Multilevel Routing with Jumper Insertion for Antenna Avoidance

Tsung-Yi Ho[1], Yao-Wen Chang[2]*, and Sao-Jie Chen[2]†

[1]Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan

[2]Graduate Institute of Electronics Engineering and Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan

Abstract

As technology advances into nanometer territory, the antenna problem has caused significant impact on routing tools. The antenna effect is a phenomenon of plasma-induced gate oxide degradation caused by charge accumulation on conductors. It directly influences reliability, manufacturability and yield of VLSI circuits, especially in deep-submicron technology using high density plasma. Furthermore, the continuous increase of the problem size of IC routing is also a great challenge to existing routing algorithms. In this paper, we propose a novel framework for multilevel full-chip routing with antenna avoidance using built-in jumper insertion approach. Compared with the state-of-the-art multilevel routing, the experimental results show that our approach reduced 100% antenna-violated gates and results in fewer wirelength, vias, and delay increase.

1 Introduction

Yield and reliability of VLSI circuit have always been an important item on the agenda of IC manufacturers. With the continuous and rapid increase in complexity of VLSI designs and fabrication technologies, manufacturing yield and product reliability is becoming one of the most important issues among the other existing ones, such as small die size, high speed, low power and so on [2]. The fine feature size of modern IC technologies is typically achieved by using plasma-based processes. As the technology enters the deep-submicron era, more stringent process requirements cause some advanced high-density plasma reactors adopted in the production lines to achieve fine-line patterns. However, these plasma-based processes have a tendency to charge conducting components of a fabricated structure. The existing experimental evidence indicates that charging may affect the quality of the thin oxide. This is called the *antenna effect* (also called "plasma-induced gate oxide damage"). During metallization, chips are usually processed "from the bulk up", each time adding an additional layer of interconnect. While the metal interconnect chip is being assembled, the interconnect of a net will consist of a number of disconnected pieces of floating metal. Long floating interconnects act as temporary capacitors to store charges gained from the energy provided during fabrication steps such as chemical mechanical polishing (CMP). A random discharge of the floating node due to subsequent process steps could permanently damage transistors, rendering the IC useless [9].

In order to reduce or prevent damage to the gate oxide from the plasma process, and thus to ensure reliability of VLSIs, a circuit layout rule that considers the antenna effect (antenna rule) is employed. The conventional antenna rule restricts a maximum antenna size or antenna ratio allowed for circuit layout. Recent studies show that the damage, considering all plasma-based manufacturing operations, increases in proportion to both the area and the perimeter of the antennas [7]. These models provide a good guideline for router or physical layer EDA tools to help reduce damage from the antenna effect and get higher yield and reliability.

Wang et al. [8] proposed a channel router which considers the antenna effect. They introduce a layer restriction to a conventional channel router, which limits the maximum length of the wires with antenna problems. Shirota et al [7] proposed a router which combines a traditional router with a modification of wires for reducing the antenna effect damage, using a rip-up and reroute method. But this method fixes the antenna after routing; It is not a built-in approach. The diode insertion method is also proposed to fix antenna problem [4]. It is the simplest way to deal with antenna problems by forcing a discharge path. But in today's high-density VLSI layouts, there is simply not enough room for "under-the-wire" diode insertion for all wires. Furthermore, it will cause congestion, add capacitance to the net, reduce room for ECO, and generate leakage power. Thus, people prefer a jumper-based solution to a diode-based solution.

Routing complexity is also an important problem for modern routers. To cope with the increasing complexity, researchers have proposed multilevel approaches to handle the problem [1, 3, 6].

The multilevel framework has attracted much attention in the literature recently. It employs a two-stage technique: coarsening followed by uncoarsening. The coarsening stage iteratively groups a set of circuit components (e.g., circuit nodes, cells, modules, routing tiles, etc.) based on a predefined cost metric, until the number of components being considered falls below a certain threshold. Then, the uncoarsening stage iteratively ungroups a set of previously clustered circuit components and refines the solution by using a combinatorial optimization technique (e.g., simulated annealing, local refinement, etc). The multilevel framework has been successfully applied to partitioning, floorplanning, placement and routing in VLSI physical design.

In this paper, we propose a multilevel router for reducing the antenna effect damage by built-in jumper insertion. The three main features of the proposed method are: (1) a bottom-up approach is used for jumper prediction; (2) a state-of-the-art multilevel routing framework [3] is adopted for run-time speedup and antenna fixing; (3) nets that failed to route or violate antenna rule are routed at the uncoarsening stage for better routing completion. Experimental results show that our algorithm is very efficient.

The rest of this paper is organized as follows. Section 2 describes the antenna effect damage. Section 3 presents our multilevel framework for reducing antenna effect damage. Experimental results are shown in Section 4. Finally, we give concluding remarks in Section 5, as well as our goals for future work.

2 Antenna Effect Damage

The mechanism of antenna damage is not fully understood, but there is experimental evidence indicating when a charging occurs and how it may affect the quality of gate oxide [9]. Charging occurs when conductor layers not covered by a shielding layer of oxide are directly exposed to plasma. The amount of such charging is proportional to this plasma-exposed area. If the charged conductor layers are connected only to the gate oxide, Fowler-Nordheim (F-N) tunneling current will discharge through the thin oxide and cause damage to it.

Process antenna rules adhere to the design requirement that the total charge accumulated on metal connected to a polysilicon gate during any stage of metalization cannot exceed a certain threshold, beyond which the excessive charge accumulation may permanently damage the gate. Let $gate - strength(g, L)$ be the maximum length of a wire of minimum width on layer L that can be directly connected to the gate g without causing an antenna violation. The larger the values of gate-strength, the easier it is to fix the antenna violation. In 0.18-micron technology and above, gate-strength of 1000 microns and above is not

*Yao-Wen Chang's work was partially supported by the National Science Council of Taiwan ROC under Grant No. NSC 91-2215-E-002-038.

†Prof. Sao-Jie Chen is with ECE Department, University of Wisconsin, Madison during his sabbatical leave; his work was partially supported by the National Science Council of Taiwan ROC under Grant No. NSC 91-2215-E-002-042.

uncommon, and fixing by post-processing suffices. In 0.13-micron and below, however, the average and worst-case gate-strength's are substantially reduced (about 20 $\sim$ 100 microns [5]). When the worst-case gate-strength is merely a handful of cellrows, antenna fixing becomes very challenging.

On the other hand, if the amount of charging collected by connected conductor layer patterns could be released through a low impedance path, such as a previously formed diffusion layer pattern (e.g., source/drain), it will not introduce the gate oxide damage.

Jumper insertion is the most popular way to solve the antenna problem. Let us show its usage in the following example. Suppose we have a two-terminal net in which a is the source node and b is the terminal node (See Figure 1(a), (b)). In this case, the approximated gate-strength of b is the sum of the length of segments 4, 5, and 6, which may violate the minimum allowable gate-strength. If we add a jumper at the long segment 5 (See Figure 1(c), (d)), the approximate gate-strength of d is just the sum of the length of segments 8, 9, and 10, which will not violate the minimum allowable gate-strength. And the sum of the length of segments 4, 5, and 6, which will not cause damages to gates because they are floating. Thus, if we add jumpers appropriately, the antenna problem can be easily solved.

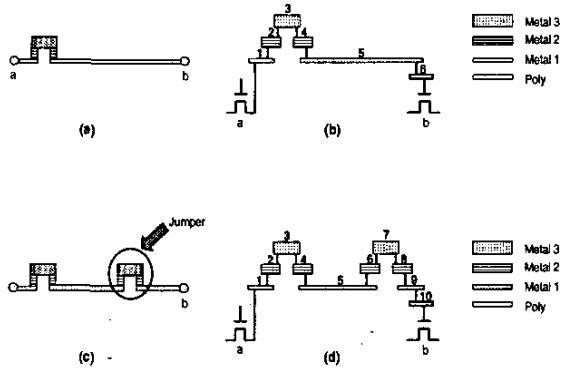

Figure 1: (a) A two pin net. (b) The cross section view of (a). (c) A two pin net with jumper insertion. (d) The cross section view of (c).

3 Multilevel Routing Framework

Our multilevel routing algorithm is inspired by the work of [3]. As illustrated in Figure 2, G_0 corresponds to the routing graph of the level 0 of the multilevel coarsening stage. Before the coarsening process, we first perform the optimal jumper prediction for every net. After that, we can indicate which two-pin nets are needed to insert jumpers by using minimum number of jumpers. Then, our congestion-driven global router first finds routing paths for the *local nets* (or *local 2-pin connections*) (those nets [connections] that entirely sit inside a tile) at each level. After the global routing is performed, we merge four adjacent tiles of G_0 into a larger tile and at the same time perform resource estimation for use at the next level (i.e., level 1 here). Coarsening continues until the number of tiles at a level, say the k-th level, is below a given threshold. After coarsening, in order to break the cumulative length from the gates, we first break in two those segments of two-pin nets that need jumpers, if the length of those segments exceeds the minimum allowable gate-strength. If they have not exceeded the minimum allowable gate-strength, then we try to assign the remaining segments to the highest layer. Segments assigned to the highest layer have the same utility as jumper. If the highest layer is too congested, we assign segments to the lower layer and fix them by adding jumpers near the gate input using distance-aware maze routing after the track assignment phase. After the layer assignment, a track assignment for fast routing completion and antenna avoidance is performed to assign straight segments to underlying routing resources. After that, an antenna check process for every terminal is performed. If nets have antenna violations, we identify them as failed nets that will be routed at the uncoarsening stage. During uncoarsening, the unroutable and antenna-violated nets are considered. Maze routing and rip-up and re-route are performed to refine the routing solution. Then we proceed to the next level (level $k-1$) of uncoarsening by dividing each tile to four finer tiles. The process continues up to level 0 when the final routing solution is obtained.

3.1 Bottom-Up Optimal Jumper Prediction

At the beginning of the interconnect fabrication process, the receiver type terminals are in poly, and driver type terminals are in diffusion. Any incomplete interconnect segments connected to one or more receivers forms the desirable antenna. The risk of gate oxide damage is proportional to the amount of charges collected by the antenna and inversely proportional to the area of the gate oxide. In order to reduce the negative impact of antenna effects, the antenna area of each terminal has to be minimized. Thus, it is natural to formulate the antenna area of a terminal as the interconnect length divided by the number of gates that spreads from it. To minimize the total antenna area, we can break signal wires with antenna violation and routes to the highest levels by inserting a jumper. This reduces the charge amount for violated nets during manufacturing. But each jumper needs at least two vias and will cause delay. Therefore, given a netlist, we first use the minimum-radius minimum-cost spanning tree algorithm (*MRMCST*) proposed in [3] to construct a performance-driven routing tree for each multi-pin net. The MRMCST can minimize its critical path and preserve minimum total wirelength at the same time. Then we decompose each net into 2-pin connections, with each connection corresponding to an edge of the MRMCST. Each net to be connected is composed of a set of terminals, one of which is a driver and the others receivers. In 0.13μ, the short gate-strength results in a dramatic increase in the number of jumpers that need to be added to the wire. Thus, it is very important to minimize antenna area and jumpers at the same time.

In this paper, we proposed a bottom-up approach to predict the jumper positions by inserting a minimum number of jumpers. Given a net and a source, we first hang the net by using the source as root. Then we compute the position of the jumper by accumulating interconnect length from each terminal in bottom-up fashion. To compute it, we have two possible scenarios as shown in Figure 4. Line 4 considers whether the cumulative length $C(v)$ of descendants of the terminal v (e.g., $\sum_{i=1}^{m} d(e_i)$ in Figure 3(a), where $d(e_i)$ denotes the length between the node v and the root of the i-th subtree and m denotes the number of subtrees) is less than the allowable antenna area ($Amax$). For this case, there are two possible scenarios. First, if the sum of $C(v)$ and the length between v and its precedent w (i.e. u(e)) does not exceed the $Amax$, we accumulate the total length and the number of gates to w for further computation. Second, if the sum of $C(v)$ and the length between v and its precedent w exceeds the $Amax$, we add a jumper at the position near v (see Figure 3(b)). After that, if the remaining length connecting to w also exceeds the $Amax$, we add a jumper at the position near w. Line 10 considers whether the cumulative length of descendants ($C(v)$) of the terminal v is greater than the allowable antenna area ($Amax$). For this case, we first rank the length of edge adjacent to v in increasing order. If the cumulative length exceeds the $Amax$, we add a jumper at the edge near v (see Figure 3(c)). If we add a jumper at the edge (v, w), then goto Line 4. The algorithm is summarized in Figure 4.

Given a net with n nodes, the best case time complexity of our jumper-prediction algorithm is $O(n)$ when the cumulative length of descendants for all nodes is less than the allowable antenna area (Step 4-9). And the worst case time complexity happens when the net topology is a star graph (all the nodes are connected to the source directly) and the cumulative length of descendants for a source is greater than the allowable antenna area (Step 10-26). Since the complexity is determined by sorting, the worst case time complexity is $O(n \log n)$.

By this algorithm, we can predict which edges are needed to insert jumpers. After that, we first break in two those segments of two-pin nets that need jumpers for reducing the amount of charging, if the length of those segments exceeds the minimum allowable gate-strength. If they have not exceeded the minimum allowable gate-strength, then we try to assign the remaining segments to the highest layer for "long jumpers". If the highest layer is too congested, we assign segments to the lower layer and fix them by adding jumpers near the gate input using distance-aware maze routing after the track assignment phase.

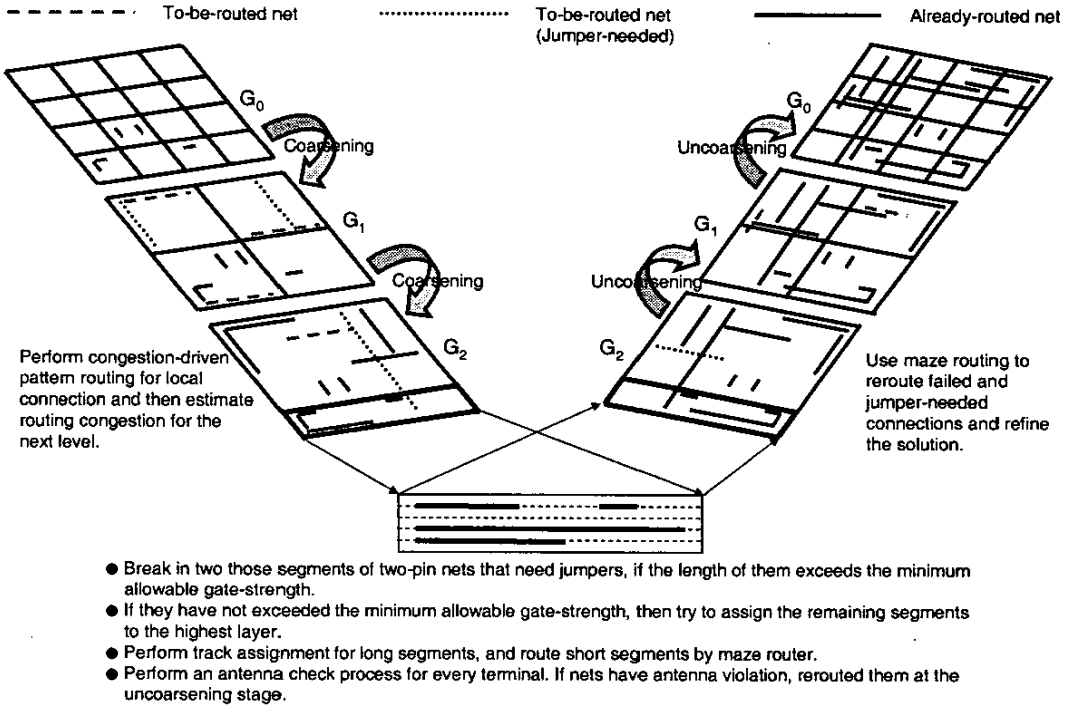

Figure 2: The multilevel framework flow.

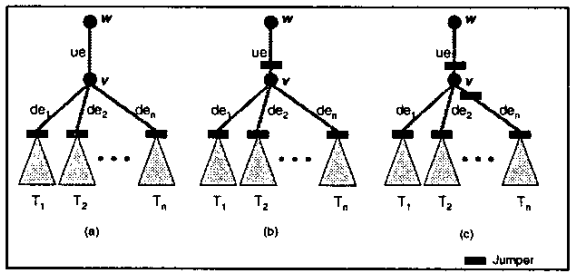

Figure 3: (a) The initial case before the jumper-prediction process (b) The case of $\sum_{i=1}^{m} d(e_i) < Amax$, but $\sum_{i=1}^{m} d(e_i) + u(e) > Amax$, we add a jumper on $u(e)$ then do SecondJumperCheck. (c) The case of $\sum_{i=1}^{m} d(e_i) > Amax$, we accumulate the length of edges adjacent to v in increasing order. If the cumulative length exceeds the $Amax$, we add a jumper at the edge near v.

3.2 Multilevel Routing with Antenna Avoidance

After the jumper positions are predicted, our multilevel framework starts by coarsening the finest tiles of level 0. At each level, tiles are processed one by one, and only local nets (connections) are routed. At each level, a fast congestion-driven pattern routing [3] is used for global routing.

After the global routing is completed, in order to break the cumulative length from the gates, we first break in two those segments of two-pin nets that need jumpers, if the length of those segments exceeds the minimum allowable gate-strength. If they have not exceeded the minimum allowable gate-strength, then we try to assign the remaining segments to the highest layer. If the highest layer is too congested, we assign segments to the lower layer and fix them by adding jumpers near the gate input using distance-aware maze routing after the track assignment phase. Then, an intermediate step of track assignment between coarsening and uncoarsening stages is used for fast routing completion and antenna avoidance.

The track assigner works on a full row or column of the global cell array at a time. To simplify the track assignment problem, we only assign segments which span more than one complete global cell in a row or a column. (We handle short segments during detailed routing.) Therefore, these segments for track assignment are long and may violate the antenna rules. If the segments need jumpers after the jumper-prediction phase, we just add jumpers at the two end sides, floating the segments so that they won't cause damage to gates. Thus, track assignment is a suitable stage to address antenna avoidance.

After the track assignment phase, the actual track position of a segment is known. Thus, we can perform maze routing to complete the routing. After that, we perform an antenna check for every terminal. Since the accumulated gate-strength is kept in every terminal, the antenna-check process can be performed quickly. An accurate damage function which considers all plasma-based manufacturing operations is adopted for the antenna check. If nets have antenna violations, we regard them as failed nets to be routed at the uncoarsening stage. The uncoarsening stage starts to refine each local failed net (connection), remaining from the coarsening stage that has not passed the antenna check. The global router is now changed to the maze router. Also, a distance aware detailed maze routing is performed after the global maze routing. Uncoarsening continues until the first level G_0 is reached and the final solution is found.

4 Experimental Results

We have implemented our multilevel system with antenna avoidance in the C++ language on a 1 GHz SUN Blade 2000 workstation with 1GB memory. See Table 1 for the benchmark circuits. The design rules for wire/via widths and wire/via separation for detailed routing are the same as those used in [3].

Table 1 describes the set of benchmark circuits. In the table, "Size" gives the layout dimensions, "#Layers" denotes the number of routing layers used, "#Nets" represents the number of two-pin connections after

Circuits	Results without antenna avoidance					Our Results				
	Wirelength	#Vias	#Violated Gate	Time	D_{avg}	Wirelength	#Vias	#Violated Gate	Time	D_{avg}
S5378	8.4e7	7451	129	10.6	1258	8.4e7	7533	0	12.5	1271
S9234	6.0e7	6239	75	8.1	1009	6.1e7	6315	0	10.9	1015
S13207	2.3e8	16003	304	22.6	1243	2.4e8	16242	0	29.9	1281
S15850	2.9e8	19126	354	62.6	1253	3.0e8	19534	0	75.8	1279
S38417	8.0e8	49816	683	71.3	1146	8.2e8	50521	0	86.9	1171
S38584	1.1e9	65798	974	255.6	1151	1.2e9	67068	0	307.0	1194

Table 2: Results of wirelength, vias, violated gates, run-time, and delay comparison.

```
Algorithm : JumperPredict(T)
   Input  : MRMCST T and source s;
   Output : Jumper positions of T
begin
1  Pick unvisited v s.t. all descendants of v have been
   visited and let w as the precedent of v.
2  for all node v, set C(v) = 0 and visit(v) = 0
3  while w ≠ s do
4     if (C(v) < Amax)
5        if (C(v) + leng(v, w) > Amax)
6           JumperAdded(v, Up);
7           SecondJumperCheck(leng(v, w) − (Amax
            −C(v)))
8        else
9           AccumulateSegment();
10    else
11       InsHeap(v, w, leng(v, w));
12       UpJumper=0;
13       while HeapSize ≠ 0
14          AccLen=AccLen+PopHeap()→length;
15          if (AccLen>Amax)
16             if (PopHeap()→node ≠ w)
17                JumperAdded(v, Dn);
18             else
19                JumperAdded(v, Up);
20                UpJumper= 1;
21          HeapSize−−;
22       if (UpJumper== 1)
23          SecondJumperCheck(leng(v, w) − 1);
24       else
25          AccumulateSegment();
26    visit(v)=1;
end
```

Figure 4: Algorithm for Jumper Prediction.

Circuits	Size (μm)	#Layer	#Nets	#Diffusions	#Gates
S5378	4330x2370	3	3124	1694	3040
S9234	4020x2230	3	2774	1486	2699
S13207	6590x3640	3	6995	3781	6781
S15850	7040x3880	3	8321	4472	8094
S38417	11430x6180	3	21035	11309	20901
S38584	12940x6710	3	28177	14753	27836

Table 1: The benchmark circuits.

net decomposition, "#Diffusions" represents the number of diffusions (drivers), and "#Gates" represents the number of receiver type terminals.

Experimental results on wirelength, vias, run-time, violated gates, and delay are listed in Table 2, where "D_{avg}" represents the average net delay. To perform experiments on timing-driven routing, we used the same resistance and capacitance parameters as those used in [3] for comparison. A via is modeled as the Π-model circuit, with its resistance and capacitance being twice those of a wire segment. And as mentioned in [5], we set $Amax$ to $100\mu m$ in this experiment. Compared with [3], the experimental results show that our router reduced 100% antenna-violated gates and resulted in fewer increases in wirelength, vias, run-time, and delay.

5 Conclusion

In this paper, we have proposed a novel framework for multilevel full-chip routing with antenna avoidance using a built-in jumper insertion approach. The experimental results have shown that our algorithm is very efficient and effective. Our future work lies in multilevel routing considering other nanometer electrical effects.

References

[1] J. Cong, M. Xie, and Y. Zhang, "An enhanced multilevel routing system," *Proc. ICCAD*, pp. 51-58, Nov. 2002.

[2] H. T. Heineken, J. Khare, W. Maly, P. K. Nag, C. Ouyang, and W. A. Pleskacz, "CAD at the design-manufacturing interface," *Proc. DAC*, pp. 321-326, Jun. 1997.

[3] T.-Y. Ho, Y.-W. Chang, S.-J. Chen, and D. T. Lee, "A Fast Crosstalk- and Performance-Driven Multilevel Routing System," *Proc. ICCAD*, 2003.

[4] L.-D. Huang, X.-P. Tang, H. Xiang, D. F. Wong, and I.-M. Liu, "A Polynomial Time Optimal Diode Insertion/Routing Algorithm for Fixing Antenna Problem," *Proc. DATE*, pp. 470-475, 2002.

[5] H. K.-S. Leung, "Advanced routing in changing technology landscape," *Proc. ISPD*, pp. 118-121, Apr. 2003.

[6] S.-P. Lin and Y.-W. Chang, "A novel framework for multilevel routing considering routability and performance," *Proc. ICCAD*, pp. 44-50, Nov. 2002.

[7] H. Shirota, T. Sadakane, M. Terai, and K. Okazaki, "A new router for reducing "Antenna effect" in ASIC design," *Proc. CICC*, pp. 27.5.1 - 27.5.4, Sep. 1998.

[8] K. P. Wang, M. Marek-Sadowska, and W. Maly, "Layout design for yield and reliability," *Proc. PDW*, pp. 190-197, Apr. 1996.

[9] H. Watanabe, et.al., "A wafer level monitoring method for plasma-charging damage using antenna PMOSFET test structure," *IEEE Trans. Semiconductor manufacturing*, vol. 10, no. 2, May. 1997.