Manufacturability-Aware Physical Layout Optimizations

David Z. Pan and Martin D. F. Wong†
Dept. of Electrical and Computer Engineering, Univ. of Texas at Austin
† Dept. of Electrical and Computer Engineering, Univ. of Illinois at Urbana-Champaign
Email: dpan@ece.utexas.edu, mdfwong@uiuc.edu

***Abstract*— Nanometer VLSI design is greatly challenged by the growing interdependency between manufacturing and design. Existing approaches in design for manufacturability (DFM) are still mostly post design, rather than during design. To really bridge the gap between design and manufacturing, it is important to model and feed proper manufacturing metrics and cost functions upstream, especially at the key physical layout optimization stages such as routing and placement, to have major impacts. In this paper, we show several aspects of the true manufacturability-aware physical design, from lithography-aware routing, to redundant-via aware routing, to CMP aware floorplanning and placement, and show their promises.**

I. Introduction

As VLSI technology continues to scale down to nanometer dimensions, the semiconductor industry is greatly challenged not only by many entangled *deep sub-micron* physical effects to reach *design closure* for timing, signal integrity, low power, etc., but also by *deep sub-wavelength* lithography and other manufacturability issues to reach the *manufacturing closure*, i.e., being manufactured reliably with high yield and robustness.

Among various manufacturing limits/issues, lithography is probably the most fundamental one. Leading edge IC manufacturers nowadays still use 193nm lithography to print 90nm or smaller feature size, relying heavily on various and even exotic resolution enhancement techniques (RETs) [1], [2], such as optical proximity correction (OPC), phase shift mask (PSM) and off axis illumination (OAI) to modify the chip mask database (GDSII) and achieve better printability, higher yield, and less variability. However, these RETs are mostly done during post-tapeout mask synthesis at fabs, which may be too late to make all the necessary corrections, or follow designer's intent. Besides, RETs are expensive. For a typical 90nm IC design process, 12 out of roughly 30 masks would require some form of resolution enhancement [2]. RET dramatically increases the mask cost, which has soared and reached $2 million per set at 90nm node. One major reason is due to extensive usage of RET. The semiconductor industry is adopting the immersion lithography, which will further extend the 193nm lithography to 45nm node or even below. Therefore, the usage of RETs will become more pervasive for future technology generations.

Compared to the traditional random defects, the yield loss due to lithography limitation is highly layout dependent. Thus different layouts will need different level of resolution enhancements. There are also other important pattern (or design) dependent yield loss mechanism, such as those caused by chemical-mechanical polishing (CMP), and via failure. As the physical layout design and manufacturing closely relate to each other for nanometer VLSI, there is clearly a growing level of interdependency between them. A *true* design for manufacturability (DFM) flow is needed to *abstract* and *predict* the downstream manufacturing effects upstream into the key layout optimization stages, such as routing, placement, and floorplanning, to have the true designer's intent preserved, maximize the overall manufacturability, and minimize the overall manufacturing cost. In this paper, we will present several key aspects of the true manufacturability-aware physical layout optimizations, guided by proper predictive functions and metrics.

The rest of the paper is organized as follows. Section II shows RET-aware routing. Section III presents redundant-via enhanced routing for yield improvement. Section IV presents a CMP-aware floorplanning, followed by the conclusion in Section V

II. RET-AWARE Detailed ROUTING

Since RETs are not cheap (e.g., data volume increase by up to 10x, more mask write and inspection time) and one may not be able to make all the corrections during mask synthesis, it will be beneficial to consider the downstream lithography impact early on during the routing stage, especially at the detailed routing stage where exact polygons are determined.

One solution is to provide more and more routingg rules by fabs to the design houses and CAD tools. However, as technology moves to 90nm and beyond, the number of rules quickly explodes [3], [4], [5], [6]. This will significantly affect the router performance. In addition, there may be exotic rules hard for routers to resolve. Since the rule-based models usually lack accuracy, very conservative or restricted rules may have to be given [7]. Since these restricted design rules will be applied globally, the physical layout may be overly pessimistic. On the other hand, lithography simulations, though more accurate, could be very CPU intensive. It could easily take hundreds of CPU hours to run a full chip simulation-based OPC. Our experience with PROLITH [8] and SIGMA-C [9], two leading industry lithography simulation tools, shows that it could take a few minutes to simulate a 5um x 5um area (in accurate mode).

Thus it is desirable to directly link *fast* lithography simulations with the routing. There are very few works on this topic so far. The work by [10] is the first attempt to our best knowledge, where the optical interferences from neighboring edges are accumulated for an entire net under consideration. It then used the maze routing with Lagrangian relaxation to satisfy the interference constraint for each net. The cumulative interference metric, however, is not a direct measurement related with the final printability.

In [11], the concept of lithography hotspot map (LHM) was introduced to for post-routing fixing on the litho hotspot regions (similar to congestion or thermal hotspots). To be more specific, fast lithography simulations with effective table look-up are performed to generate the edge placement error (EPE) map to reflect lithography hotspots. Fig. 1 shows an example of the EPE, which is the edge difference between the mask and the wafer. Since EPE is often used by RET tools to guide the mask synthesis [12], less EPE usually corresponds to less RET effort. The EPE guided correction technique naturally fits into existing design flows and is capable of handling *full-chip* capacity.

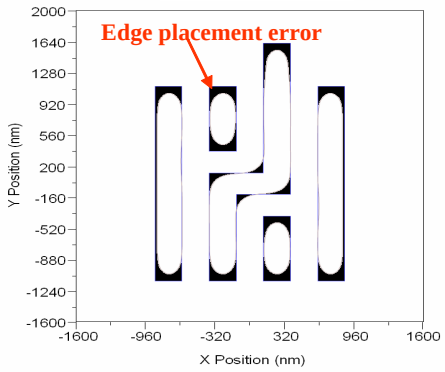

Fig. 1. Edge placement error map.

To fast generate EPE map, efficient table-look-up techniques can be used [11]. When the EPE map is generated, a ranked list of interfering neighboring edges can be stored. This information will be useful for RET-aware detailed routing (RADAR) with EPE guided wire spreading and rip-up and reroute [11]. The method requires only one full-chip fast litho-simulation to filter out the EPE hot spots. Then re-simulations are needed only when necessary in the litho hot spots when routing changes are made. Compared to [10] which performed simulations during the entire maze routing, it is more suited for full-chip optimizations. An EPE guided post-routing optimization flow is shown in Fig. 2.

The algorithm and flow in Fig. 2 are implemented in an industrial strength router and validated on some real 65nm industry designs. Fig. 3 shows the EPE hotspots (i.e., EPE bigger than certain threshold) before and after post routing optimization, such as wire spreading and rip-up and reroute. The number of EPE hotspots is reduced by 40% after both wire spreading and rip-up & reroute, by comparing Fig. 2 (a) and (c).

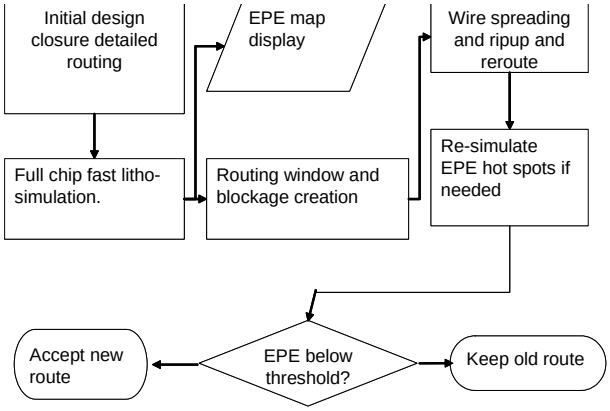

Fig. 2. EPE based detailed routing flow.

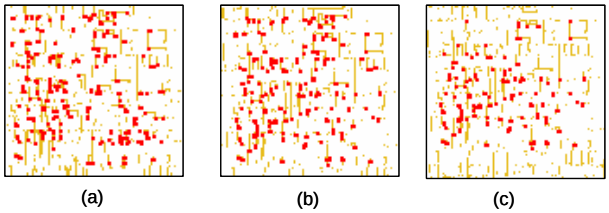

Fig. 3. (a) EPE map for the initial routing; (b) EPE map after the wire spreading; (c) EPE map after the rip-up & reroute.

III. REDUNDANT-VIA AWARE ROUTING

Among various yield loss mechanisms, via failure in nanometer IC manufacturing is another important one, especially for copper interconnects. A via may fail completely or partially due to various reasons such as cut misalignment, electromigration, or thermal stress induced voiding. A complete via failure will cause a broken net, while a partial via failure will increase the resistance of the signal net and lead to delay penalty and timing problems.

A common solution to reduce the via failure is to add a redundant via adjacent to a single via as a backup (Fig. 4), so that it is less likely to fail. Redundant vias also decrease the resistance of via and alleviate the delay penalty by partial via failures. In fact, redundant via insertion has been strongly recommended by major foundries in their 130nm and 90nm processes [13] to improve the yield. Major EDA vendors such as Cadence and Synopsys have added the feature of redundant via insertion to their latest routers. There are also third-party EDA tools such as Nannor Acuma [14] and Prediction EYE/PEYE [15] specially designed to insert redundant vias. However, all these tools insert redundant vias at the post-routing stage, e.g., by greedy ad hoc insertion and wire spreading. Thus the redundant via insertion flexibility is greatly constrained. An example is given in Fig. 4 where for routing in Fig. 4 (a), no redundant via can be added for via B, but it can be done in Fig. 4 (b). The via B in Fig. 4 (a) is called a dead via, as it cannot have any redundant via. One can also observe from Fig. 4 that different vias have different freedoms for redundant via insertion.

To improve the redundant via coverage, it will be more effective to consider the feasibility of redundant via insertion

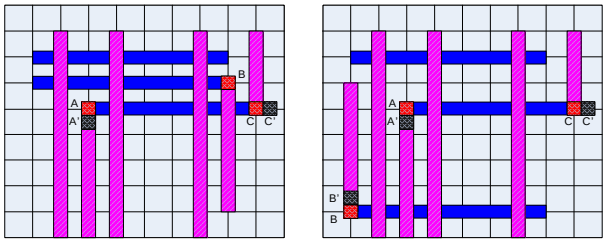

Fig. 4. (a) Three different vias with different degrees of freedom to insert a redundant via. (b) an alternative route where a redundant via B' can be added for via B.

during the routing, instead of post-routing insertion [16]. In [16], this problem is formulated as a *maze routing with redundant via constraint* (MRRVC) problem, i.e., each net has some upper bound on the number of the dead vias. It is a flexible and generic formulation since some nets are critical and may require 100% redundant via coverage, while some other nets can tolerate partial via failure (if it is not timing critical). More redundant vias may hurt the overall routability, if the routing resources are limited. This is a routing tradeoff issue. The redundant via constraints can be user-specified.

By assigning proper cost (e.g., the number of dead vias for the current net under consideration, as well as those for previously routed nets) to each routing edge during the maze routing, the MRRVC problem can be transformed into a multi-constraint shortest path problem. It can then be solved using the Lagrangian relaxation technique [16].

Table I shows the comparison of the tradeoff from MRRAV, where several different algorithms are compared during maze routing: DV0, DV1 and DV2 have constraints of up to 0, 1 and 2 dead vias per net, while Conv is the conventional maze routing without redundant via awareness during routing, but add redundant via insertion post routing. As we can see, the MRRAV algorithm can trade off between the number of routable nets (#RNets) and the redundant via coverage (%RVia). By relaxing the constraint to allow no more than two dead vias per net, one may still achieve almost the same routability, yet improve the percentage of redundant via covereage remarkably by 36% (66.1% versus 48.5%). Thus good tradeoff can be obtained.

TABLE I

COMPARISON OF MRRAV ALGORITHM WITH CONVENTIONAL MAZE ROUTING.

Alg	#RNets	%RNets	#Via	%RVia
DV0	546	68.25%	1937	100%
DV1	689	86.13%	2401	84.0%
DV2	773	96.6%	2504	66.1%
Conv	782	97.7%	2377	48.5%

IV. CMP-AWARE FLOORPLANNING AND PLACEMENT

Chemical mechanical polishing (CMP) is a fundamental manufacturing enhancement step to obtain global planarization, on both front end process steps such as shallow trench isolation (STI) and back end multi-level copper interconnection. Control of post-CMP topography variation is crucial in meeting the ever decreasing depth-of-focus requirement in photolithography and the ever increasing levels of interconnect demands due to routing complexity. Since the post-CMP topography variation is strongly dependent on the layout patterns [17], [18], it should be considered during the physical layout optimization. In this section, we will present a CMP-aware floorplanner for shuttle mask optimization [19]. The same principle, however, can be used during IC floorplanning and placement to achieve better overall chip planarization and less variability.

To guide effective layout optimization, one needs to understand the post-CMP modeling. Intuitively, higher feature density will lead to higher post-CMP topography. But the relationship is not that simple. Ouma et al's 2-D low-pass filter model [20] is well accepted and widely used model to estimate the post-CMP topography variation with respect to the feature density. It is inexpensive to compute, easy to calibrate, and reasonably accurate.

To further reduce the topography variation, dummy features are often inserted before performing the CMP process. Tian et al [21] developed a convolution model for optimal dummy feature insertion for oxide. It was later extended to handle the shallow trench isolation (STI) process [22]. The STI model is more complex than the oxide CMP model, as it requires modeling of dual-material polish and local pad compression. Thus, nonlinear programming formulations and iterative methods were proposed to minimize topography variation with dummy features [22].

These linear/nonlinear programming formulations give the exact and best dummy insertion results for post CMP topography variation. However, they are too slow to be used inside a floorplanner or placement engine. For CMP-aware floorplanning and placement, one needs to be able to *predict* the post-CMP effect efficiently. In [19], three *predictive* models and cost functions are proposed to guide the CMP awareness. They are

1) $MaxDiff = max\{\rho_{i,j}\} - min\{\rho_{i,j}\}$. This function represents the maximum effective density (ρ) difference without dummy insertion [20]. The (i, j) represents the layout grid index.
2) The second predictive function *SDH*, meaning "sigma delta height", attempts to predict the dummy insertion effect.
   $SDH = \sum_{i,j}(1 - c_{i,j})(\rho_{i,j} - min\{\rho_{i,j}\})$
   Essentially, for those locations with low effective density ($\rho_{i,j}$), a higher capacity (more white space for dummy feature insertion) at that grid $c_{i,j}$ will be desirable.
3) The third function is a normalized *SDH*.
   $NSDH = \sum_{i,j}(2 - c_{i,j})[1 + (\rho_{i,j} - min\{\rho_{i,j}\})/(max\{\rho_{i,j}\} - min\{\rho_{i,j}\})]$

A CMP aware floorplanner based on [23] was implemented [19], using the above predictive models to estimate the post-CMP topography variation. The objective function is a weighted combination of the area and the post-CMP topography variation of the floorplan, inside the simulated annealing (SA) engine. The exact and more computationally

expensive algorithm [22] is used get the final optimal dummy insertion on the best floorplanning solution from SA engine. Since it is called only once, its computational time is still acceptable.

The CMP-aware floorplanning algorithm [19] is tested on a data set from a real industry mask for the 90nm technology node which consists of 10 chips. Table II shows the comparison among different predictive functions using Area, Area+MaxDiff, Area+SDH, and Area+NSDH. In the table, WS is the white space ratio, VwoD is the variation without dummy insertion. The unit of the variation is angstrom. VwithD is the variation with dummy insertion, from solving the exact method [22] with minimum variation objective. DAmount represents the minimum dummy fill amount obtained by solving [22] with the minimum fill amount objective.

TABLE II

COMPARISON CMP-AWARE FLOORPLANNING USING DIFFERENT PREDICTIVE FUNCTIONS.

Cost function	WS	VwoD	VwithD	DAmount
Area only	2.82%	818	92	340
Area+MaxDiff	6.87%	612	67	338
Area+SDH	8.27%	588	64	298
Area+NSDH	6.04%	751	67	298

As we can see, all three predictive functions work well in effectively reducing the post CMP topography variation. The variation is improved by around 30% in all three functions. With the same amount of dummy feature insertion, area+NSDH obtains slightly larger variation than that of area+SDH. However, it obtains the minimum white space. If we consider all three metrics of area, topography variation, and amount of dummy feature insertion, area+NSDH performs the best. Fig. 5 shows the floorplan layout obtained by using the cost function of area+NSDH.

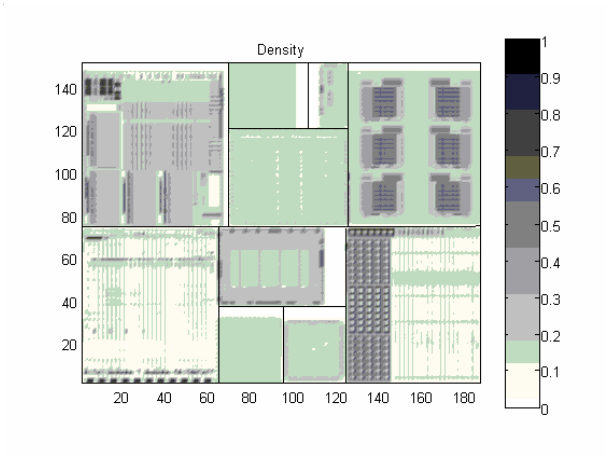

Fig. 5. A shuttle mask floorplan by area+NSDH

V. CONCLUSIONS

For nanometer designs and manufacturing closure, a *true* design for manufacturability (DFM) flow is needed. It should be able to predict the downstream manufacturing effects upstream into the key layout optimization stages to maximize the overall manufacturability/yield, and provide design and manufacturing tradeoff. Since DFM is a rather broad area, we do not intend to be exhaustive in this paper but focus on several key aspects on the manufacturability-aware physical layout optimizations to validate concepts and show promises, such as lithography-aware routing, redundant via enhanced routing, and CMP aware floorplanning. We believe many research opportunities lie in the true manufacturability-aware physical design.

VI. ACKNOWLEDGMENT

This work was partially supported by an IBM Faculty Award, National Science Foundation under grant CCR-0306244, an Intel Equipment Grant, and the lithography simulation software donations from KLA-Tencor and Sigma-C. The authors would like to thank the current and former students at UT Austin, Lida Huang, Joydeep Mitra, Ruiqi Tian, Gang Xu, and Peng Yu for their contributions to various aspects in this paper. We also thank Warren Grobman, Chris Mack, Lueng Heng, Ruchir Puri, and Patrick McGuinness for their helpful discussions.

REFERENCES

[1] L. W. Liebman, "Layout impact of resolution enhancement techniques: impediment or opportunity?," in *Proc. of the International Symposium on Physical Design*, pp. 110–117, 2003.

[2] F. M. Schellenberg, "Resolution enhancement technology: The past, the present and extension for the future," in *SPIE Microlithography Symposium*, 2004.

[3] H. K.-S. Leung, "Advanced Routing in Changing Technology Landscape," in *ISPD '03: Proceedings of the 2003 international symposium on Physical design*, pp. 118–121, ACM Press, 2003.

[4] A. B. Kahng, "Research directions for coevolution of rules and routers," in *Proc. ACM/IEEE Intl. Symp. on Physical Design*, pp. pp. 122–125, April 2003.

[5] P. Gupta and A. Kahng, "Manufacturing-aware physical design," in *Proc. IEEE/ACM Intl. Conference on Computer-Aided Design*, pp. pp. 681–687, November 2003.

[6] L. Scheffer, "Physical CAD changes to incorporate design for lithography and manufacturability.," in *ASP-DAC*, pp. 768–773, 2004.

[7] M. Lavin, F.-L. Heng, and G. Northrup, "Backend CAD flows for "restrictive design rules"," in *Proc. of the International Conference on Computer-Aided Design*, pp. 739–746, 2004.

[8] *PROLITH (version 8.0)*. KLA-Tencor Corporation.

[9] *SOLID-CTM (version 6.4.1)*. Sigma-C Software.

[10] L.-D. Huang and M. D. F. Wong, "Optical Proximity Correction (OPC)-Friendly Maze Routing," in *Proc. of Design Automation Conference*, pp. 186–191, ACM Press, 2004.

[11] J. Mitra, P. Yu, and D. Z. Pan, "RADAR: RET-Aware Detailed Routing Using Fast Lithography Simulations," in *Proc. of Design Automation Conference*, ACM Press, 2005.

[12] N. B. Cobb, *Fast Optical and Process Proximity Correction Algorithms for Integrated Circuit Manufacturing*. PhD thesis, University of California, Berkeley, 1998.

[13] *TSMC Symposium*, 2004.

[14] *http://www.nannor.com/redundant.htm.*

[15] G. A. Allan and A. J. Walton, "Automated redundant via placement for increased yield and reliability," in *Proc. SPIE Vol. 3216, p. 114-125, Microelectronic Manufacturing Yield, Reliability, and Failure Analysis III, Ali Keshavarzi; Sharad Prasad; Hans-Dieter Hartmann; Eds.*, pp. 114–125, Sept. 1997.

[16] G. Xu, L.-D. Huang, D. Z. Pan, and M. D. Wong, "Redundant-Via Enhanced Maze Routing for Yield Improvement," in *Proc. of Asia and South Pacific Design Automation Conference*, 2005.

[17] S. Prasad, W. Loh, A. Kapoor, E. Chang, B. Stine, D. Boning, and J. Chung, "Statistical metrology for characterizing CMP processes," *Microelectron. Eng.*, vol. 33, no. 1-4, pp. 231–240, 1996.

[18] B. Stine, D. Ouma, R. Divecha, D. Boning, and J. Chung, "Rapid Characterization and Modeling of Pattern Dependent Variation in Chemical Mechanical Polishing," *IEEE Trans. Semiconduct. Manufact.*, vol. 11, pp. 129–140, Feb. 1998.

[19] G. Xu, R. Tian, D. Z. Pan, and M. D. Wong, "CMP aware shuttle mask floorplanning," in *Proc. of Asia and South Pacific Design Automation Conference*, Jan. 2005.

[20] D. Ouma, D. Boning, J. Chung, G. Shinn, L. Olsen, and J. Clark, "An integrated characterization and modeling methodology for CMP dielectric planarization," in *Proc. Int. Interconnect Technology Conf*, pp. 67–69, June 1998.

[21] R. Tian, D. F. Wong, and R. Boone, "Model-based dummy feature placement for oxide chemical-mechanical polishing manufacturability," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 20, pp. 902–910, July 2001.

[22] R. Tian, X. Tang, and D. F. Wong, "Dummy feature placement for chemical-mechanical polishing uniformity in a shallow trench isolation process," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 21, pp. 63–71, Jan. 2002.

[23] D. F. Wong and C. L. Liu, "A new algorithm for floorplan design," in *Proc. of Design Automation Conference*, pp. 101–107, IEEE Press, 1986.