# Recent Topics in CMP-Related IC Design for Manufacturing


## ABSTRACT

CMP is a planarization technique for multilevel VLSI metallization processes. CMP fills are inserted as a design for manufacturability (DFM) methodology to improve pattern-dependent post-CMP topography, i.e., to improve interconnect planarity. Design-driven fill synthesis seeks to optimize CMP fill with respect to objectives beyond mere density uniformity. Design-driven fill synthesis minimizes the impact of CMP fill on function-driven performance and parametric yield metrics, while satisfying manufacturing-driven density criteria. This paper reviews CMP fill techniques as a DFM process, then describes a power-aware methodology wherein CMP fill synthesis is integrated within a holistic design- and manufacturing-driven flow. Experimental validation with production 65nm CMP models shows that this proposed methodology obtains power reductions on the design side, while satisfying density requirements imposed by manufacturing.


## INTRODUCTION

In the past decade, *chemical-mechanical polishing* (CMP) has emerged as the predominant planarization technique for shallow trench isolation (STI) and back-end-of-the-line (BEOL) metallization. Failure to meet surface topography variation constraints leads to degradation of transistor characteristics in the case of STI, and electrical shorts or increased wire resistance in the case of BEOL interconnects, all of which are detrimental to circuit yield. As technology scaling advances into 45nm and beyond, variations due to lithography and CMP processes have to be accurately considered in the design flow. As an example, on-chip variation (OCV) from the interconnect thickness variation due to CMP becomes relatively larger and needs to be taken into consideration in the post-layout RC extraction and timing flow [10].

Design for manufacturability (DFM) targets design analysis, modeling, optimization and automation to enable manufacture of working chips, and improve chip performance and reliability, in the face of mounting challenges (variability, leakage, etc.) in semiconductor manufacturing. As the number of metal layers increases and line widths shrink, tolerance for topographical imperfections decreases. This is due to tight depth-of-focus variation requirements and high sensitivity of resistance to metal thickness. Despite improvements in CMP technology, layout pattern sensitivities are significant, causing certain regions to have higher topographies than others due to differences in underlying densities. One DFM technique that improves surface planarity in today's advanced technologies is the insertion of CMP *fill* shapes, which are non-functional features that do not directly contribute to the logic implementation and can either be grounded or left floating (Figure 3). Fill shapes impact electrical performance of the designs and hence should be optimized for circuit performance in addition to layer topography uniformity. Efficient characterization, modeling and optimization of CMP fills are necessary to help reduce design pessimism and improve suboptimal design performance.

The remainder of this paper is organized as follows. In the next section, we present a brief overview of CMP modeling, including for copper dual-damascene / low-k processes. We

then review the state of the art in density control through fill insertion. Next, we present an overview of design-driven CMP fill synthesis. Specific examples of design-driven fill synthesis, namely, timing- and power-driven fill optimization methodologies are presented next, followed by a summary of CMP fill synthesis design flow issues and the paper's conclusions.

## CMP MODELING

CMP is the technique of choice to satisfy local and global planarity constraints imposed by today's advanced photolithography methods, as well as interconnect performance margins. However, CMP heavily depends on the underlying pattern density and therefore, various degrees of metal dishing and dielectric erosion happen at different metal densities (Figure 1) [10]. To reduce the post-CMP thickness variation dummy fill shapes (CMP fill) are routinely inserted into the layout. Due to complex nature of CMP process (pad, interaction between slurry and metal/oxide, multi-level effects, etc.), a physical model is required to simulate CMP process on a full chip level. Such a model needs to address the CMP process at three scales of particle-, feature- and die-/wafer-level.

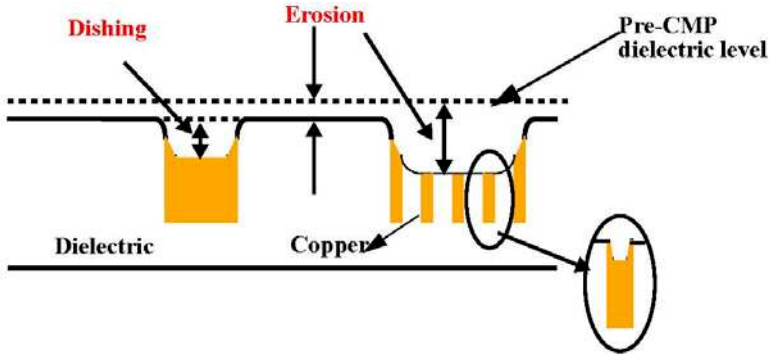



**Figure 1.** Metal dishing and dielectric erosion [2].

Particle-scale modeling addresses the roles and interactions of slurry particles, slurry chemicals, polishing pads and wafer materials. The two most important issues at the particle scale are the material removal rate (MRR) and the surface quality (surface roughness). The material removal rate determines the production rate of the process. Surface qualities determine the yield of the process. A widely known model is the experimental Preston's equation which was initially introduced for glass polishing [11]. Models at the feature and die scales are needed to address the topography evolution of IC chips as a function of pattern density, line width, and polish time. The MIT semi-empirical model is based on a framework first proposed by Stine *et al.* [5]; it characterizes and models the polishing behavior of inter-layer dielectric (ILD) as a function of structure pattern density. Experimental data indicates that the effective pattern density, which is defined as the ratio of the raised area to the total area for a window of a given size and shape, is the dominant factor for within-die nonuniformity, while the structure area, pitch, and perimeter/area play only minor roles. The effect of pattern density on within-die nonuniformity is due to pressure nonuniformity. In a high-density area, where the oxide area contacted by the pad is larger, the effective pressure is lower and hence, the material removal rate is smaller. Analytical equations to address this are developed by Stine *et al.* [6] based on Preston's equation of material removal rate. While the semi-empirical models of Stine *et al.* [6] and Ouma *et al.* [7] are the most successful at feature and die scales, others have tried to develop

physical models to explain the topography evolution. Ouyang *et al.* [8] developed equations based on fluid mechanics to predict pressure at any given location on the die. Preston's equation is then used to evaluate the topography evolution [11]. Another fluid mechanics based feature-scale topography model was developed by Yao *et al.* [9].

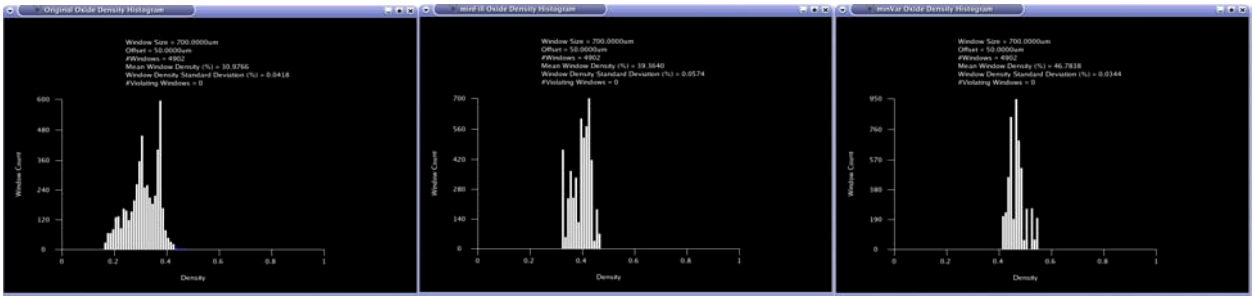
With dual-damascene copper interconnects, the low mechanical strength of low-k films can cause defects when using conventional CMP. An alternative solution to this problem is to use *stress-free polishing* (SFP). Stress-free polishing is a low viscosity fluid contact alternative to CMP that enables the removal of copper overburden without exposing barrier or dielectric films to mechanical stress. SFP applies a voltage across the wafer with an electrolyte held as anode. The process removes copper through control of wafer rotation, recirculating electrolyte flow, and voltage and current without the use of slurry. Since the SFP process is inherently isotropic with no planarizing capability, a partial CMP step will be added at the end to achieve surface planarity [12]. Hence, the above-mentioned models can be applied to predict the resultant topography after the partial CMP process.

## LAYOUT DENSITY CONTROL THROUGH FILL INSERTION

Pattern-density uniformity is achieved by insertion ("filling") or, to a lesser extent, partial deletion ("slotting") of features in the layout. The foundry's design rules specify upper and lower bounds on feature density for $w \times w$ windows of the layout. In practice, these bounds are verified and enforced only for overlapping windows of size $w \times w$ within a so-called *fixed dissection* of the layout. More precisely, windows of size $w \times w$ are stepped across an $n \times n$ layout region, with step size of $w/r$; this induces a dissection of the layout region into *tiles* of size $(w/r) \times (w/r)$, and density bounds are verified in all windows of $r \times r$ tiles having top-left corners at locations $(i \cdot (w/r), j \cdot (w/r))$, for $i, j = 0, 1, \ldots, r((n/w)-1)$, as shown in Figure 3a. Kahng *et al.* [16] developed an optimal solution approach for fill synthesis, using linear programming (LP). The LP approach finds the optimum total area of fill shapes to be inserted into each tile, such that the difference between maximum and minimum post-fill window densities in the layout is minimized. In a follow-up work, Tian *et al.* [17] developed a minimum-fill LP formulation, incorporating an elliptical weighting function (from the planarization model of Ouma [7]) which has better correspondence to the pad deformation. The works of [16] and [17] contrast (1) a 'manufacturer's objective', which is to minimize the post-fill density variation across the layout to improve manufacturing uniformity, versus (2) a 'designer's objective', which is to minimize the total amount of fill inserted into the layout so as to minimize added capacitance and coupling, timing, noise and power impacts on the design. Manufacturability goals also dictate the creation of "smooth" fill, e.g., through application of the LP approach with explicit upper bounds on the density difference between any pair of adjacent windows. Table 1 [19] compares relevant metrics of minimum-variation, minimum-fill, and maximum-smoothness solutions for a challenging image sensor chip. Figure 2 additionally shows layout density histograms for the original layout, a minimum-fill solution, and a minimum-variation solution.

**Table 1.** Comparison of different fill insertion formulations and post-fill layout metrics.

| | Min. D | Max. D | Delta D | # fill | Avg. Smoothness |
|---|---|---|---|---|---|
| Original Solution | 0.1652 | 0.4717 | 0.3065 | - | 0.0508 |
| minVar | 0.4153 | 0.5448 | 0.1295 | 784,968 | 0.0234 |
| minFill | 0.3234 | 0.4717 | 0.1483 | 416,773 | 0.0317 |
| maxSmoothness | 0.3945 | 0.5243 | 0.1298 | 711,429 | 0.0174 |



a) Original Density Histogram (ΔD = 31%)

(b) Min-Fill Density Histogram (ΔD = 15%)

(c) Min-Variation Density Histogram (ΔD = 13%)

**Figure 2.** Post-fill layout density histograms for original, minimum-fill, and minimum-variation solutions.

## DESIGN-DRIVEN FILL SYNTHESIS

As the industry moves into the 45*nm* node and beyond, traditional fill synthesis methods have reached their limits of usefulness. One indication of this is the emergence of "recommended rules" in design rule manuals, e.g., "it is better to have a small difference between the density values of adjacent windows" (a smoothness objective), or "it is better to maximize the overlap of fill shapes on adjacent layers to enable dummy via insertion" (since dummy vias enhance low-k dielectric material stability). The impact of fill synthesis on parasitics and timing also continues to be a key concern for the designer. It is increasingly difficult for a design rule check (DRC) platform to obtain an *optimal, design-driven* fill synthesis solution that meets all basic CMP design rules and as many recommended rules as possible, while minimizing impact on timing, signal integrity and power. With this in mind, we now sketch the capabilities of more sophisticated, "design-driven" fill syntheses. Such techniques can potentially reduce engineering effort while enhancing manufacturability through increased process and design latitudes. Design-driven fill embodies such features as the following [2].

- *Global Optimization:* Typical foundry rules specify upper and lower bounds on density in windows of the layout region. These windows are "stepped" to improve density uniformity, e.g., a physical verification run deck might verify density in 200µm × 200µm windows, stepped in increments of 50µm (Figure 3(a)). In this scenario, windows have $w$ = 200µm, each window is divided into $r$ = 4 "steps", and the step distance of $w/r$ = 50µm defines a *tile* size. A large ASIC 20mm on a side would have 160000 tiles on each layer, and a 10-layer

metal process implies 1.6 million tiles in the chip. Modern design-driven fill should be able to compute optimal amounts of fill to be added into each tile, simultaneously for all tiles on all layers of the chip. This is enabled by, e.g., highly scalable nonlinear optimization technology.

- *Model-Based Fill Synthesis:* Rule-based fill synthesis is based on concepts such as density or *keep-off distance* rules (i.e., no fill shape is inserted less than this distance away from any layout feature, as shown in Figure 3(b)), which may be applied to wiring segments belonging to nets with less than given threshold amounts of timing slack. Model-based fill synthesis, on the other hand, would use CMP models to, e.g., identify regions where maintaining planarity and hence wire cross-section is important (e.g., next to timing-critical segments and below critical segments). The model-based approach has implicit tight coupling to extraction and timing engines, particularly to model the impact of fill on post-polish wire thickness and RC values (Figure 4).
- *Timing-Driven Fill Synthesis:* One of the largest concerns in fill synthesis, apart from meeting the CMP design rules, is the impact of fill insertion on capacitances of signal nets. Excessive increase in wire capacitance can cause a net to violate setup timing constraint (or, on the other hand, added capacitance can increase hold timing slack). A large value for keep-off distance reduces the setup slack danger but reduces available area for fill insertion and therefore can make it impossible to meet minimum density constraints. With timing-driven fill, the impact of inserting fills on timing is continually assessed, and the minimum keep-off distance for each net to meet the setup time constraint can be computed to avoid a wastefully large "one size fits all" keep-off distance. In a more advanced, timing-driven fill flow, the impact of fill insertion on both wafer topography and timing would be analyzed and optimized concurrently. One additional advantage of timing-driven fill is that it can improve the hold timing slack of a net by deliberately and selectively introducing capacitance to that net.
- *Power-Driven Fill Synthesis:* Power consumption is also a very important consideration in today's integrated circuit (IC) designs. As CMP fill changes interconnect capacitance it significantly affects the dynamic power consumption of these nets. Power-driven fill synthesis consists of identifying the power-critical nets – i.e., those with high switching activity and large capacitance – and optimizing the fill shapes around them to minimize power consumption [3].

In the next section, we present our recently proposed timing- and power-driven CMP fill insertion methodologies as examples of design-driven fill synthesis. Experimental assessments in a production 65nm technology suggest that using our proposed methodologies can reduce performance guardbanding during the design process, and increase product performance.

## TIMING AND POWER-DRIVEN FILL

Fill shapes must be placed to improve not only CMP uniformity, but circuit performance as well. Insertion of fills needs to be automated utilizing available fill insertion guidelines. For timing- and power-critical circuits, timing- and power-aware fill synthesis methodologies are needed. We target these problems and provide an automated fill synthesis framework by which circuit performance can also be improved. In our fill synthesis framework, which is based on a physical-analog paradigm originating in [13, 14], the possible locations for fill insertion have

allocated energies. When a fill is to be inserted, the location with lowest energy is sought. Fills are inserted into a given region one by one, until a prescribed density constraint is satisfied while performance (timing or power) is improved.

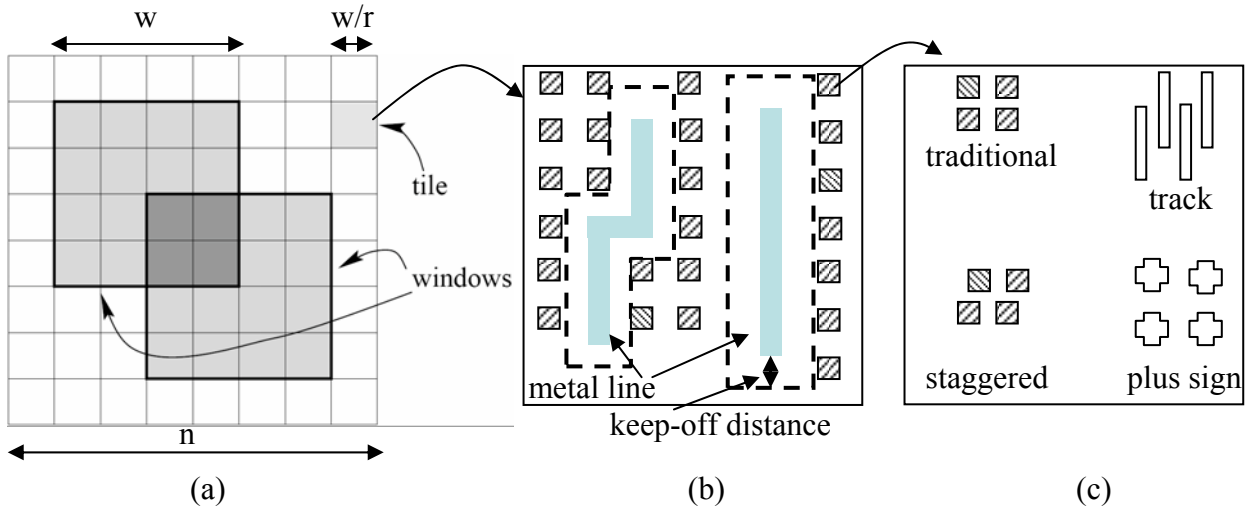



**Figure 3.** (a) Windows and tiles used in fill synthesis; (b) sample layout within a tile, showing keep-off distance and inserted floating fill shapes; and (3) different shapes and patterns of floating fill.

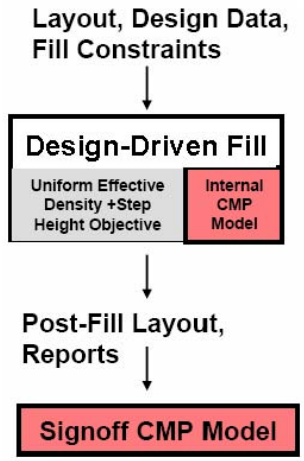



**Figure 4.** Design-driven fill with embedded CMP model.

## Fill optimization methodology

We now describe our fill optimization methodology, covering the steps of region definition, grid definition, energy modeling within a grid, and fill insertion within a grid.

**Adaptive Regions.** Given a metal layer, we use a scanline algorithm to determine regions between facing interconnects with no other interconnect in between. A region consists of a maximal rectangular area such that two interconnects directly face each other in the horizontal

(vertical) direction, when the main routing orientation in the given interconnect layer is vertical (horizontal). To obtain a uniform post-fill metal density while also considering performance impact of fill, we use a uniform target density for each region. Our region definition is adaptive in the sense that it follows the specific configuration of the interconnect design, in contrast to methods that simply dissect the layout into uniform square regions.

**Converting a Region to a Grid.** In converting a region into a grid, we can define possible grid rectangles for fill insertion. Each such region is converted into a grid after keep-off distances are stripped (i.e., made unavailable for filling). For each region, we form a grid of rectangles into which fills will be placed. A single region is shown in Figure 5, where square boxes are the grid rectangles into which metal fills can be placed. The auxiliary frame forms the outer edge of the grid. The grid rectangles and the frame are connected together with bonds, which have adjustable energy values between them. Bonds touching a grid rectangle are its incident bonds.

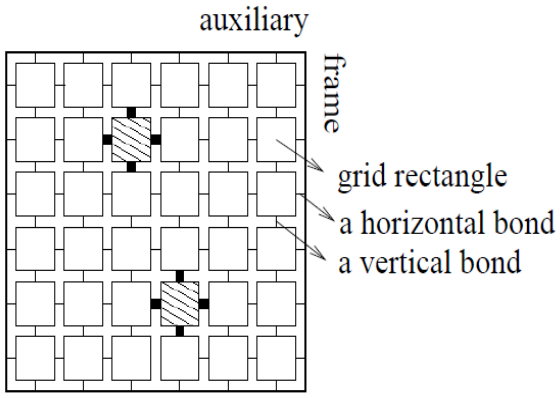



**Figure 5.** A grid inside a region consisting of grid rectangles to be filled.

**Energy Modeling in a Grid.** We have developed a parameterized model [3] to choose bond values according to a given fill insertion guideline. Fills may need to be inserted away from the critical interconnects. We design the bond energies so that they decrease as we move away from critical nets.

**Insertion of Fills into a Grid.** We map the fill insertion problem to an energy-minimization problem. Fills are placed in available locations to minimize an energy criterion. The rectangles are fixed in the grid, and hence the locations of bonds are fixed. We find the grid rectangle for fill insertion with minimum sum of energies for its incident bonds. We implement metal fill insertion using greedy optimization.

**Timing-oriented optimization**

Interconnect delay has been catching up with gate delay and even exceeding it in certain contexts. While STA (Static Timing Analysis) with gate delays alone would be sufficient previously, sub-90nm technologies require accurate incorporation of critical nets into the STA. Hence, similar to critical gates being important in a critical path, critical nets are important during fill optimization and require special attention. In fact, chip performance can be improved by optimizing the fills around the critical nets. Table 2 shows our critical net-aware fill flow.

The first step in timing critical net-aware fill is to identify the timing critical nets. We handle this step by using the STA timing reports and extracting critical net names. Fills should be placed away from critical nets. To account for this, we update bond energies so that the energies favor locations away from the critical nets.

**Power-oriented optimization**

Power consumption is a dominant consideration in integrated circuits. Dynamic power increases as larger capacitances are charged and discharged at every clock cycle. Power can be reduced by reducing the capacitances of power-critical nets. Similar to the critical-aware timing optimization, we determine the power-critical nets and optimize the fills around them. A net is power-critical if its dynamic power (proportional to the product of switched capacitance and switching activity factor) is high. To determine such nets, we rank the nets according to their power consumption. We select a number of the most power-critical nets and optimize the fills around these nets so that the capacitances seen by such nets are minimized.

**Table 2.** Timing critical net-aware fill flow.

| |
|---|
| 1. Place, synthesize clock network, and route the design. |
| 2. Extract SPEF parasitics from post-routing DEF file. |
| 3. Run static timing analysis using SPEF file from Step 2. |
| 4. Use Perl scripts to obtain top critical net names. |
| 5. Input net names to metal fill optimization (MFO) for energy calculation. |
| 6. Insert timing critical net-aware fills. |

**Experimental setup, protocol and results**

Our experiments validate the efficiency and relevance of the proposed fill synthesis method by comparing layouts filled with the proposed method against those filled with a traditional fill method. We have implemented the proposed method using approximately 4000 lines of custom C++ code. Input is a GDSII (Graphic Data Stream) file and output is the fill-optimized GDSII and DEF (Design Exchange Format). In our flow, we use Cadence SOC Encounter v5.2 for placement, clock tree synthesis and routing, and Synopsys Star RCXT 2007.06 to extract post-fill parasitics in SPEF format. We compare layouts filled with the proposed scheme versus Blaze DFM, Inc. Intelligent Fill (IF) or Mentor Graphics Corp. Calibre-implemented fill scheme representing the merits of traditional fills; these tools are otherwise powerful industrial tools. We use Synopsys Primetime 2005.12 for static timing analysis. For CMP prediction, we use Cadence CCP 1.4 with models optimized for the TSMC 65nm process. Our analysis flow is given in Table 3. For the optimized case, fills are inserted using our code. The SPEF file on Line 4 is in Standard Parasitic Exchange Format, and the DEF file is in Design Exchange Format. For the power-oriented fill, instead of running static timing analysis, we run scripts to compute the interconnect switching dynamic power. We have used TSMC 65nm GPlus 8-layer metal technology in our experiments. We have used the ISCAS89 S38417 and OpenCores ALU and AES benchmark circuits and a 2.2GHz server to conduct our experiments. The testcases are summarized in Table 4.

**Table 3.** Analysis flow.

| |
|---|
| 1. Obtain filled (i.e., post-fill insertion) GDS |
| 2. Use Perl scripts to obtain DEF file containing fills. |
| 3. Extract SPEF parasitics from DEF containing fills. |
| 4. Run static timing analysis using SPEF file from Step 3. |
| 5. Use Mentor Calibre scripts to obtain density histograms. |
| 6. Use Perl scripts to obtain timing slack histograms. |

Figure 6 shows the worst 200 critical path setup slacks for the S38417 benchmark. Slack is a parameter which gives the difference between the actual signal arrival time at a flip-flop input or primary output, and the corresponding required arrival time. A distribution towards the right implies better circuit timing. We present the results using histograms to show that not only a single path is modified by optimization, but rather that the whole distribution shifts and its shape changes in most cases. As can be observed from the figure, our MFO approach results in improved timing as compared to the fill options available from the Blaze IF tool.

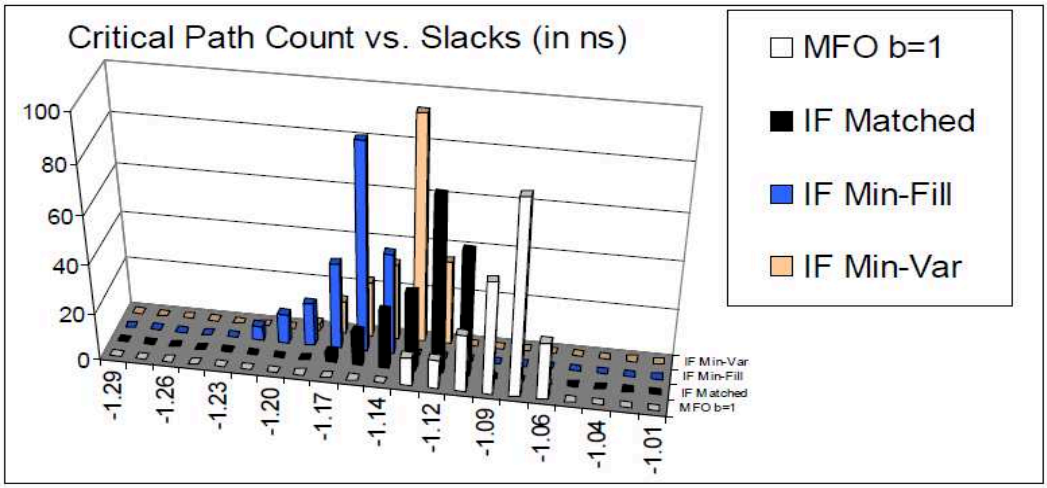



**Figure 6.** Timing slacks for S38417 ISCAS'89 benchmark with 43,076 interconnect segments for MFO, Blaze IF matched to the fill number of MFO, Blaze IF min-fill, and Blaze IF min-var cases for the case of 30% target density. Runtime is 172 seconds including printing out the GDS.

Table 4 summarizes our experimental results for power-oriented fill. We use 100 power-critical nets for each testcase. Power is given in the table for 1GHz operating frequency. For large circuits, a non-negligible The S38417 testcase contains many flip-flops with short critical paths, hence its optimization results (an observed 35.8% dynamic power savings) are highly optimistic. With longer critical paths, there will be more interconnects involved, not all of which can be optimized simultaneously, and fill optimization will result in smaller timing and power impact on average.

**Table 4.** Power-oriented fill results.

| Circuit | Power with fills | Power with power-oriented fills | % Improvement |
|---|---|---|---|
| S38417 | 1746 | 1121 | 35.8 |
| ALU | 6337 | 6185 | 2.4 |
| AES | 11670 | 11350 | 2.7 |

## CMP FILL SYNTHESIS: FLOW ISSUES AND FUTURES

As we look to the future, several aspects of the CMP fill synthesis flow are worth reviewing. We first consider the "platform" for delivery of future CMP fill syntheses that enhance manufacturability, as well as parametric yield, of the IC product. A natural scenario is for fill synthesis and timing to be integrated within the detailed routing tool. This is intuitively reasonable for such reasons as: (1) routers create geometries and close timing, and hence are the natural candidates to perform fill synthesis; (2) timing closure is more certain when fill is synthesized by the design team before handoff to manufacturing; and (3) grounded fill, which reduces timing uncertainty and improves IR drop, is a natural extension of power/ground routing. At the same time, there are significant barriers to emergence of such a fill synthesis solution: (1) complicated (wraparound, full-chip, width-distribution-dependent, etc.) density analyses that support high-quality CMP modeling are not easily performed by a router; (2) tools that are optimized for batch solution of full-chip detailed routing may not be able to deliver high-quality fill without a significant runtime hit; (3) with the possible exception of hold-time slack and coupling-induced delay uncertainty issues, grounded fill can be harmful to chip performance, and the more preferable floating-fill approach is less natural for a router; and (4) at the design-manufacturing handoff, better passing of design intent [15] (and, acceptance of the foundry's role in performing litho-driven reticle enhancement) could allow responsibility for CMP fill to remain with the foundry, and reduce any need for the router to assume this role.

Second, the industry will require new techniques to mitigate the impact of CMP fill on manufacturing closure. For example, depth of focus (DOF) is a major contributor to lithographic-process margin, and a major cause of focus variation is imperfect planarization of fabrication layers. Currently, optical proximity correction (OPC) methods are oblivious to the predictable nature of focus variation arising from wafer topography. As a result, designers suffer from manufacturing-yield loss, as well as loss of design quality through unnecessary guardbanding. Figure 7 shows how post-CMP-thickness variation can result in loss of critical dimension (CD) control [18]. To minimize the impact of pattern-dependent effects of the CMP process, the OPC methods should be aware of the post-CMP topography to assign appropriate defocus value for all the features with the same topography. Gupta *et al.* [18] propose a flow and methodology to drive OPC with a topography map of the layout that is generated by CMP simulation. The experimental results show that the proposed topography-aware OPC can yield up to 67% reduction in edge-placement errors and 12% reduction in timing uncertainty at the cost of little increase in mask cost.

Finally, we observe that any future CMP fill methodology will involve some subset of four potential elements: (1) CMP simulation, (2) topography-aware RC extraction, (3) timing and signal integrity awareness, and (4) multilayer fill synthesis. A "smart" fill synthesis, if it

accurately optimizes planarity and accounts for multilayer topographic effects, minimizes the need for (2). And, while efforts toward (1) and (2) only address the analysis side of fill pattern design, a combination of (1), (3) and (4), with planarity assumptions validated by best possible CMP simulation, may provide a very strong replacement for today's physical-verification-based floating fill or router-based grounded fill [2].

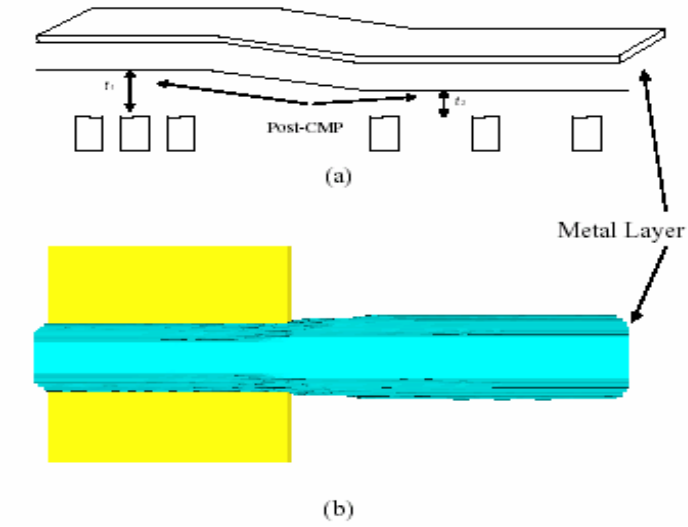



**Figure 7.** (a) Side view showing thickness variation over regions with dense and sparse layout. (b) Top view showing CD variation when a line is patterned over a regions with uneven wafer topography, i.e., under conditions of varying defocus [18].

## CONCLUSIONS

We have reviewed the role of CMP fill in reducing manufacturing variation (specifically, post-polishing wafer topography). We have also pointed out the recent trend toward design-driven fill synthesis that integrates scalable global optimization, handling of a rich array of constraint types, CMP process modeling and simulation, and incremental parasitic extraction and timing analysis. In this paper, we introduced a fill synthesis framework that heuristically accepts and realizes complex, performance-driven CMP fill guidelines. Our heuristic approach uses an energy minimization framework to achieve metal fill insertion in a given region. We have used our tool to optimize fill in both timing criticality-aware and power-aware use contexts. With large 65nm industrial testcases, we recover up to 96.1% of timing slack that otherwise would have been lost due to critical net capacitance increase caused by added floating fills. We have also demonstrated up to 35.8% reduction in total interconnect switching power using power-aware fill insertion.

## REFERENCES


1. A. B. Kahng, K. Samadi and P. Sharma, “Study of Floating Fill Impact on Interconnect Capacitance,” in *Proc. ISQED*, 2006, pp. 691-696.
2. A. B. Kahng and K. Samadi, “CMP Fill Synthesis: A Survey of Recent Studies,” in *IEEE Trans. on CAD*, 27(1), 2007, pp. 3-19.
3. R. O. Topaloglu, “Characterization, Modeling and Optimization of Fills and Stress in Semiconductor Integrated Circuits,” *Ph.D. thesis, Dept. of Computer Science and Engineering, UC San Diego*, 2008.
4. D. White and A. Gower, “Impact of Multi-Level Chemical Mechanical Polishing on 90nm and Below,” in *Proc. VMIC*, 2005.
5. B. Stine, D. Ouma, R. Divecha, D. Boning, J. Chung, D. L. Hetherington, I. Ali, G. Shinn, J. Clark, O. S. Nakagawa and S. -Y. Oh, “A Closed-Form Analytic Model For ILD Thickness Variation In CMP Processes,” in *Proc. CM P-MIC*, 1997, pp. 266.273.
6. B. Stine, D. O. Ouma, R. R. Divecha, D. S. Boning, J. E. Chung, D. L. Hetherington, C. R. Harwood, O. S. Nakagawa and S.-Y. Oh, “Rapid Characterization and Modeling of Pattern-Dependent Variation in Chemical-Mechanical Polishing,” in *IEEE Trans. on Semiconductor Manufacturing,* 11(2), 1998, pp. 129-140.
7. D. Ouma, “Modeling of Chemical-Mechanical Polishing for Dielectric Planarization,” *Ph.D. Dissertation*, Dept. of Electrical Engineering and Computer Science, MIT, Cambridge, MA, 1998.
8. C. Ouyang, K. Ryu, L. Milor, W. Maly, G. Hill and Y.-K. Peng, “An Analytical Model of Multiple ILD Thickness Variation Induced by Interaction of Layout Pattern and CMP Process,” in *IEEE Trans. on Semiconductor Manufacturing*, 13(3), 2000, pp. 286-292.
9. C.-H. Yao, D. L. Feke, K. M. Robinson and S. Meikle, “Modeling of Chemical Mechanical Polishing Processes Using a Discretized Geometry Approach,” in *J. Electrochemical Society,* 147(4), 2000, pp. 1502-1512.
10. H. Liao, L. Song, N. Jakatdar and R. Radojcic, “Integration of CMP Modeling in RC Extraction and Timing Flow,” in *Proc. CICC*, 2007, pp. 249-252.
11. J. Luo and D. Dornfeld, *Integrated Modeling of Chemical Mechanical Planarization for Sub-micron IC Fabrication: From Particle Scale to Feature, Die and Wafer Scales*, Springer, 2004.
12. D. Wang, W. Barth, W. Catabay, S. Lakshminaryanan, P. Burke and J. Pallinti, “Removing Copper Over low-k Films Using Stress-Free Polishing,” in *Solid State Technology Magazine*, 2004.
13. A. B. Kahng and R.O. Topaloglu, "Performance-Aware CMP Fill Pattern Optimization," **Invited Paper**, in *Proc. VMIC,* 2007.
14. R. O. Topaloglu, "Energy-Minimization Model for Fill Synthesis," in *Proc. ISQED*, 2007, pp. 444-451.
15. Y. Chen, P. Gupta and A. B. Kahng, “Performance-Impact Limited Area Fill Synthesis,” in *Proc. DAC*, 2003, pp. 22-27.
16. A. B. Kahng, G. Robins, A. Singh, H. Wang and A. Zelikovsky, “Filling and Slotting: Analysis and Algorithms,” in *Proc. ISPD*, 1998, pp. 95–102.
17. R. Tian, D. F. Wong, R. Boone and A. Reich, “Dummy feature placement for oxide chemical–mechanical polishing manufacturability,” *Technical Report*, Dept. Computer Science, Univ. Texas, Austin, TX, pp. 9-19, 1999.
18. P. Gupta, A. B. Kahng, C.-H. Park, K. Samadi and X. Xu, “Wafer Topography-Aware Optical Proximity Correction,” in *IEEE Transactions on CAD*, (25)12, 2006, pp. 2747-2756.
19. A. Balasinski, J. Cetin and A. B. Kahng, “Intelligent Fill Pattern and Extraction Methodology for SoC,” in *Proc. IWSOC*, 2006, pp. 156–159.