percentage of material in each thermal cell to determine the thermal conductivity, the volumetric heat sources inside each thermal cell of the meshed structure, and the runtime for solving the partial differential equations. The worst case complexity for the former is $O(n^2)$ and the average is $O(n \log n)$, n being the number of layout geometries. FLUENT [21] uses the finite volume method and its runtime varies between $1/40^{th}$ and $1/25^{th}$ of the time it takes to determine the percentage of the material in each thermal cell and the volumetric heat sources. Note that once the thermal analysis has been done, the layout statistics are generated, whose runtime is given in Section VI. They are integrated with the thermal profile of the chip.

2) Results

Figure 13 shows the results for our circuits incorporating their temperature profiles for a given signal activity level.

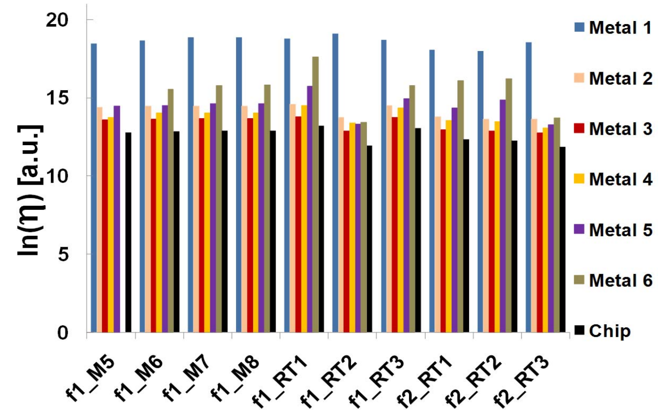


Figure 13. η for each layer and η for the chip for the circuits using the E Model and temperature profiles.

The trend among the models and the circuits remains the same after integrating the temperature profiles. Only the magnitudes of the characteristic lifetimes change. Figure 14 shows lifetimes with and without temperature profiles for the circuits for both the E model and the $\sqrt{E}$ model.

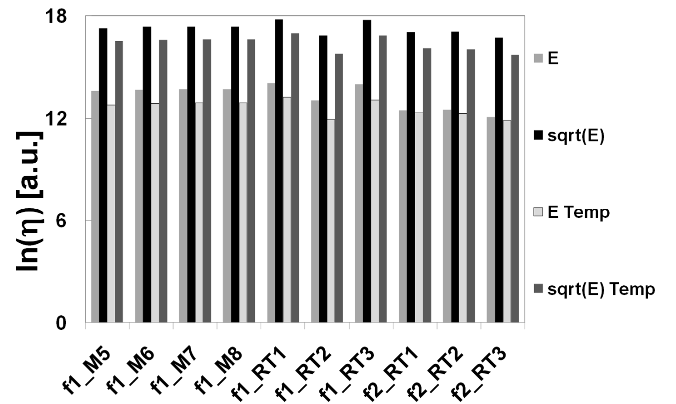


Figure 14. Characteristic lifetimes for the circuits according to the E model and the $\sqrt{E}$ model.

Figure 15 shows the characteristic lifetime for all layers of *f1_M5* for both the E model and the $\sqrt{E}$ model, with and without the thermal map.

D. Temperature Map

Integrating the temperature profile in our methodology takes into account the variation in characteristic lifetime caused by the variation in on-chip temperature. However, for large layouts the best-case complexity for generating the thermal map is greater than the worst-case complexity for generating the layout statistics. Moreover, different input vectors will affect the thermal map differently, thus requiring exhaustive thermal profiling, unless there are formal methods to generate thermal profiles. Hence the efficacy of including thermal maps

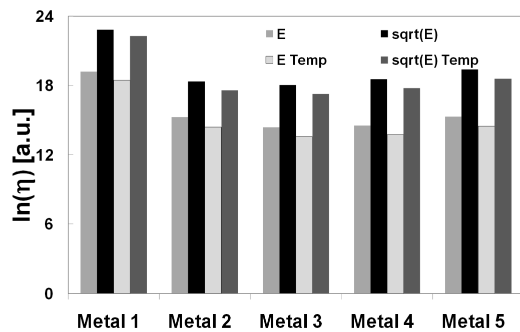


Figure 15. Characteristic lifetime for all layers of *f1_M5* with and without temperature according to the E Model and the $\sqrt{E}$ model.

ultimately depends on the intended use of the simulator. If the simulator is being used for accurate reliability estimates, then thermal maps must be integrated. The same will be the case if the intention is to observe the effect of a particular class of input signals. However, if the intended use for the designer is to get some quick reliability numbers, then results without thermal maps can give somewhat of an accurate guess at best, or describe the range of lifetimes at worst.

VIII. Conclusion

A methodology was proposed to assess backend TDDB chip reliability. The methodology has been developed in a way that other failure mechanisms can be integrated into it. Results from the simulator, built upon our methodology, showed the feasibility of our approach. In doing so, we also analyzed the effect of layout on backend TDDB reliability. We showed the absence of any correlation between timing performance and reliability. We also showed that greater wire coverage will result in smaller lifetimes, as expected. We demonstrated that the narrowest linespace group may not impact the chip lifetime critically. Instead, it is the linespace group with the highest coverage that is most instrumental in determining lifetime. We demonstrated that integrating temperature maps result in lower, though accurate, TF estimates. This study is a step towards our eventual goal of developing a tool to help designers build in reliability not just for backend low-k dielectric failures, but also for other failure mechanisms.

Our methodology does not assume the design to be “as drawn” or that the failures are being caused by the layout features. Of course, if such were the case, then we would not have been using an extreme value distribution. We assume that the layout is manufactured from the geometries in our test structures and modeling takes into account the invariance of Weibull statistics to area scaling, thus justifying extrapolations. It is assumed that any failure causing mechanisms that manifest themselves in test structures are reflected in the characteristic lifetimes used in the methodology.

References

[1] T. Pompl, *et al.*, "Practical aspects of reliability analysis for IC designs," in *Proc. Design Automation Conf.*, 2006, pp. 193-198.

[2] S. M. Alam, *et al.*, "Reliability computer-aided design tool for full-chip electromigration analysis and comparison with different interconnect metallizations," *Microelectronics Journal,* vol. 38, pp. 463-73, 2007.

[3] Synopsys Inc, "FAMMOS."

[4] E. Y. Wu, *et al.*, "On the Weibull shape factor of intrinsic breakdown of dielectric films and its accurate experimental determination. Part I: theory, methodology, experimental techniques," *IEEE Transactions on Electron Devices,* vol. 49, pp. 2131-2140, 2002.

[5] M. Bashir and L. Milor, "Towards a chip level reliability simulator for copper/low-k backend processes," in *Proc. Design, Automation & Test in Europe (DATE),* , 2010, pp. 279-282.

[6] G. S. Haase and J. W. McPherson, "Modeling of Interconnect Dielectric Lifetime Under Stress Conditions and New Extrapolation Methodologies for Time-Dependent Dielectric Breakdown," in *Proc.Int. Reliability Physics Symposium (IRPS)*, 2007, pp. 390-398.

[7] J. Kim, *et al.*, "Time Dependent Dielectric Breakdown Characteristics of Low-k Dielectric (SiOC) Over a Wide Range of Test Areas and Electric Fields," in *Proc. Int. Reliability Physics Symposium (IRPS)*, 2007, pp. 399-404.

[8] F. Chen, *et al.*, "A Comprehensive Study of Low-k SiCOH TDDB Phenomena and Its Reliability Lifetime Model Development," in *Proc. Int. Reliability Physics Symposium (IRPS)*, 2006, pp. 46-53.

[9] F. Chen, *et al.*, "Cu/low-k dielectric TDDB reliability issues for advanced CMOS technologies," *Microelectronics Reliability,* vol. 48, pp. 1375-1383, 2008.

[10] M. Bashir and L. Milor, "A methodology to extract failure rates for low-k dielectric breakdown with multiple geometries and in the presence of die-to-die linewidth variation," *Microelectronics Reliability,* vol. 49, pp. 1096-102, 2009.

[11] M. Bashir and L. Milor, "Analysis of the Impact of linewidth variation on Low-k Dielectric Breakdown," in *Proc. Int. Reliability Physiscs Symp.*, 2010, pp. 895 - 902.

[12] R. A. Gottscho, *et al.*, "Microscopic uniformity in plasma etching," *Journal of Vacuum Science and Technology* vol. 10, pp. 2133-47, 1992.

[13] J. Sung-Yup, *et al.*, "The characteristics of Cu-drift induced dielectric breakdown under alternating polarity bias temperature stress," in *Proc. Int. Reliability Physics Symp.*, 2009, pp. 825-827.

[14] M. Vilmay, *et al.*, "Copper line topology impact on the SiOCH low-k reliability in sub 45nm technology node. From the time-dependent dielectric breakdown to the product lifetime," in *Proc Int. Reliability Physics Symp.*, 2009, pp. 606-612.

[15] NCSU EDA. *NCSU Free PDK45*. Available: http://www.eda.ncsu.edu/wiki/FreePDK

[16] Launchbird Design Systems Inc. *CF FFT*. Available: http://www.opencores.org

[17] Synopsys Inc. *Design Compiler*.

[18] Cadence Design Systems Inc, "SoC Encounter RTL-to-GDSII."

[19] Synopsys Inc, "PrimeTime."

[20] M. Bashir, *et al.*, "Methodology to determine the impact of linewidth variation on chip scale copper/low-k backend dielectric breakdown," *Microelectronics Reliability,* vol. 50, pp. 1341-6, 2010.

[21] ANSYS Inc.,"FLUENT."