Design of Integrated-Circuit Interconnects with Accurate Modeling of Chemical-Mechanical Planarization

Lei He[a], Andrew B. Kahng[b,c], King Ho Tam[a] and Jinjun Xiong[a]

[a] EE Dept., University of California, Los Angeles, CA 90095, USA
[b] ECE Dept., University of California, San Diego, CA 92093, USA
[c] Blaze DFM, Inc., Sunnyvale, CA 94089, USA *

ABSTRACT

Dummy fill insertion in Chemical-mechanical Planarization (CMP) can change the coupling and total capacitance of interconnect. Moreover, dishing and erosion phenomena change interconnect cross-sections and hence significantly affect interconnect resistance. This work first studies interconnect parasitic variations due to (1) different fill patterns that are nominally "equivalent" with respect to foundry rules; and (2) dishing and erosion of conductors and dielectric using an accurate density-step-height model for multi-step CMP from the literature.[1] Our results show that for long parallel wires the variation of coupling capacitance between adjacent wires can be up to 25% and 300% for wires that are 3x and 6x minimum space apart respectively, and the variation of total wire capacitance can be more than 10%. We also show that the variation of wire resistance due to dishing and erosion can be over 30%. This work also evaluates how CMP effects (fill insertion, dishing and erosion) impact the achievable delay of buffered global on-chip interconnects. We obtain the delay of buses from accurate SPICE simulations considering CMP-related parasitic variation. Our studies show that the interconnect design considering fill and buffer insertion simultaneously with CMP effects reduces the unit length delay of global interconnect bus by up to 3.3% over the design which does not consider any CMP effects.

1. INTRODUCTION

Chemical-mechanical planarization (CMP) is an enabling technique to achieve wafer planarity in BEOL manufacturing processes. However, CMP also causes design variations due to *dummy fill* insertion[2] and *dishing* and *erosion*.[3] Dummy fill insertion improves the uniformity of metal feature density and enhances the planarization achieved by CMP, but it changes the coupling and total capacitance of interconnects.[2,4,5] Dishing and erosion phenomena change interconnect cross-sections,[6] and hence affect interconnect parasitics and performance. A work from the literature[7] reports more than 35% delay variation on long interconnect due to copper CMP process.

The first contribution of this paper is a study of interconnect parasitic variations due to (1) different fill patterns that are nominally "equivalent" with respect to foundry rules; and (2) dishing and erosion of conductors and dielectric similar to those predicted by ITRS.[8] We show that the variation of coupling capacitance between two adjacent wires and and total capacitance of one wire can be more than 300% and 10%, respectively, due to "pattern-dependent" fill insertion. Moreover, the variation of wire resistance due to dishing and erosion can be over 30%, but have limited impact on interconnect capacitance.

The second contribution of this paper explores possible improvement in the interconnect performance through accurate modeling of RC parasitics under CMP variation. As an example, wide bus structures are designed to minimize the unit length delay via simultaneous buffer insertion and buffer sizing with accurate modeling of the parasitic variation due to CMP. We define CMP-aware interconnect design as one which simultaneously considers the buffer insertion and the CMP effects with a specific fill pattern. Compared to the interconnect design under "nominal" RC parasitics without considering either fill insertion or dishing and erosion, we show

*Research at UCLA is partially supported by NSF CAREER award CCR-0401682, SRC grant 1100, a UC MICRO grant sponsored by Analog Devices, Fujitsu Laboratories of America, Intel and LSI Logic, and a Faculty Partner Award by IBM. Dr. Kahng is currently with Blaze DFM, Inc., Sunnyvale, CA 94089. Research at UCSD was partially supported by the MARCO Gigascale Silicon Research Center and the National Science Foundation. Address comments to Dr. He at lhe@ee.ucla.edu or Dr. Kahng at abk@cs.ucsd.edu.

that the CMP-aware interconnect design under the *best* fill pattern (which has the smallest coupling capacitance) has up to 3.3% smaller unit length delay.

The remainder of this paper is organized as follows. Section 2 presents our study of interconnect RC parasitic variations due to either nominally "equivalent" fill patterns or dishing and erosion phenomena. Section 3 discusses our experiments on buffered global interconnect design considering the above CMP-induced RC parasitic variations. We conclude this paper with discussion of our future work in Section 4.

2. MODELING OF CMP VARIATION

The following two types of CMP effects are considered in this paper: dummy fill insertion, and dishing and erosion. Dummy fill insertion improves the uniformity of metal feature density and enhances the planarization that can be obtained by CMP, but may also change the coupling and total capacitance of interconnects. Dishing and erosion phenomena change interconnect cross-sections,[3] and hence may affect interconnect capacitance and resistance.

2.1. Fill Patterns

We assume rectangular, isothetic fill features aligned horizontally and vertically between two adjacent interconnects as shown in Figure 1. In the figure, conductors A and B are *active* interconnects and the metal shapes between them are dummy fills. We assume all dummy fills are implemented as floating metals in the final layout. Each distinct *fill pattern* is specified by: (1) the number of fill rows (M) and columns (N); (2) the series of widths $\{W_i\}_{i=1,\ldots,N}$ and lengths $\{L_j\}_{j=1,\ldots,M}$ of fills; (3) the series of horizontal and vertical spacings, $\{S_{x,i}\}_{i=1,\ldots,N}$ and $\{S_{y,j}\}_{j=1,\ldots,M}$, between fills. We denote a fill pattern by $P(M, N, W_i, L_j, S_{x,i}, S_{y,j})$ for simplicity.

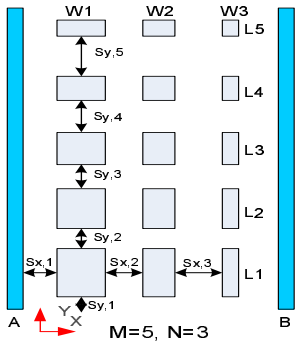


Figure 1. Fill pattern definition.

To specify the amount of fill metal needed in the space and the resulting metal density between two adjacent interconnects, we need the following two definitions.

DEFINITION 2.1. *Local metal density ρ_f – the proportion of the oxide area between two neighboring interconnects that dummy fill metal occupies.*

DEFINITION 2.2. *Effective metal density ρ_{Cu} – the proportion of the area in a planarization window[3] that all metal features (interconnect + dummy fill metal) occupies.*

To achieve CMP planarity and yield optimization, the foundry usually requires an effective metal density ρ_{Cu} to be satisfied in a "fixed-dissection" regime.[2,4] Fixed-dissection fill synthesis typically results in a number of tiles (i.e., square regions of layout, usually several tens of microns on a side) wherein prescribed amounts of fill features are to be inserted to meet individual tile's metal density requirement. This translates to assigning the amount dummy fill feature to the space between interconnects, and such amount is expressed in terms of local metal density ρ_f as defined in Definition 2.1. The inserted fill features subject to at least two foundry-dependent constraints: (1) each fill feature dimension is within the bounds $[\overline{W_l}, \overline{W_u}]$, and (2) the spacing between any two neighboring fill shapes is at least $\overline{S_l}$. A *valid* fill pattern $P(M, N, W_i, L_j, S_{x,i}, S_{y,j})$ between two adjacent interconnects achieves the required fill feature area and satisfies all design rules.

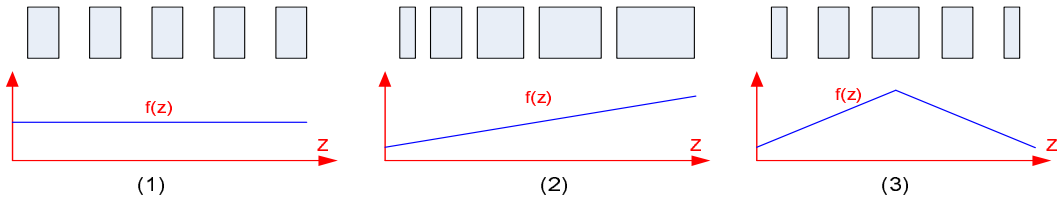


Figure 2. Geometrical interpretation of DCF.

The required fill area A is computed by $\sum_i W_i \cdot \sum_j L_j = W_b \cdot L_b$, with W_b and L_b as the total fill width budget and length budget, respectively. Hence the total horizontal (or vertical) spacing budget is computed by $S_{x,b} = \sum_j S_{x,i} = W_t - W_b$ (or $S_{y,b} = \sum_j S_{y,j} = L_t - L_b$), where W_t is the spacing between active interconnects and L_t is the length of the active interconnects. For choosing M and N , finding a valid fill pattern is equivalent to distributing the budgets of W_b, L_b, $S_{x,b}$, and $S_{y,b}$ among their respective series $\{W_i\}$, $\{L_j\}$, $\{S_{x,i}\}$, and $\{S_{y,j}\}$. To solve this problem, we define a positive *distribution characteristic function* (DCF) $f(z)$, where z is an integer variable that takes the index of the element in the series. The i^{th} element of the series is obtained by $f(i)$ plus the lower bound value as specified by filling rules. For example, the value of the i^{th} width $W_i = f(i) + \overline{W_l}$. If the so-obtained W_i exceeds the upper bound $\overline{W_u}$, we take the upper bound value. Therefore, we can obtain a DRC-clean series under the given budget for a chosen DCF; and different DCFs allow us to systematically explore different fill patterns. To illustrate this point, we take the width series $\{W_i\}$ as an example. If we define $f(z)$ as a constant number, all W_i will have the same value, i.e., all fills have uniform width. If we define $f(z)$ as a linear increasing function, the fills will have a progressively increasing width along the x-axis. If we define $f(z)$ as a triangular function with a convex shape, the center fills will have the largest width, and fills further away from the center will have a progressively decreasing width along the x-axis. Figure 2 shows three DCFs and their corresponding geometrical interpretation. In addition to defining different DCFs, we can also try different DCF combinations for $\{W_i\}$, $\{L_j\}$, $\{S_{x,i}\}$, and $\{S_{y,j}\}$ to obtain more versatile fill patterns.

Figure 3 shows the overall algorithm for searching different valid fill patterns for a given interconnect pair.

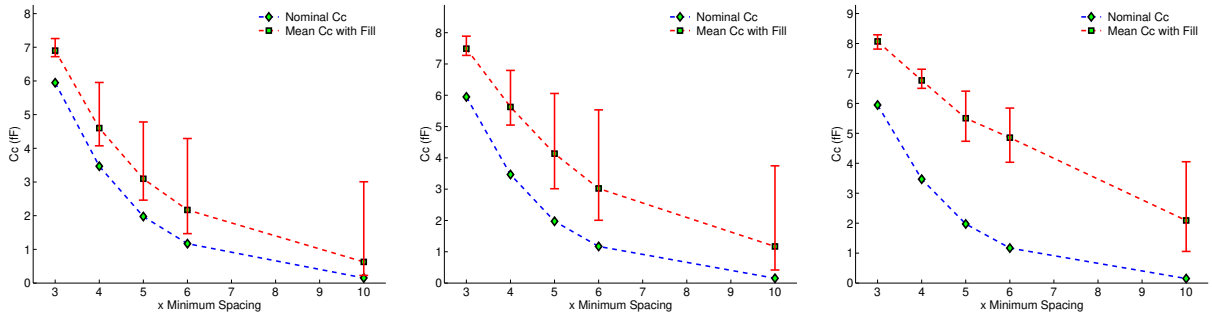
```
Pattern-Explore-Alg(T)
  Input: interconnect pair.
  Output: valid fill patterns in T.

for (all (W_b,L_b), such that W_b · L_b = T.A)
  S_{x,b} = T.W_t - W_b;
  S_{y,b} = T.L_t - L_b;
  for (all valid N,M)
    for (all valid length DCF)
    {L_j} = lengthDCF(T,L_b,N);
      for (all valid width DCF)
      {W_i} = widthDCF(T,W_b,N);
      for (all valid y spacing DCF)
        {S_{y,j}} = spaceYDCF(T,S_{y,b},N);
        for (all valid x spacing DCF)
          {S_{x,j}} = spaceXDCF(T,S_{x,b},M);
          P_v = genFillPattern(M, N, W_i, L_j, S_{x,i}, S_{y,j});
          T.fillList.push(P_v);
```

Figure 3. The overall algorithm for fill pattern exploration.

2.2. Fill Pattern Induced Variation

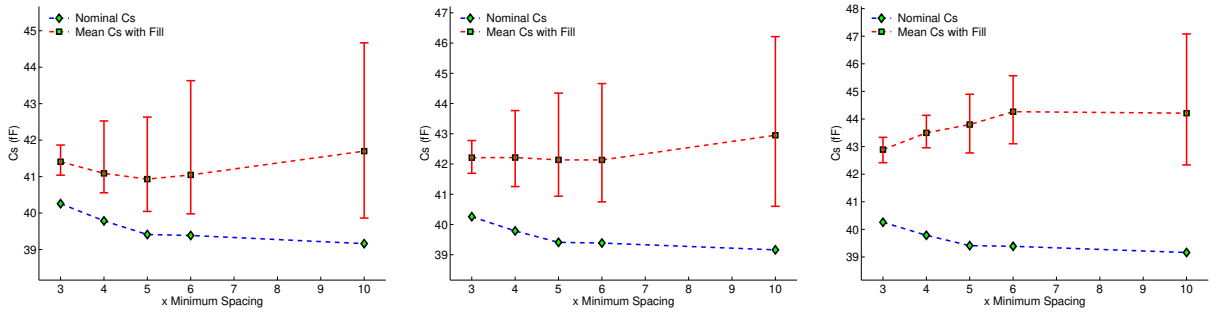
In the following, we examine the impacts of fills and fill patterns on interconnect capacitance. We consider the coupling capacitance (C_c) between active interconnects and total capacitance (C_s) of an individual interconnect. We use QuickCap,[9] a commercial signoff-quality tool, to extract C_c and C_s. The on-chip interconnect is

modeled as a stripline where the interconnect layer is sandwiched between two ground planes. We study global interconnects in the $65nm$ technology node, with conductor dimensions and spacing derived from the ITRS.[8] For each layout, the interconnect width is set to the minimum width while the spacing between two active interconnects varies from $3\times$ to $10\times$ minimum spacing[†]. Interconnect length is $1000\mu m$ for all layouts. For a given layout structure, we first extract the nominal C_c and C_s under the nominal geometries, without considering effects of either fill insertion or dishing and erosion. We then extract C_c and C_s under the same nominal geometry values but with fill insertion.



(a) ρ_f=0.3 (b) ρ_f=0.5 (c) ρ_f=0.7

Figure 4. Distribution of coupling capacitance C_c.



(a) ρ_f=0.3 (b) ρ_f=0.5 (c) ρ_f=0.7

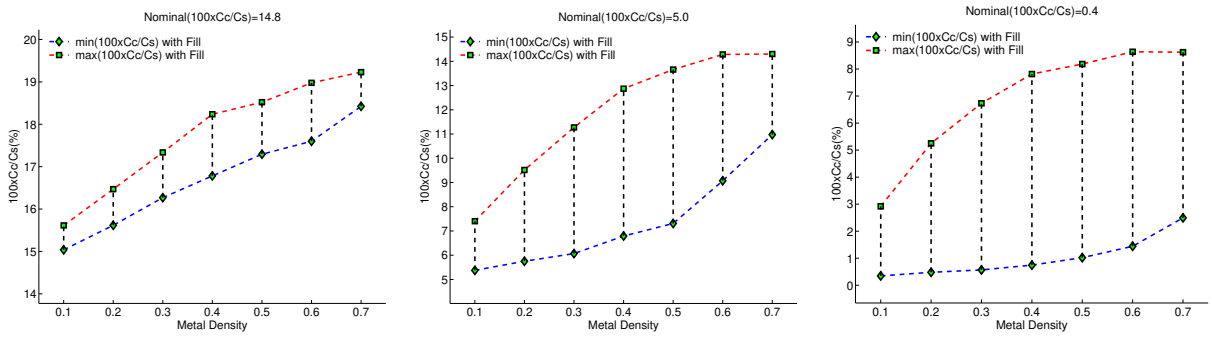
Figure 5. Distribution of total capacitance C_s.

Figures 4 and 5 plot the variation of coupling capacitance C_c and total capacitance C_s, respectively, when fills are inserted to satisfy the required local metal density ρ_f. We examine the cases where $\rho_f = 0.3, 0.5, 0.7$. We vary the spacing between interconnects from $3\times$ to $10\times$ minimum spacing. The curves with diamond symbols are the nominal C_c or C_s without fill insertion. For each interconnect configuration (given the interconnect spacing and local metal density requirement), there are many valid fill patterns and each results in different C_c and C_s. In both Figure 4 and Figure 5, the curves with square symbols represent the mean values of C_c and C_s, respectively. The ranges of C_c and C_s are represented by their respective maximum and minimum values among all the fill patterns that we have explored; these are shown in Figure 4 and 5 as well.

From Figure 4, we observe that different fill patterns indeed result in different coupling capacitances, and that

[†]To have fill insertion between active interconnect without violating design rules, the minimum spacing between active interconnect is $3\times$ minimum spacing rule.

fill insertion always increases the coupling capacitance when compared to the nominal case without considering fill insertion. Furthermore, the gap between the nominal C_c curve and the mean value C_c curve shows the average increase of C_c due to fill insertion. When the local metal density requirement increases, C_c increase since fill insertion also grows. Moreover, for the same local metal density, the relative change of C_c increases as metal spacing increases. For example, when local metal density $\rho_f = 0.5$, the relative C_c change is about 25% on average when the spacing between interconnect is 3× minimum spacing, and is more than tripled when the spacing becomes 6× minimum spacing. Similar observations hold for the total capacitance C_s data in Figure 5, except that the relative change of C_s due to fill insertion is less dramatic than that of C_c. Still, we observe more than 10% relative change of C_s. We conclude that fill insertion significantly increases both C_c and C_s when compared to the nominal case without considering fill insertion; that the relative change is more prominent for C_c than for C_s; and that different fill patterns yield different C_c and C_s values.



(a) 3× minimum spacing (b) 5× minimum spacing (c) 10× minimum spacing

Figure 6. The percentage of C_c over C_s for different local metal density requirement ρ_f.

To study the relative importance of the coupling capacitance variation versus the total capacitance variation due to fill insertion, in Figure 6 we plot the percentage of C_c over C_s with respect to different local metal densities ρ_f (0.1 to 0.7) between active interconnects, whose spacing is chosen as 3×, 5× and 10× minimum spacing, respectively. Because different fill patterns have different C_c and C_s, we only report results for the fill pattern that results in either minimum or maximum C_c over C_s among all fill patterns studied. The gap between the maximum and minimum percentage curves shows the potential variation due to fill insertion. According to Figure 6, we see that fill insertion increases the relative percentage of C_c over C_s compared to the nominal percentage of C_c over C_s without fill insertion as shown in the title of each plot, and that the relative percentage increase becomes larger as the local metal density increases. Moreover, when the metal spacing becomes larger, the relative percentage of C_c over C_s is also increasingly larger compared to the nominal case. On the other hand, because the coupling capacitance decreases as the metal spacing increases, the combined C_c increase is not very significant. In our study, we find that the coupling capacitance is no more than 20% of the total capacitance among all test cases we have studied.

In summary, fill insertion has a very substantial impact on C_c and different fill pattern densities can result in widely varying C_c. Even though variation of C_s is less dramatic, we still see a spread of more than 10% in relation to the nominal C_s. Therefore, to obtain robust designs that will meet requirements (e.g., delay and parametric yield) after insertion of dummy fill, the variation (increase) of both C_c and C_s must be considered by the design flow.

2.3. Dishing and Erosion Induced Variation

Figure 7 illustrates dishing and erosion phenomena due to CMP.[1] Step height is defined as the difference of height between different area on the surface of the wafer. Dishing is a special case of step height that it specifically refers to the difference between the height of the copper in the trench of the metal interconnect and that of

the dielectric in the space surrounding the trenches. Erosion is defined as the difference between the dielectric thickness before CMP and that after CMP. Both dishing and erosion cause loss of metal thickness.

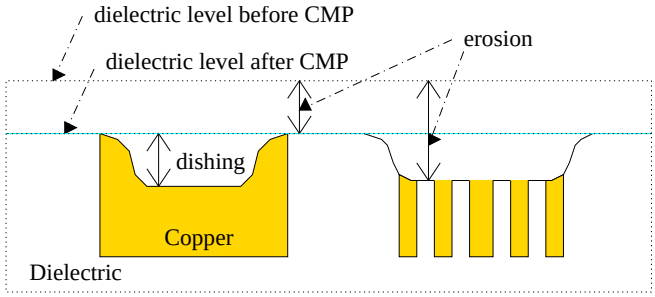


Figure 7. Dishing and Erosion in Copper CMP.

We employ the dishing and erosion model[1] for the multi-step CMP process to calculate post-CMP interconnect geometries. During interconnect formation, trenches are etched on the oxide, followed by barrier deposition on the etched surface to prevent copper diffusion into the oxide. Then a thick layer of copper are deposited on the wafer. CMP removes both the bulk copper above the trenches and the barrier on the area between the trenches. The multi-step model consists of three steps which correspond to three different polishing pads. We assume that Step 1 eliminates all the local step heights and is therefore irrelevant to the modeling of dishing and erosion. We also assume that Step 2 completely removes all the remaining copper so that there is no dishing and erosion at the moment when the polishing pad reaches the barrier. We use the same assumption as in Gbondo-Tugbawa's model[1] that the polishing time of Step 2 after reaching the barrier layer is $20s$ and that of the entire Step 3 is $65s$.

To model barrier/copper simultaneous polishing in Steps 2 and 3 and oxide/copper simultaneous polishing in Step 2, we use

$$d = d_p \cdot e^{\frac{-t}{\tau}} + d_{ss} \cdot \left(1 - e^{\frac{-t}{\tau}}\right) \quad (1)$$

$$E = X_1 \cdot t + X_2 \cdot (d_{ss} - d_p) \cdot \left(e^{\frac{-t}{\tau}} - 1\right) \quad (2)$$

where d_p is the amount of dishing at time $t = 0$, d and E are the amount of dishing and erosion respectively after polishing time t. Note that the amount of E is not counted towards the final amount of erosion as long as the barrier is not cleared. The other terms are defined as

$$d_{ss} = \frac{d_{max} \cdot (r_{Cu} - r_{up}) \cdot (1 - \rho_{Cu})}{r_{Cu} \cdot (1 - \rho_{Cu}) + r_{up} \cdot \rho_{Cu}} \quad (3)$$

$$\tau = \frac{d_{max} \cdot (1 - \rho_{Cu})}{r_{Cu} \cdot (1 - \rho_{Cu}) + r_{up} \cdot \rho_{Cu}} \quad (4)$$

$$X_1 = \frac{r_{Cu} \cdot r_{up}}{r_{Cu} \cdot (1 - \rho_{Cu}) + r_{up} \cdot \rho_{Cu}} \quad (5)$$

$$X_2 = \frac{r_{up} \cdot \rho_{Cu}}{r_{Cu} \cdot (1 - \rho_{Cu}) + r_{up} \cdot \rho_{Cu}} \quad (6)$$

where ρ_{Cu} is the effective metal density, r_{Cu} is the blanket copper removal rate, r_{up} is the effective removal rate of the "up" area (i.e., barrier in barrier/copper polishing and oxide in oxide/copper polishing). r_{up} is obtained by scaling the blanket removal rate by the factor Ψ to account for the edge rounding effect. Ψ is given by

$$\Psi = C \cdot e^{\frac{-s}{s_C}} + 1 \quad (7)$$

with process-dependent constants C and s_C. d_{max} is also a layout feature-dependent parameter and is given by

$$d_{max} = B \cdot \left(\frac{w}{w_0}\right)^{\alpha} \cdot \left(\frac{s}{s_0}\right)^{\beta} \quad (8)$$

where w and s are the wire width and the wire spacing, B, α and β are process-dependent constants, and $w_0 = s_0 = 1\mu m$. All process-dependent constants are taken from the original model.[1]

The model for oxide/copper simultaneous polishing in Step 3 is much more complicated since the removal rate of oxide (the up-area) is larger than the removal rate of copper (the down-area), which leads to more boundary conditions. The amount of dishing and erosion is given by

$$d = \begin{cases} d_p - \frac{r_{ox}}{1-\rho_{Cu}} \cdot t & 0 \le t < t_{cr}, d_p > d_{cr} \\ d_{cr} \cdot e^{\frac{-t}{\tau_3}} + D_{ss} \cdot \left(1 - e^{\frac{-t}{\tau_3}}\right) & t \ge t_{cr}, d_p > d_{cr} \\ d_p \cdot e^{\frac{-t}{\tau_3}} + D_{ss} \cdot \left(1 - e^{\frac{-t}{\tau_3}}\right) & t \ge 0, d_p \le d_{cr} \end{cases} \quad (9)$$

$$E = \begin{cases} \frac{r_{ox}}{1-\rho_{Cu}} \cdot t & 0 \le t < t_{cr}, d_p > d_{cr} \\ \frac{r_{ox}}{1-\rho_{Cu}} \cdot t_{cr} + X_3 \cdot t + Z_3 \cdot \left(1 - e^{\frac{-t}{\tau_3}}\right) & t \ge t_{cr}, d_p > d_{cr} \\ X_3 \cdot t + Y_3 \cdot \left(1 - e^{\frac{-t}{\tau_3}}\right) & t \ge 0, d_p \le d_{cr} \end{cases} \quad (10)$$

where d_p is the amount of dishing at $t = 0$, ρ_{Cu} is the effective metal density, r_{Cu} is the blanket removal rate of copper, and r_{ox} is the effective removal rate of oxide which is again obtained by scaling the blanket removal rate with Ψ as defined in Equation 7. d_{cr} is the critical dishing and is defined exactly as in Equation (8) for d_{max}. The other terms are defined as

$$t_{cr} = \frac{(d_p - d_{cr}) \cdot (1 - \rho_{Cu})}{r_{ox}} \quad (11)$$

$$D_{ss} = \frac{d_{max}^3 \cdot (r_{Cu} - r_{ox}) \cdot \rho_{Cu}}{r_{Cu} \cdot (1 - \rho_{Cu}) + r_{ox} \cdot \rho_{Cu}} \quad (12)$$

$$\tau_3 = \frac{d_{max}^3 \cdot \rho_{Cu}}{r_Cu \cdot (1 - \rho_{Cu}) + r_{ox} \cdot \rho_{Cu}} \quad (13)$$

$$X_3 = r_{ox} + \frac{r_{ox} \cdot D_{ss}}{d_{max}^3} \quad (14)$$

$$Z_3 = \frac{r_{ox} \cdot \tau_3 \cdot (d_p - D_{ss})}{d_{max}^3} \quad (15)$$

$$Y_3 = \frac{r_{ox} \cdot \tau_3 \cdot (d_p - D_{ss})}{d_{max}^3} \quad (16)$$

$$d_{max}^3 = d_{cr} \cdot \left(\frac{s}{w}\right) \quad (17)$$

Table 1 shows the RC parasitics for a $1000\mu m$ long global interconnect bus structure under the $65nm$ technology node. R_0 is the resistance computed from the geometry values obtained from ITRS specifications, i.e., dishing and erosion effects are not taken into account. R_f is the resistance after "best" fill insertion which fulfills 50% metal density requirement (i.e. $\rho_{Cu} = 0.5$). Based on this, we include the metal loss due to dishing and erosion when computing R_f. From Table 1, we can see that resistance variation due to dishing and erosion is significant, and that resistance is always increasing, potentially by more than 30%. As width increases, the resistance variation becomes increasingly severe. For example, when conductor width increases from $0.24\mu m$ to $4.75\mu m$, the resistance variation increases from 29% to 32%.

All capacitance values in Table 1 are extracted using QuickCap.[9] $C_{c,0}$ and $C_{s,0}$ are the coupling capacitance and total capacitance without considering fill insertion or dishing and erosion effects. $C_{c,1}$ and $C_{s,1}$ are the coupling capacitance and total capacitance for the same assumed structure as in Section 2.2, taking geometry variations due to dishing and erosion effects (but no fill insertion) into account. Finally, $C_{c,f}$ and $C_{s,f}$ are the coupling capacitance and total capacitance when effects due to dummy fill, dishing and erosion are all taken into consideration. From Table 1, we observe that dishing and erosion alone have marginal impact on capacitance for most design contexts. In light of these results, we do not consider dishing and erosion effects on capacitance.

Width	Space	wo/CMP	w/CMP	wo/CMP		Dishing/Erosion		Fill+Dishing/Erosion	
μm	μm	$R_0(\Omega)$	$R_f(\Omega)$	$C_{c,0}$	$C_{s,0}$	$C_{c,1}$ (%)	$C_{s,1}$ (%)	$C_{c,f}$ (%)	$C_{s,f}$ (%)
0.24	0.95	186	239 (28.7)	6.99	79.46	6.80 (-2.63)	79.20 (-0.33)	9.30 (33.06)	79.38 (-0.11)
2.61	0.95	16.9	22.1 (30.6)	7.24	268.56	6.96 (-3.78)	268.05 (-0.19)	9.14 (26.33)	264.92 (-1.35)
4.75	0.95	9.29	12.3 (31.4)	7.01	433.29	7.22 (2.97)	436.25 (0.68)	8.87 (26.51)	432.29 (-0.23)
0.24	1.43	186	239 (28.8)	2.32	78.82	2.38 (2.54)	78.72 (-0.13)	5.63 (142.71)	80.31 (1.88)
2.61	1.43	16.9	22.1 (30.9)	2.41	265.79	2.31 (-4.35)	265.01 (-0.29)	5.84 (141.81)	266.76 (0.36)
4.75	1.43	9.29	12.2 (31.7)	2.17	437.34	2.34 (8.11)	431.37 (-1.36)	5.39 (148.81)	434.32 (-0.69)

Table 1. RC parasitic comparison for $65nm$ global interconnects.

2.4. CMP-aware Table-based RC Model

Based upon our study of CMP-induced RC parasitic variations, we tabulate the extracted capacitance in a table indexed by active interconnect width, spacing and local metal density. As different fill patterns under the same pattern density result in different capacitance values as shown in Section 2.2, the capacitance table only saves the capacitance under the *best* (*worst*) fill pattern, which gives the minimum (maximum) C_c among all patterns. We use local metal density as index to the table and it represents the amount of fill features required in the wire spacing. We only refer to either the *best* or the *worst* fill pattern. We use formulae of Section 2.3 to compute the resistance under dishing and erosion effects. In the following, we denote the resulting RC models as *CMP-aware* RC parasitic models. In contrast, interconnect parasitics without consideration of fill pattern insertion, dishing or erosion effects are called *CMP-oblivious* RC models.

3. CMP IMPACT ON BUFFERED INTERCONNECT PERFORMANCE

To understand the impact of CMP on interconnect performance, we design a wide interconnect bus using detailed SPICE simulations under the CMP-oblivious and the CMP-aware RC parasitic models, respectively. For simplicity, we assume the bus structure has uniform wire width w and spacing s, with w equals to s. The wide bus structure is modeled by four parallel, capacitively-coupled wires as shown in Figure 8, where V_+ and V_- are two opposite sets of input ramp waveforms. This circuit model results in the minimum number of elements yet still captures the necessary elements which cause the "worst" case coupling effects between interconnects. We set the number of segments n to 10 in our experiment. Buffer size S_{buf} and buffer insertion length L_{buf} are the variables subjected to optimization. The objective is to minimize the unit length delay D_L through simultaneous buffer insertion and buffer sizing. D_L is calculated by $(t_2^{0.5} - t_1^{0.5})/L_{buf}$, where $V_{out}(t_2^{0.5}) = V_{out}(t_1^{0.5}) = 0.5 \cdot V_{dd}$ for one switch at the input.

ITRS[8] $65nm$ global interconnect parameters and BSIM 4 device models[10] are assumed in this study. Given the wire width w in terms of integer multiples of the minimum width (which is $0.2375\mu m$ in our experiment), the local metal density ρ_f and the effective metal density ρ_{Cu}, we search for the minimum unit length delay D_L by varying the buffer size S_{buf} and the length L_{buf}. The searching granularity of optimal L_{buf} and S_{buf} are to the accuracy of $100\mu m$ and $10\times$ of the minimum buffer size, respectively.

Table 2 shows experimental results for minimizing the unit length delay D_L under both the CMP-oblivious RC model and the CMP-aware RC model with the best (minimum coupling capacitance) fill patterns. For fair comparison, the final unit length delay D_L under the CMP-oblivious RC model is computed after the best fill insertion has been performed. Column 1 is the metal width; Columns 2 and 3 are the local metal density ρ_f and the effective metal density ρ_{Cu}, respectively. Columns 4 and 5 (or Columns 7 and 8) are the optimal buffer insertion length L_{buf} and buffer sizing S_{buf}, respectively.

We report the unit length delay D_L through SPICE simulation in Columns 6 and 10 for both CMP-oblivious and CMP-aware designs, respectively. By comparing Columns 6 and 10, we observe that CMP-aware designs always result in smaller unit length delay than CMP-oblivious designs, and the relative improvement can be up to 3.3% (see Column 11). To measure the buffer area penalty in achieving the so-obtained unit length delay, we normalize the buffer area with respect to interconnect length, i.e., $S_{buf,L} = S_{buf}/L_{buf}$. Column 9 reports the relative increase of $S_{buf,L}$ for the CMP-aware design compared to the CMP-oblivious design. According to Column 9, we note that under most cases, a CMP-aware design tends to use more buffer area compared to a CMP-oblivious design, and the relative increase of the normalized buffer area is no more than 22%.

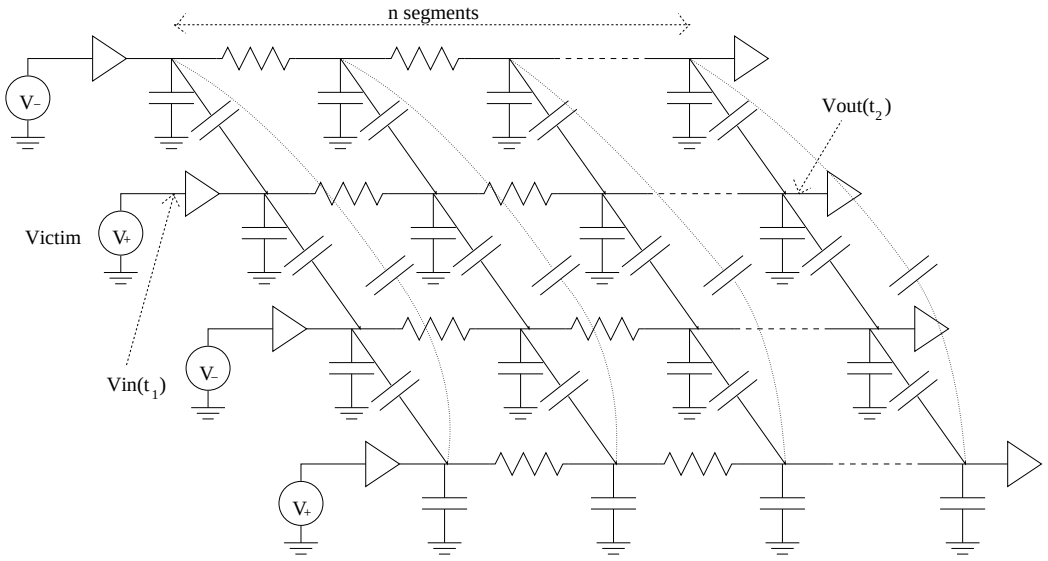


Figure 8. SPICE model of a wide parallel bus.

For fixed wire width w and ρ_f, the improvement of unit length delay for the CMP-aware design over the CMP-oblivious design decreases as ρ_{Cu} increases. This is because of the diminishing amount of erosion that in turn causes resistance to increase. For example, when $w = 3\times$ and $\rho_f = 0.5$, the relative reduction of unit length delay for CMP-aware designs over CMP-oblivious designs decreases from 3.3% to 2.2% when the effective metal density ρ_{Cu} increases from 0.3 to 0.7. We further note that for fixed w and ρ_{Cu}, the improvement of unit length delay for the CMP-aware design over the CMP-oblivious design increases as ρ_f increases. This increasing trend is mainly due to the increase in coupling capacitance between interconnects when ρ_f becomes larger. For example, when $w = 4\times$ and $\rho_{Cu} = 0.5$, the relative reduction of unit length delay for CMP-aware designs over CMP-oblivious designs increases from 1% to 1.7% when the effective metal density ρ_f increases from 0.1 to 0.9.

Although the above two trends are generally observed throughout the table, there exists a few exceptions, which are mainly due to the discretization errors in the search of optimal buffer insertion length L_{buf} and buffer size S_{buf}. Completely eliminating such errors is infeasible as it would take enormous simulation time.

Table 3 shows another set of experimental results for minimizing the unit length delay D_L under both CMP-oblivious RC model and CMP-aware RC model with the worst fill patterns (maximum coupling capacitance). Columns 4 and 5 (or Columns 8 and 9) are the optimal buffer insertion length L_{buf} and buffer sizing S_{buf}, respectively. Only designs with $w = 4\times$ are shown for brevity. We report the unit length delay D_L through SPICE simulation in Columns 6 and 11 for both CMP-oblivious and CMP-aware designs, respectively, under the worst fill pattern. Similarly, Columns 7 and 13 are the unit length delay D_L under the the best fill insertion for both CMP-oblivious and CMP-aware designs, respectively, under the best fill pattern. By comparing Columns 6 and 11, we observe that CMP-aware designs under the worst fill insertion consistently result in smaller unit length delay than CMP-oblivious designs, and the relative improvement can be up to 4.1% (see Column 12). By comparing Columns 7 and 13, we observe that CMP-aware design assuming the worst fill pattern are not necessarily better than CMP-oblivious designs when in fact other fill patterns are inserted, as the former sometimes result in a higher unit length delay (see positive percentages in Column 14). Therefore, there exists no single design that is CMP-variation optimized; designers must design the interconnect according to the specific dummy fill pattern in order to attain optimality.

4. CONCLUSION

This paper studies CMP-induced interconnect parasitic variations due to (1) different fill patterns that are nominally "equivalent" with respect to foundry rules; and (2) dishing and erosion of conductors and dielectric similar to those predicted by ITRS.[8] We show that the variation of coupling and total capacitance can be

			CMP-oblivious			CMP-aware w/the best fill				
1	2	3	4	5	6	7	8	9	10	11
w	ρ_f	ρ_{Cu}	L_{buf} (μm)	S_{buf}	D_L ($fs/\mu m$)	L_{buf} (μm)	S_{buf}	$\Delta S_{buf,L}$	D_L ($fs/\mu m$)	ΔD_L
3	0.1	0.3	2137	310	21.2	1862	310	14.8%	20.6	-2.8%
3	0.1	0.5	2137	310	20.4	1962	310	8.9%	19.8	-2.5%
3	0.1	0.7	2137	310	19.9	1962	310	8.9%	19.5	-1.6%
3	0.5	0.3	2137	310	21.6	1862	310	14.8%	20.8	-3.3%
3	0.5	0.5	2137	310	20.7	1962	310	8.9%	20.2	-2.4%
3	0.5	0.7	2137	310	20.2	1962	310	8.9%	19.8	-2.2%
3	0.9	0.3	2137	310	21.3	1862	310	14.8%	21.1	-0.7%
3	0.9	0.5	2137	310	21.0	1862	310	14.8%	20.5	-2.4%
3	0.9	0.7	2137	310	20.5	1962	310	8.9%	20.0	-2.5%
4	0.1	0.3	2637	350	18.8	2137	310	9.3%	18.4	-2.1%
4	0.1	0.5	2637	350	18.1	2137	310	9.3%	17.9	-1.0%
4	0.1	0.7	2637	350	17.7	2237	310	4.4%	17.4	-1.4%
4	0.5	0.3	2637	350	19.3	2062	310	13.3%	18.8	-2.6%
4	0.5	0.5	2637	350	18.4	2137	310	9.3%	18.2	-1.4%
4	0.5	0.7	2637	350	18.0	2437	320	-1.1%	17.9	-1.0%
4	0.9	0.3	2637	350	19.9	2262	300	-0.1%	19.2	-3.2%
4	0.9	0.5	2637	350	19.0	2337	320	3.2%	18.7	-1.7%
4	0.9	0.7	2637	350	18.5	2137	340	19.9%	18.3	-1.3%
5	0.1	0.3	2812	400	17.7	2237	350	10.0%	17.3	-2.3%
5	0.1	0.5	2812	400	17.0	2337	350	5.3%	16.8	-1.5%
5	0.1	0.7	2812	400	16.7	2637	350	-6.7%	16.5	-1.0%
5	0.5	0.3	2812	400	17.9	2237	350	10.0%	17.5	-2.7%
5	0.5	0.5	2812	400	17.2	2537	350	-3.0%	17.0	-1.2%
5	0.5	0.7	2812	400	16.8	2337	360	8.3%	16.6	-1.2%
5	0.9	0.3	2812	400	18.8	2137	370	21.7%	18.2	-3.3%
5	0.9	0.5	2812	400	17.9	2237	370	16.3%	17.6	-1.8%
5	0.9	0.7	2812	400	17.4	2537	370	2.5%	17.3	-0.9%

Table 2. CMP-oblivious vs CMP-aware w/the best fill optimization.

more than 300% and 10%, respectively, between two adjacent wires due to "pattern-dependent" fill insertion. Moreover, the variation of wire resistance due to dishing and erosion can be over 30%, but have limited impact on interconnect capacitance.

This paper also explores possible improvement in the interconnect performance through accurate modeling of RC parasitics under CMP variation. Wide bus structures are designed to minimize the unit length delay via simultaneous buffer insertion and buffer sizing with accurate modeling of the parasitic variation due to CMP. We define CMP-aware interconnect as one which simultaneously considers the buffer insertion and the CMP effects with a specific fill pattern. Compared to the interconnect design under "nominal" RC parasitics without considering either fill insertion or dishing and erosion, we show that the CMP-aware interconnect design under the *best* fill pattern (which has the smallest coupling capacitance) has up to 3.3% smaller unit length delay. We show that the improvement of CMP-aware designs over the design without CMP consideration increases with increasing amount of dummy fill between interconnects and decreasing effective metal density (including dummy fill). We also demonstrate that CMP-aware design under any fill pattern differing from that inserted do not necessarily lead to a better design than those which do not consider CMP at all; optimization must be performed simultaneously with buffer insertion and CMP-awareness with the fill pattern-to-insert to achieve optimality.

We have demonstrated in another work[11] that the CMP effect has notable impact on the buffer insertion problem. Based on this, we have proposed an algorithm which solves the buffer insertion problem optimally under

			CMP-oblivious				CMP-aware w/the worst fill						
1	2	3	4	5	6	7	8	9	10	11	12	13	14
w	ρ_f	ρ_{Cu}	L_{buf} (μm)	S_{buf}	worst fill D_L ($fs/\mu m$)	best fill D_L ($fs/\mu m$)	L_{buf} (μm)	S_{buf}	$\Delta S_{buf,L}$	worst fill D_L ($fs/\mu m$)	worst fill ΔD_L	best fill D_L ($fs/\mu m$)	best fill ΔD_L
4	0.1	0.3	2637	350	20.6	18.8	1962	320	22.9%	19.8	-4.1%	18.7	-0.4%
4	0.1	0.5	2637	350	19.7	18.1	2262	330	9.9%	19.2	-2.7%	18.0	-0.5%
4	0.1	0.7	2637	350	19.2	17.7	2062	340	24.2%	18.8	-2.1%	17.8	0.7%
4	0.5	0.3	2637	350	21.0	19.3	2162	330	15.0%	20.4	-2.7%	19.0	-1.5%
4	0.5	0.5	2637	350	20.5	18.4	1962	340	30.6%	19.8	-3.5%	18.6	0.8%
4	0.5	0.7	2637	350	20.0	18.0	2262	340	13.2%	19.4	-2.7%	18.0	-0.3%
4	0.9	0.3	2637	350	21.1	19.9	2162	330	15.0%	20.6	-2.6%	19.6	-1.5%
4	0.9	0.5	2637	350	20.6	19.0	1962	340	30.6%	19.8	-3.7%	18.9	-0.4%
4	0.9	0.7	2637	350	20.1	18.5	2262	340	13.2%	19.5	-2.8%	18.6	0.5%

Table 3. CMP-oblivious vs CMP-aware w/the worst fill optimization.

the CMP variation. In the future we intend to study the impact from more sources of variations on interconnect performance and design. We will focus on (1) studying the performance impact of these new variation sources; (2) understanding the potential room for improvement by considering these new variation sources during design; and (3) developing algorithm to incorporate variation models in the interconnect optimization flow.

REFERENCES

1. T. E. Gbondo-Tugbawa, *Chip-Scale Modeling of Pattern Dependencies in Copper Chemical Mechanical Polishing Process*. PhD thesis, Massachusetts Institute of Technology, 2002.
2. Y. Chen, P. Gupta, and A. B. Kahng, "Performance-impact limited area fill synthesis," in *Proc. DAC*, Jun 2003.
3. T. Tugbawa, T. Park, D. Boning, T. Pan, P. Li, S. Hymes, T. Brown, and L. Camilletti, "A mathematical model of pattern dependencies in cu cmp processes," in *Proc. CMP Symposium, Electrochemical Society Meeting*, Oct 1999.
4. P. Gupta and A. B. Kahng, "Manufacturing-aware physical design," in *Proc. ICCAD*, Oct 2003.
5. W. Grobman, M. Thompson, R. Wang, C. Yuan, R. Tian, and E. Demircan, "Reticle enhancement technology: Implications and challenges for physical design," in *DAC*, Jun 2001.
6. R. Chang, *Integrated CMP Metrology and Modeling with Respect to Circuit Performance*. PhD thesis, University of California, Berkeley, 2004.
7. V. Mehrotra, S. Sam, D. Boning, A. Chandrakasan, R. Vallishayee, and S. Nassif, "A methodology for modeling the effects of systematic within-die interconnect and device variation on circuit performance," in *Proc. DAC*, Jun 2000.
8. Semiconductor Industry Association, *International Technology Roadmap for Semiconductors*, 2003.
9. "Quickcap User Manual," in *http://www.magma-da.com/*.
10. "Berkeley predictive technology model," in *http://www-device.eecs.berkeley.edu/~ptm*.
11. L. He, A. B. Kahng, K. Tam, and J. Xiong, "Simultaneous buffer insertion and wire sizing considering systematic cmp variation and random leff variation," in *Proc. ISPD*, Apr 2005.