Multilevel Full-Chip Routing With Testability and Yield Enhancement

Katherine Shu-Min Li, *Member, IEEE*, Yao-Wen Chang, *Member, IEEE*, Chung-Len Lee, *Member, IEEE*, Chauchin Su, *Member, IEEE*, and Jwu E. Chen, *Member, IEEE*

***Abstract*—We propose a multilevel full-chip routing algorithm that improves testability and diagnosability, manufacturability, and signal integrity for yield enhancement. Two major issues are addressed. 1) The oscillation ring test (ORT) and its diagnosis scheme for interconnects based on the popular IEEE Standard 1500 are integrated into the multilevel routing framework to achieve testability enhancement. We augment the traditional multilevel framework of coarsening and uncoarsening by introducing a preprocessing stage that analyzes the interconnect structure for better resource estimation before the coarsening stage, and a final stage after uncoarsening that improves testability to achieve 100% interconnect fault coverage and maximal diagnosability. 2) We present a heuristic to reduce routing congestion to optimize the multiple-fault probability, chemical–mechanical polishing- and optical proximity correction-induced manufacturability, and crosstalk effects, for yield improvement. Experimental results on the Microelectronics Center for North Carolina benchmark circuits show that the proposed ORT method achieves 100% fault coverage and the optimal diagnosis resolution for interconnects. Further, the multilevel routing algorithm effectively balances the routing density to achieve 100% routing completion.**

***Index Terms*—Interconnect, routing, signal integrity, yield.**

I. Introduction

With ever-decreasing feature sizes and increasing chip dimensions, the integration complexity in system-on-a-chip (SOC) designs grows dramatically [1]. The high integration complexity is not only caused by the huge number of transistors and interconnects fabricated in a single chip but also by the modern SOC design issues in testability, manufacturability, and signal integrity. In particular, interconnect delay dominates the circuit performance for nanometer IC designs. Therefore, it is desirable to handle the large-scale interconnect integration considering testability, manufacturability (including process variation control, chemical–mechanical polishing (CMP), optical proximity correction (OPC), etc.), and signal integrity issues simultaneously.

Testability and diagnosability are very important issues for interconnect design in SOC ICs. Many research works on interconnect testing can be found in the literature. Earlier works on interconnect testing were targeted for board-level testing. However, it is very difficult to apply these interconnect testing methods under the SOC environment without design-for-testability support. The popular IEEE Standard 1500 [2] provides a structural support for core testing as well as interconnect testing in SOC. IEEE Standard 1500 SOC test environment consists of a centralized test access mechanism (TAM) and wrappers around cores. The TAM defines the test control, whereas the wrappers provide a standardized interface for test data transmission. An interconnect oscillation ring test (IORT) [3] method was proposed to detect not only stuck-at and open faults but also delay and crosstalk glitch faults. Many testing and diagnosis problems are incurred by particular interconnect structures, which can be partly solved by a carefully designed router. Furthermore, to reduce the probability of multiple faults, it is desirable to reduce wiring congestion in a specific area. This approach is specifically important as the probability of back-end-of-line defects (i.e., high-resistance via and interconnect defects) increases [4]. Therefore, the wire congestion is directly related to multiple fault probability and should be addressed in the routing stage.

The manufacturing process increasingly constrains physical layout design [5]. With the move to nanometer design, the semiconductor industry has seen an increase in defects and a decrease in yield. Defects arise from sources such as lithography errors, environmental variations, and process variations, and contribute to yield loss. Since defect mechanisms are ever more dependent on design, yield must now be treated throughout the design flow. To address yield loss effectively, a systemic solution in the design flow [6] is required.

One source of process variation comes from the CMP process [7], [8]. CMP-induced variation is kept within acceptable limits by controlling local feature (interconnect) density, relative to a process-specific "window size," to achieve global planarization for manufacturability and performance. As a result, balancing interconnect density minimizes the CMP-induced variation, and thus routing plays an important role in determining CMP variation.

Manuscript received December 16, 2004; revised June 20, 2005 and September 16, 2006. The work of K. S.-M. Li was supported in part by the National Science Council of Taiwan under Grant NSC 95-2218-E-110-009. The work of Y.-W. Chang was supported in part by the National Science Council of Taiwan under Grants NSC 93-2215-E-002-009, NSC 93-2215-E-002-029, and NSC 93-2752-E-002-008-PAE. The work of C. Su was supported in part by the National Science Council of Taiwan under Grants NSC 95-2221-E-009-328 MY3 and NSC 95-2221-E-009-334. This paper was recommended by Associate Editor T. Yoshimura.

K. S.-M. Li is with the Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung 804, Taiwan, R.O.C. (e-mail: smli@cse.nsysu.edu.tw).

Y.-W. Chang is with the Department of Electrical Engineering and Graduate Institute of Electronics Engineering, National Taiwan University, Taipei 106, Taiwan, R.O.C. (e-mail: ywchang@cc.ee.ntu.edu.tw).

C.-L. Lee is with the National Chiao Tung University, Hsinchu 300, Taiwan, R.O.C.

C. Su is with the Department of Electrical and Control Engineering, National Chiao Tung University, Hsinchu 300, Taiwan, R.O.C.

J. E. Chen is with the Department of Electrical Engineering, National Central University, Jhungli 32001, Taiwan, R.O.C.

Digital Object Identifier 10.1109/TCAD.2007.895587

0278-0070/$25.00 © 2007 IEEE

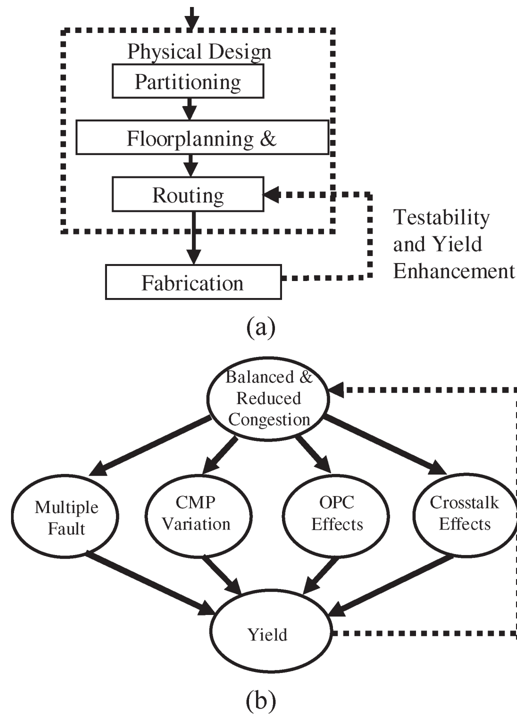


Fig. 1. (a) Testability and yield enhancement in the routing stage. (b) Reducing and balancing routing congestion improves multiple fault probability, CMP-induced variation, OPC effects, and crosstalk effects, all of which lead to yield enhancement.

OPC is one of the most effective methods adopted to compensate for the light diffraction effect, typically used as a postlayout process to improve manufacturability. Recently, Huang and Wong [9] proposed an algorithm that considers the OPC effect during routing by utilizing a symmetrical property. However, the process is time-consuming, and its results are still limited by the original layout quality. Again, balancing interconnect density can improve the OPC effects efficiently, since it may not always be possible to include OPC patterns in an overly congested area.

Signal integrity is also an important factor that affects yield in nanometer IC technology [1], and the crosstalk fault is a major source that contributes to the loss of signal integrity in nanometer IC [10]. The crosstalk effect is caused by coupling capacitance between adjacent wires, in which a signal change on an aggressor net interferes with the signal on a victim net. There are two types of crosstalk effects: 1) glitch and 2) delay. The glitch may induce malfunctioning in the logic values of circuit nodes, whereas the crosstalk-induced delay slows down both signals. In a heavily congested routing region, smaller wire spacing leads to larger coupling capacitance and thus greater crosstalk. Thus, reducing routing congestion also reduces crosstalk and the probability of crosstalk faults.

In this paper, we simultaneously consider the interconnect design issues of testability and diagnosability, manufacturability, and signal integrity in the routing stage for yield improvement [see Fig. 1(a)]. Traditionally, these issues are tackled at the postlayout stage. With the increasing design complexity, it is very difficult and even infeasible to handle those issues at the postlayout stage when most interconnect layouts are fixed and not flexible to be changed. In particular, the design-for-manufacturability issues can all be improved through reducing and balancing the routing congestion [see Fig. 1(b)]. Therefore, it is desirable to develop a congestion-driven routing algorithm for yield improvement.

Since our goal is to enhance testability and yield in the multilevel full-chip routing, we shall first review some important routing work. Traditionally, the complex routing problem is often solved by using the two-stage approach of global routing, followed by detailed routing. Global routing first partitions the routing area into tiles and decides tile-to-tile paths for all nets, whereas detailed routing assigns actual tracks and vias for nets. Many routing algorithms adopt a flat framework of finding paths for all nets. Those algorithms can be classified into sequential and concurrent approaches. Sequential routing algorithms include maze-searching approaches [11], [12] and line-searching approaches [13], which route net by net. Most concurrent algorithms apply network-flow [13] or linear-assignment formulation [14], [15] to route a set of nets at one time.

The major problem of the flat framework lies in their scalability for handling larger designs. As technology advances, technology nodes are getting smaller, and circuit sizes are getting larger. To cope with the increasing complexity, researchers proposed the use of hierarchical approaches in handling the problem by dividing a routing region into subregions and independently routing each subregion. Marek-Sadowska [15] proposed a hierarchical global router based on linear assignment. Chang *et al.* [14] applied linear assignment to develop a hierarchical concurrent global detailed router for FPGAs.

The two-level hierarchical routing framework, however, lacks information for the interactions among the subregions and is thus still insufficient in handling the dramatically growing complexity in current and future IC designs [16]. Therefore, it is desired to employ more levels of routing for very large-scale IC designs. The multilevel framework has attracted much attention in the literature recently. It employs a two-stage technique, namely, coarsening, followed by uncoarsening. The coarsening stage iteratively groups a set of circuit components (e.g., circuit nodes, cells, modules, and routing tiles) based on a predefined cost metric until the number of components being considered is smaller than a given 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). The multilevel framework has been successfully applied to the VLSI physical design. For example, the famous multilevel partitioners ML [17] and hMETIS [18], the multilevel placer mPL [19], and the multilevel floorplanner/placer MB*-tree [20] all show the promise of the multilevel framework for large-scale circuit partitioning, placement, and floorplanning. A framework similar to multilevel routing was presented in [21] and [22]. Lin *et al.* [22] and Hayashi and Tsukiyama [21] presented hybrid hierarchical global routers for multilayer VLSIs, in which both the bottom-up (coarsening) and top-down (uncoarsening) techniques were used for global routing.

Recently, Cong *et al.* proposed a pioneering multilevel global-routing approach for large-scale full-chip

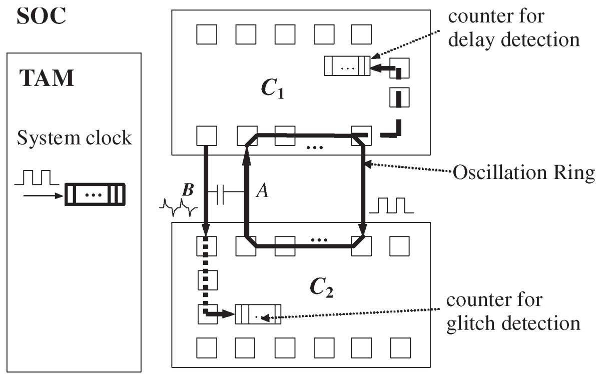


Fig. 2. Test architecture for delay and crosstalk detection and delay measurement.

routability-driven routing [16]. Cong *et al.* later proposed an enhanced multilevel routing system, which is named MARS [23]. Lin and Chang proposed a novel multilevel framework for full-chip routing, which considers both routability and performance [24]. This framework integrates global routing, detailed routing, and resource estimation together at each level, leading to more accurate routing resource estimation during coarsening and thus facilitating the solution refinement during uncoarsening. Recently, Ho *et al.* proposed yet another multilevel framework by introducing an intermediate layer and track assignment stage between coarsening and uncoarsening to handle crosstalk minimization [26].

In this paper, we propose a multilevel full-chip routing framework considering testability and diagnosability, multiple fault probability, manufacturability, and signal integrity simultaneously. Different from the previous works, our approach has the following distinguished features.

1) Simultaneously consider testability and diagnosability, multiple fault probability, manufacturability, and signal integrity in the multilevel routing framework.
2) Propose a new testability-driven multilevel routing framework, consisting of a preprocessing stage for *o*scillation *r*ing *t*est (ORT) generation for *i*nterconnect (IORT), a coarsening stage, an intermediate stage for optimization, an uncoarsening stage, and a postprocessing stage to process diagnosis patterns for interconnects (i.e., interconnect oscillation ring diagnosis, IORD).
3) Provide testability and yield enhancement solutions in the routing stage to both diagnose interconnects and improve density flexibility.
4) Present heuristics to balance and reduce congestion in routing for yield improvement (by reducing multiple fault probability, CMP variation, OPC effects, and crosstalk).

Experimental results on the Microelectronics Center for North Carolina (MCNC) benchmark circuits show that the proposed ORT method achieves 100% fault detection coverage and the optimal diagnosis resolution for interconnects. Further, the multilevel routing algorithm effectively balances the routing density to achieve 100% routing completion. Experimental results show that our method significantly improves routing quality for testability and yield enhancement. Compared with [24], the experimental results show that our router improves the maximal congestion by 1.24–6.11× in runtime speedup by 1.08–7.66× and improves the average congestion by 1.00–4.52× with the improved congestion deviation by 1.37–5.55×. Compared with [26], the experimental results also show that our router improves the maximal congestion by 1.54–1.84× and the average congestion by 1.17–1.34× with the congestion deviation being improved by 1.13–1.63×.

This paper is organized as follows. Section II gives preliminary backgrounds on ORT and diagnosis and the CMP and crosstalk models. Section III presents the integrated multilevel routing framework. Experimental results are reported in Section IV, and concluding remark follows in Section V.

II. Preliminaries

In this section, we give preliminary backgrounds on ORT and diagnosis, and the CMP and crosstalk models.

A. ORT Architecture for Interconnect

We first discuss the ORT for interconnects. ORT is an efficient method to detect faults in SOC interconnects [3]. An oscillation ring (OR) is a closed loop of a circuit under test with an *odd number of signal inversions*. Once the ring is constructed during test mode, oscillation signal appears on the ring. Fig. 2 illustrates a global counter-based test architecture for both delay and crosstalk glitch detection for SOC ICs. This test architecture implements the IEEE Standard 1500 core test standard, in which each input/output pin of a core is attached with a *wrapper cell*, and a centralized TAM is provided to coordinate all test process. In addition to the normal input/output connections, all wrapper cells in a core can also be connected with a shift register, which is usually referred to as a scan path, to facilitate test access.

A modified wrapper cell design has been proposed to provide extra connections and inversion control so that the ORs can be constructed through the wires and the boundary scan paths in cores [3]. For example, the ORT architecture in Fig. 2 consists of one OR and a neighboring net, and two scan paths in cores C_1 and C_2 are part of the OR.

This test architecture can detect stuck-at, open, delay, and crosstalk glitch faults. If an OR fails to oscillate, it implies that there exists stuck-at or open fault(s) in the OR. The period of the oscillation signal can also be measured by using a delay counter in a core to test delay faults, and a similar approach can be used for crosstalk glitch detection.

The graphics will stay in the "second" column, but you can drag them to the first column. Make the graphic wider to push out any text that may try to fill in next to the graphic.

A local counter is included in each core, and a central counter is in the TAM of the chip. The central counter in the TAM is enabled by signal $OscTest$ and triggered by the system clock. A local counter is connected to one wrapper cell in each core; however, it can be accessed by every wrapper cell through the wrapper cell chain. When an OR passes a core, an internal scan path is formed to connect the oscillation signal to the local counter. For example, consider core C_1, in which the OR pass by (see Fig. 2). The oscillation signal is fed to the

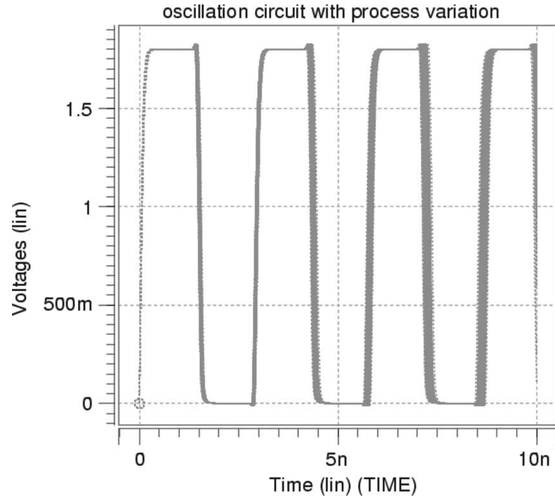


Fig. 3. Simulation waveform with process variation effects on the ORT scheme.

local counter through a series of modified wrapper cells that are configured as SI → SO. When an oscillation test session starts ($OscTest = 1$), the TAM enables its own central counter as well as all local counters in cores. After the central counter in the TAM counts to a specific number n, the oscillation test session terminates and all local counters are disabled ($OscTest = 0$). Then, all the local counter contents can then be scanned out to automatic test equipment (ATE) for inspection.

Assume that m ORs are tested. Let the frequency of the system clock be f, and the delay counter contents of the rings be $n_1, n_2, \ldots, n_m$, respectively. An estimation of the ith ring's oscillation frequency f_i can be approximated by

$$f_i = f \times n_i / n. \tag{1}$$

Since the frequency of each ring is predetermined during the design phase, a delay fault can thus be detected and measured as compared with the result of the counters.

B. Process Variation Effects on Oscillation Signals

In order to consider the process variation effect on this proposed ORT scheme, we conducted the experiments for a ring consisting of seven inverters (plus transmission gates) and 20-μm lines. The Monte Carlo simulation was conducted by changing the W/L ratio of all transistors and the R, C parameters of the nets. The mean is the nominal value, whereas the distribution is Gaussian with $3\sigma = 20\%$ of the nominal value. Four hundred simulation runs were performed, and the simulation results were shown in Fig. 3, in which all oscillation signals started at time 0.

At the end of the first cycle, there is a small variation in the cycle period, and the variations are less than 1.2% of the nominal cycle period. The simulation results show that: 1) this scheme can oscillate with an odd number of inversions and 2) the process variation effects with 20% variance contribute to less than 1.2% in the frequency and oscillation period.

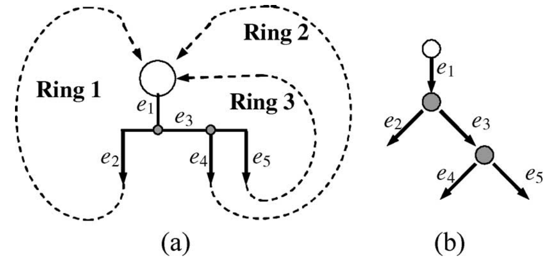


Fig. 4. (a) Hypernet and (b) diagnosis graph model.

C. Interconnect Model in ORT

In this section, we show the interconnect model in the ORT scheme for interconnect detection. We give a graph modeling for interconnect detection. A multiterminal net is usually modeled by a *hypernet*. The circuit structure of an SOC can be directly transformed into a *hypergraph*, in which each vertex denotes a pin, whereas each hypernet represents a signal net. However, this graph model is not good enough for the ORT problem, as two branches of a net should belong to two different rings, and they cannot be simultaneously tested [3]. Therefore, we consider each branch of a hypernet separately, instead of treating them as a whole. Each branch of a hypernet thus corresponds to a two-pin net, which connects the source vertex to one of its sink vertices. An n-terminal hypernet is thus broken into $(n - 1)$ two-pin nets. The result is a normal graph $G = (V, E)$, where E is the set of two-pin nets.

A complete test for all interconnections is thus reduced to the problem of finding a set of rings that cover all edges corresponding to the interconnection structure in the graph G. This is equivalent to finding a set of subcircuits (rings) $R = \{G_1, G_2, \ldots, G_n\}$, such that:

1) $\forall G_i, G_i \subseteq G, G_i = (V_i, E_i)$, G_i is a ring;
2) $\bigcup_{i=1}^{n} E_i = E$.

If delay fault is considered, signal delay on each net along the ring should also be considered. The period of the oscillation signal is thus the summation of the path delay on all wires and scan paths. A large delay on an interconnect wire can be detected by observing the frequency of an oscillation signal that passes the wire under test. The detection can be masked by the variation of delays on other wires in the same ring, and thus the control of process variation is crucial for the correct detection.

D. Interconnect Diagnosis With ORT

Diagnosis is the process of locating the exact fault site. The ORT can also be used for interconnect diagnosis. For interconnect diagnosis, the two-pin net model is also not sufficient. Consider the four-terminal net shown in Fig. 4(a), which is divided into five edge segments e_1 to e_5. If edge e_1 is faulty, all three rings will not oscillate correctly. A faulty e_3 affects rings 2 and 3, whereas faults on edges e_2, e_4, and e_5 affect rings 1, 2, and 3, respectively. For diagnosis purposes, all five segments are different.

From the aforementioned discussion, hypernets cannot be used for diagnosis. Therefore, the interconnect structure is transformed into a diagnosis graph model as follows. The scan

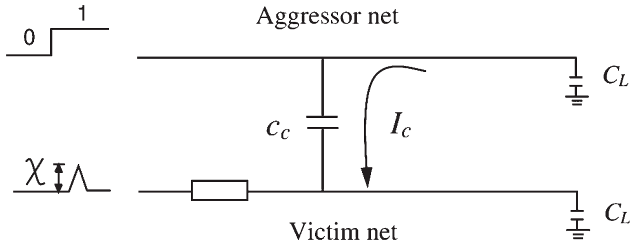


Fig. 5. Noise due to crosstalk-induced current.

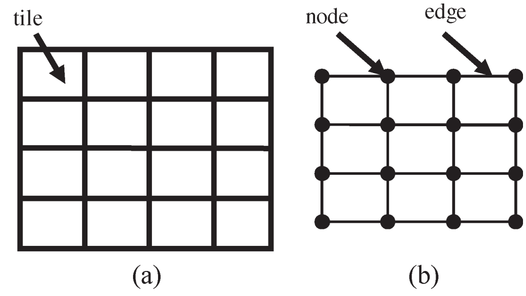


Fig. 6. (a) Partitioned layout. (b) Routing graph.

path and wrapper cells in a core are lumped into a single *terminal node*, as we assume that they are fault free. The fanout points of a hypernet form dummy *intermediate nodes*, and a wire segment connecting two nodes is an edge. For example, the diagnosis graph model for the hypernet in Fig. 4(a) is shown in Fig. 4(b), in which the white node is a terminal node and gray nodes are intermediate nodes. An edge is the smallest unit of a wire segment to be uniquely diagnosed. From the aforementioned discussion, any stem affects all the downstream nodes and edges.

E. CMP Model

With the IC industry rapidly switching to copper interconnects, the lack of etching techniques to remove copper has led manufacturers to use CMP to remove excess copper and associated barrier metals. The challenge is to optimize *CMP process uniformity*, which means minimizing copper dishing and dielectric erosion and ensuring that structures with widely varying feature densities are polished evenly. In order to improve manufacturability, the variation induced by CMP should be kept within acceptable limits.

Several models for oxide planarization via CMP have been proposed in [7]. Among them, the model in [27] is neither computationally expensive nor difficult to calibrate. In this model, the interlevel dielectric (ILD) thickness z at location (x, y) is calculated as follows:

$$z = \begin{cases} z_0 - \left(\frac{K_i t}{\rho(x,y)}\right), & t < (\rho_0 z_1)/K \\ z_0 - z_1 - K_i t + \rho_0(x,y) z_1, & t > (\rho_0 z_1)/K \end{cases}. \quad (2)$$

In this model, the most important factor that determines the value of z is the effective pattern density $\rho(x, y)$. In other words, we improve the variation of dielectric thickness z by keeping the effective pattern density $\rho(x, y)$ relatively constant across the routing surface. Balancing the wiring congestion can effectively achieve this goal.

F. Crosstalk Model

In this section, we show the crosstalk model. Fig. 5 shows the noise model. The noise χ on the victim net is induced by a rising transition on the aggressor net through the coupling capacitance c_c. The coupling capacitance, which is defined as follows, is proportional to the fringing capacitance c_f and the coupling length l_c, and inversely proportional to the distance d between the aggressor and the victim nets

$$c_c = \frac{l_c c_f}{d}. \quad (3)$$

In (4), consider a wire $e = (u, v)$, where u and v are two nodes in a routing tree. Let the length of the wire segment e be l_e, and $T(v)$ be the subtree rooted at v. $I_{T(v)}$ is the total downstream current seen at v and is the current induced by aggressor nets on downstream wires of v. The current on a unit-length wire induced by aggressor nets is $i_0 = \lambda p c_0$ [28], where λ is the fixed ratio of coupling to total wire capacitance, p is the slope (i.e., power supply voltage over input rise time) of the aggressor net's signal, and c_0 is the unit-length wire capacitance. In deep submicron process, a major part of the wire capacitance is attributed to the coupling capacitance if the wire spacing is kept minimum (e.g., $\lambda = 0.7$ [28]). The resulting noise $\chi(u, v)$ induced from the coupling current is the voltage pulse coupled from aggressor nets in the victim net for a wire segment $e = (u, v)$. The induced noise can be expressed as

$$\chi(u, v) = r_b I_{T(v)} + r_0 l_e \left(\frac{i_0 l_e}{2} + I_{T(v)}\right). \quad (4)$$

The crosstalk effect can be effectively reduced by increasing wire spacing, which decreases the unit-length coupling capacitance according to (3). To achieve this goal, a router should constrain or limit the coupled number and length of adjacent wires in any area. Therefore, it is desirable to minimize the routing congestion to improve the crosstalk effects.

G. OPC-Driven Routing Mechanism

OPC offers basic corrections and a useful amount of yield improvement. The goal of OPC is to produce smaller features in an IC to enhance the "printability" of a wafer pattern. OPC applies systematic changes to photomask geometries to compensate for nonlinear distortions caused by optical diffraction and resist process effects. Specifically, these distortions include linewidth variations, dependent on pattern density; the variations affect a device's operation speed and line-end shortening that breaks connections to contacts. A mask incorporating OPC is thus a system that negates undesirable distortion effects during pattern transfer.

OPC works by making small changes to the IC layout that anticipate the distortions. To compensate for line-end shortening,

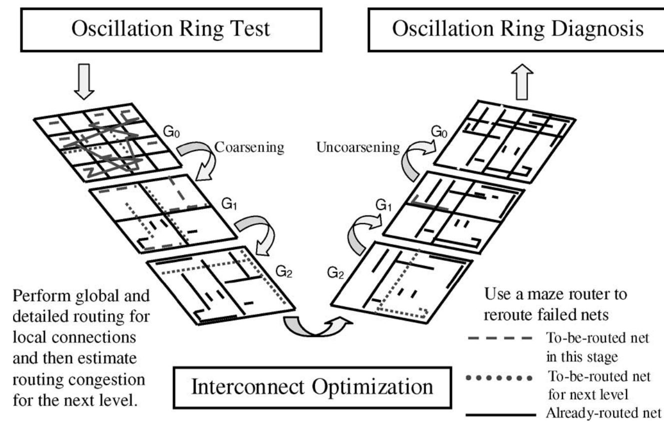


Fig. 7. Integrated multilevel routing framework.

the line is extended using a hammerhead shape that results in a line in the resist that is much closer to the original intended layout. To compensate for corner rounding, serif shapes are added to (or subtracted from) corners to produce corners in the silicon that are closer to the ideal layout.

The inclusion of OPC patterns in the interconnect structure may not be always possible when the wires are heavily congested, as the minimum wire spacing could be violated due to the added pattern. A balanced and less congested routing area thus increases the probability of successful OPC.

III. Integrated Multilevel Routing Framework

We propose a new integrated multilevel routing framework in this section. The router considers routability, performance, testability, diagnosability, congestion, process variation, and crosstalk simultaneously.

The ORs for test are based on circuit connectivity, and thus they can be constructed before routing. However, when delay fault is considered, the routing structure must also be considered, since the wire delay is mainly decided by the wire length. On the other hand, the diagnosis process has to consider the actual net layout, and they must be considered after the routing process.

A. Routing Model

Our global routing algorithm is based on a graph search technique guided by the congestion information associated with routing regions. The router assigns higher costs to route nets through congested areas (or those of higher delay and/or crosstalk costs) to balance the net distribution among routing regions. Before we apply the graph search technique to multilevel routing, we first model the routing architecture as a graph such that the graph topology represents the chip structure. Fig. 6 illustrates the routing graph model.

For the modeling, we first partition a chip into an array of rectangular subregions. These subregions are called *global cells* (GCs). A node in the graph represents a GC in the chip, and an edge denotes the boundary between two adjacent GCs. Each edge is assigned a weight/capacity according to the physical area or the number of tracks of a GC. The graph is used to represent the routing area and is called a *multilevel routing graph*, which is denoted by G_k, where k is the level ID. A global router finds GC-to-GC paths for all nets on a routing graph to guide the detailed routing. The goal of global routing is to route as many nets as possible while meeting the capacity constraint of each edge and any other constraints, if specified.

As the process technology advances, multiple routing layers are possible. The number of layers in a modern chip can be more than eight. Wires in each layer can run either horizontally (H) or vertically (V) in a grid style.

As illustrated in Fig. 7, G_0 corresponds to the routing graph of the level 0 of the multilevel coarsening stage. At each level, our global router first finds routing paths for the *local nets* (or *local two-pin connections*) (those nets that entirely sit inside a GC). After the global routing is performed, we merge 2×2 of GC into a larger G_i 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 GCs at a level, for example, the kth level, is below a threshold. The uncoarsening stage tries to refine the routing solution of the unassigned segments of the level k. During uncoarsening, the unroutable nets are performed by point-to-path maze routing and rip-up and reroute to refine the routing solution. Then, we proceed to the next level (level $k-1$) of uncoarsening by expanding each G_k to four finer G_{k-1}'s. The process continues until we reach level 0 when the final routing solution is obtained.

B. Testability-Aware Multilevel Routing

In the coarsening stage of multilevel routing, shorter nets are routed first, and a congestion-driven heuristic is used to guide a pattern router. For all the nets that can be successfully routed, both global route and detailed route are conducted. All the nets that fail to complete will be handled at the uncoarsening stage. At the uncoarsening stage, the failed nets are routed by a global router with a different cost function to avoid heavily congested area, and a detailed maze router is used to determine the final routing paths.

In addition to the traditional multilevel framework, we incorporate an ORT in the preprocessing stage to guide the resource estimation for interconnects and 100% fault detection coverage, an intermediate stage for interconnect optimization, and an oscillation ring diagnosis in the postprocessing stage to guarantee maximal interconnect diagnosability (see Fig. 7).

C. Diagnosability-Aware Routing Structure

The minimum spanning tree (MST) topology leads to the minimum total wire length, and thus congestion is often easier to be controlled for MST than other topologies. This topology may result in longer critical paths and thus degrade circuit performance. In contrast, a shortest path tree may result in the best performance, but its total wire length (and congestion) may be significantly larger than that constructed by the MST algorithm.

The diagnosis problem also affects the routing structure. For instance, consider the four-terminal net example shown in Fig. 8. With the *spanning tree* connection given in Fig. 8(a),

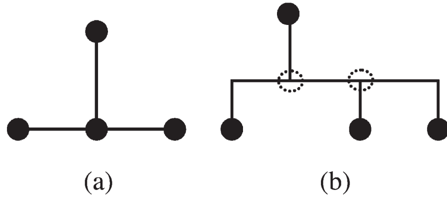


Fig. 8. Two routing trees. (a) Spanning tree with three segments. (b) Steiner tree with the minimum number of intermediate nodes, resulting in five segments.

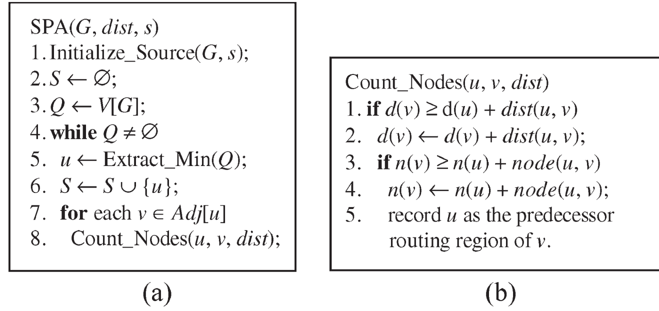


Fig. 9. (a) Shortest path algorithm. (b) $n(v)$ computation.

there are three different net segments to be diagnosed. On the other hand, as the diagnosis graph model shown in Fig. 4(b), for the *Steiner tree* connection given in Fig. 8(b), there are two intermediate nodes (indicated by the two dotted circles) and thus five net segments to be diagnosed. In general, a spanning tree connection employed fewer wire segments to be diagnosed, and thus it is favored in our router. Our algorithm first constructs the MST structure whenever possible, which is best for diagnosability. Otherwise, it will find a routing tree with the least number of intermediate nodes.

In order to route a net with the minimum number of intermediate branch nodes and the shortest path, we apply the algorithm shown in Fig. 9(a) for the routing tree construction. The algorithm, which is based on Dijkstra's shortest path algorithm, finds the shortest path with the minimum number of intermediate nodes. It associates each basic detailed routing region u with two labels: $d(u)$ and $n(u)$, where $d(u)$ is the distance of the shortest path from source s to u, and $n(u)$ is the minimum number of intermediate nodes along the shortest path from s to u. Initially, $d(u) = \infty, n(u) = \infty, \forall u \neq s, d(s) = 0$, and $n(s) = 0$. The computation of label d's is the same as the original Dijkstra's algorithm. The computation of $n(v)$ is shown in Fig. 9(b), where $dist(u, v)$ and $node(u, v)$ are the distance and the number of intermediate nodes between nodes u and v, respectively.

D. Cost Metric for Routing Density Control

A router that incurs imbalanced routing density may degrade system performance in many ways.

1) Crosstalk effects are the results of signal coupling between adjacent wires, and the coupling capacitance is usually inversely proportional to the distance between wires. In a heavily congested area, the distance between adjacent wires is small and thus the probability of crosstalk faults is increased.
2) Physical defects in a congested area may create multiple faults, which are difficult to be detected and diagnosed.
3) Process variation due to CMP effects is usually caused by unbalanced routing congestion or density.
4) OPC pattern compensation is usually difficult and even impossible to be performed in an overly or imbalanced dense area.

Therefore, it is desirable to simultaneously balance and reduce routing congestion/density in all areas for router design. The global routing is based on the approach used in the pattern router [25] and first routes local nets on the tiles of level 0. Let the multilevel routing graph of level i be $G_i = (V_i, E_i)$. Let $R_e = \{e \in E_i | e \text{ be the edge chosen for routing}\}$. In order to balance the routing density, we use the following cost function $\alpha : E_i \to R$ to guide the routing

$$\alpha(R_e) = \sum_{e \in R_e} c_e. \tag{5}$$

The parameter c_e is the congestion of edge and it is defined as

$$c_e = \begin{cases} \frac{1}{2^{\lceil (p_e/t) - d_e \rceil}}, & d_e < (p_e/t) \\ 1, & d_e \geq (p_e/t) \end{cases}$$

where p_e and d_e are the capacity p_e and the number of nets assigned to edge $e(d_e)$, respectively. The parameter t is used to define the target level of the maximum density, and determined either by the user or by averaging over all routing areas. For example, if the goal is to make the average routing density to be half of the maximum acceptable density, then t is set to 2.

After the global routing is completed, we perform detailed routing with the guidance of the global-routing results and find a real path in the chip. Our detailed router is based on the maze-searching algorithm. Pattern routing uses an L-shaped or a Z-shaped route to make the connection, which gives the shortest path length between two points. Therefore, the wire length is minimized, and we do not include wire length in the cost function at this stage. We measure the routing congestion based on the commonly used channel density. After the detailed routing finishes routing a net, the channel density associated with an edge of a multilevel graph is updated accordingly.

If both L-shaped and Z-shaped patterns of our global router fail, we give up routing the connection, and an overflow occurs. We refer to a *failed net* (*failed connection*) as that causes an overflow. The failed nets (connections) will be reconsidered (refined) at the uncoarsening stage.

The uncoarsening stage starts to refine each local failed net (connection), left from the coarsening stage. The global router is now changed to the maze router with the following cost function $\beta : E_i \to R$

$$\beta(R_e) = \sum_{e \in R_e} (a \cdot c_e + b \cdot o_e) \tag{6}$$

where a, b, are user-defined parameters, and $o_e \in \{0, 1\}$. If an overflow happens, o_e is set to 1; otherwise, it is set to 0.

TABLE I
EXPERIMENTAL RESULTS BASED ON THE MCNC BENCHMARKS FOR TESTABILITY ENHANCEMENT OF INTERCONNECT DETECTION AND DIAGNOSIS

Circuit	Statistics				#rings constructed for testability $\|R_t\|$ & diagnosis $\|R_d\|$	
	#core	#pad	#hyp	#2-pin	$\|R_t\|$	$\|R_d\|$
ac3	27	75	211	416	133(33.3ms)	374(93.5ms)
ami33	33	42	117	343	242(60.5ms)	303(75.8ms)
ami49	49	22	361	475	156(39ms)	386(96.5ms)
apte	9	73	92	136	73(18.3ms)	122(30.5ms)
hp	11	45	72	195	81(20.3ms)	164(41ms)
xerox	10	2	161	356	218(54.5ms)	342(85.5ms)

TABLE II
ROUTING BENCHMARK CIRCUITS

Circuit	Size (μm)	#Layers	#Nets	#Pins
Mcc1	39000×45000	4	1694	3101
Mcc2	152400×152400	4	7541	25024
Struct	4903x4904	3	3551	5717
Primary1	7552x4988	3	2037	2941
Primary2	10438x6468	3	8197	11226
S5378	4330x2370	3	3124	4734
S9234	4020x2230	3	2774	4185
S13207	6590x3640	3	6995	10562
S15850	7040x3880	3	8321	12566
S38417	111430x6180	3	21035	32210
S38584	12940x6710	3	28177	42589

There is a tradeoff between minimizing congestion and overflow. At the uncoarsening stage, we intend to resolve the overflow in a tile. Therefore, we make b much larger than a. In addition, a detailed maze routing is performed after the global maze routing. Iterative refinement of a failed net is stopped when a route is found. Uncoarsening continues until the first level G_0 is reached and the final solution is found.

IV. EXPERIMENTAL RESULTS

The multilevel routing system was implemented in the C programming language on a 900-MHz SUN Blade 2500 workstation with 1-GB memory. We conducted three sets of experiments: 1) testability enhancement; 2) congestion control for routing considering multiple faults, manufacturability, and crosstalk; and 3) process variation and crosstalk improvement due to congestion. Three types of benchmarks were used in our experiments: the first type is for intermodule interconnects only (see Table I); the second is the full-chip benchmarks (only mcc1 and mcc2), which include both intermodule interconnections and intramodule interconnections; and the third type contains only intramodule interconnections which are local interconnections within standard-cell modules. The statistics of type-2 and -3 benchmarks are given in Table II.

A. Testability Enhancement

For testability enhancement, the experimental results of the embedded ORT scheme in the proposed multilevel routing framework are reported in Table I. We have presented both a detection (the preprocessing stage) and a diagnosis scheme (the postprocessing stage), as shown in Fig. 7, for OR-based interconnect testing in SOC in a predetermined design flow. Thus, $f_{\min} \leq f_i \leq f_{\max}$ gives the timing specification for this scheme, where f_i is the estimated oscillation frequency for the ith ring. Since our target of this ORT scheme is for interconnects among modules, our experiments are conducted based on the MCNC benchmark circuits with intermodule connections.

Table I gives the names of the circuits, the statistics for the circuits (the number of cores, #core; the number of pads, #pad; the number of hyperedges, #hyp; the number of two-pin nets, #2-pin), the number of rings constructed for detection, $|R_t|$, and the number of rings constructed for diagnosis, $|R_d|$. Thus, $|R_t|$ is the testability-driven cost in the preprocessing stage, and $|R_d| - |R_t|$ is the additional cost for the postprocessing stage. In addition to the 100% fault coverage of the OR detection scheme, we also obtain 100% net segment diagnosability.

To show the feasibility of this scheme, we include the actual estimated ATE measurement time in the parentheses in Table I. Since the frequency of each ring is predetermined during the design phase, a delay fault can thus be detected and measured by inspecting the contents of the local core counters (see Fig. 2). Let the oscillation frequency of the rings, according to the timing specification, be $f_{\min} \leq f_i \leq f_{\max}$, with the unit time of measuring $T_0 (= n/f)$. Thus, we have the delay counter contents of $n_{\min} \leq n_i \leq n_{\max}$, where $n_{\min} = f_{\min} \times T_0$ and $n_{\max} = f_{\max} \times T_0$. Let ξ be the resolution of delay measurement, and ε be the maximum measurement error. Since a counter's maximum measurement error is ± 1, the requirement for ε, which is defined as follows, should be the reciprocal of $f_{\min} \times T_0$

$$\varepsilon = \frac{1}{f_{\min} \times T_0} \leq \zeta. \tag{7}$$

We show an example of the delay measurement. Let the frequency specification of the ORs be 4 to 400 MHz, and ξ is 0.001, which implies that the counter content $n_{\min}$ is at least 1000. From (7), we have the required T_0 250 μs. Thus, we get the estimated detection and diagnosis time in the parentheses. For example, for the ac3 circuit, we need 133 rings to detection and 374 rings to diagnose; therefore, $133 \times 250\ \mu\text{s} = 33.25$ ms for interconnect detection, and $374 \times 250\ \mu\text{s} = 93.5$ ms for interconnect diagnosis. This shows the effectiveness and efficiency of the testability enhancement.

B. Congestion Control for Multiobjective Optimization

As mentioned in Section II, the CMP variation is controlled by reducing the variation of pattern density ρ (see Section II-E), whereas the signal integrity problem can be alleviated by reducing the routing congestion (see Section II-F). It will be clear later that our router effectively reduces both parameters (i.e., the congestion variation for CMP and the maximum congestion for the crosstalk effect).

The statistics of type-2 and -3 benchmarks are given in Table II , including the circuit size, the number of layers, the number of nets, and the number of pins. Table III reports the results for multilevel routing considering multiple faults, manufacturability, and crosstalk. We compare three different

TABLE III
COMPARISON OF ROUTING RESULTS OF MAXIMUM DENSITY WITH BOTH MAXIMUM DELAY AND AVERAGE DELAY

Circuit	(A) Performance-Driven [24]				(B) Routability-Driven [24]				(C) Proposed Balanced Density with 100% routability			
	d_{max}	d_{avg}	$\#Net_{PEAK}$	CPU	d_{max}	d_{avg}	$\#Net_{PEAK}$	CPU	d_{max}	d_{avg}	$\#Net_{PEAK}$	CPU
Mcc1	4.65e+7	1.08e+7	181	223.68	2.03e+8	3.32e+7	61	77.11	2.03e+8	3.33e+7	40	72.63
Mcc2	7.26e+7	5.07e+6	274	5964.2	8.46e+7	5.12e+6	135	2855.5	8.51e+7	5.11e+6	96	2592.34
Comparison	0.41	0.41	3.35	2.32	1	1	1.44	1.10	1	1	1	1
Struct	1.13e+6	6.93e+4	32	307.91	1.52e+6	7.13e+4	9	56.33	1.52e+6	7.13e+4	7	56.53
Primary1	3.01e+5	3.33e+4	51	241.96	7.00e+5	5.51e+4	17	63.9	6.99e+5	5.50e+4	15	64.36
Primary2	3.91e+6	2.08e+5	91	1808.56	3.92e+6	2.09e+5	28	298.17	3.91e+6	2.09e+5	25	295.32
S5378	8.89e+4	6.38e+3	49	23.28	8.91e+4	6.39e+3	17	4.13	8.94e+4	6.41e+3	15	4.29
S9234	1.02e+5	9.4e+3	61	16.78	2.53e+5	1.19e+4	15	2.91	2.53e+5	1.19e+4	14	2.9
S13207	3.96e+5	2.04e+4	114	65.45	4.64e+5	2.04e+4	30	14.44	4.64e+5	2.03e+4	27	14.57
S15850	6.03e+5	2.89e+4	140	181.82	2.66e+6	6.68e+4	30	22.04	2.66e+6	6.67e+4	26	21.77
S38417	5.22e+5	2.93e+4	272	741.53	8.52e+6	3.94e+5	27	50.02	8.52e+6	3.94e+5	23	50.08
S38584	1.64e+6	5.83e+4	295	1453.8	1.76e+8	1.25e+7	31	127.8	1.76e+8	1.25e+7	29	122.5
Comparison	0.45	0.35	6.11	7.66	1	1	1.24	1.08	1	1	1	1

TABLE IV
COMPARISON OF ROUTING RESULTS OF STATISTICAL DENSITY WITH [24]

Circuit	(A) Performance-Driven [24]				(B) Routability-Driven [24]				(C) Proposed Balanced Density with 100% routability			
	$\#Net_{avg_v}$	$\#Net_{avg_h}$	σ_v	σ_h	$\#Net_{avg_v}$	$\#Net_{avg_h}$	σ_v	σ_h	$\#Net_{avg_v}$	$\#Net_{avg_h}$	σ_v	σ_h
Mcc1	28.19	31.78	20.59	24.35	10.03	11.50	10.45	10.82	9.91	11.33	7.58	7.33
Mcc2	39.35	44.05	37.26	46.98	19.39	21.65	23.53	25.80	18.74	20.88	17.30	18.54
Comparison	2.36	2.35	2.33	2.76	1.03	1.03	1.37	1.42	1	1	1	1
Struct	4.97	4.86	4.62	5.03	1.42	1.41	1.24	1.67	1.42	1.41	1.07	1.59
Primary1	2.29	1.74	3.00	5.67	0.70	0.60	1.05	1.95	0.70	0.60	1.20	1.80
Primary2	7.22	7.49	5.56	18.23	2.05	1.85	1.59	4.57	2.05	1.85	1.56	4.45
S5378	12.53	13.46	9.16	8.40	4.38	3.44	3.45	2.13	4.40	3.46	3.44	2.10
S9234	14.16	9.99	12.91	7.04	3.95	2.56	3.25	1.62	3.95	2.56	3.24	1.60
S13207	28.43	20.49	18.40	11.08	9.30	5.93	5.77	2.76	9.29	5.92	5.23	2.81
S15850	36.61	34.48	23.89	20.42	10.29	7.41	5.63	2.92	10.31	7.41	5.39	2.91
S38417	44.58	27.38	37.36	27.94	7.31	4.27	4.75	2.17	7.3	4.27	4.44	2.18
S38584	43.99	30.53	35.93	20.12	9.06	5.80	5.74	2.86	9.05	5.79	5.43	2.88
Comparison	4.02	4.52	4.87	5.55	1.00	1.00	1.17	1.23	1	1	1	1

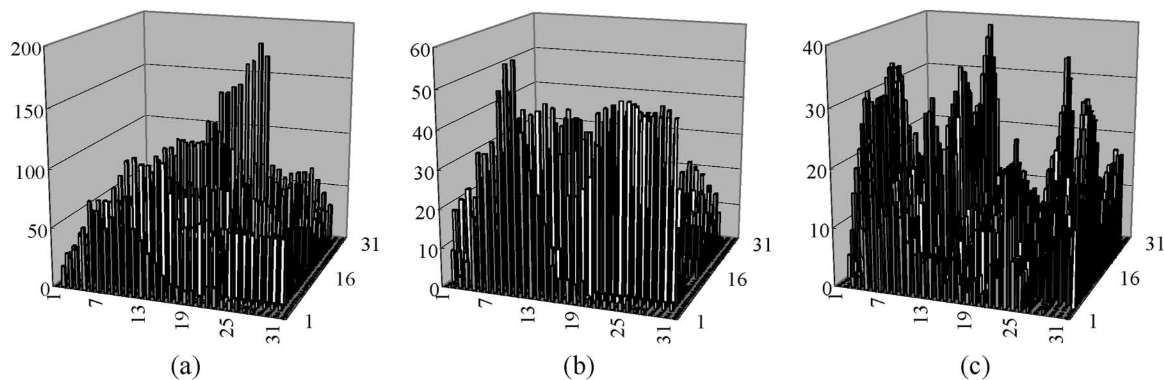


Fig. 10. Routing density distribution for mcc1 for (a) performance-driven MR, (b) routability-driven MR, and (c) proposed algorithm.

routing algorithms: 1) performance-driven MR [24]; 2) routability-driven MR [24]; and 3) our proposed method (with MST routing and balanced density).

In each case, we give the maximum (critical path) delay $d_{\max}$, average delay d_{avg}, and the maximum number of nets crossing a level-0 tile $\#\text{Net}_{\text{PEAK}}$, which is a good estimate for the maximum routing density. In our experiment, we set the parameter $t = 4$ for the ISCAS89 circuits, whereas other benchmarks are set to $t = 2$. The completion rate is 100% for all cases. The proposed method achieves about the same level of performance as the routability-driven method does by up to 0.2% increase in $d_{\max}$ and d_{avg}, but the maximum density is much smaller. Compared with [24], the experimental results show that our router improves the maximal congestion ($\#\text{Net}_{\text{PEAK}}$) by 1.24–6.11× in runtime speedup by 1.08–7.66×.

In Table IV, we show some statistical density results. The average number of nets crossing a level-0 tile is denoted by $\#\text{Net}_{\text{avg}}$, and we also list those of vertical tiles and horizontal tiles $\#\text{Net}_{\text{avg}_v}$ and $\#\text{Net}_{\text{avg}_h}$, respectively. In addition, σ_v is denoted for the standard deviation from the vertical tile prospect and σ_h for that of the horizontal tile prospect. The results show

TABLE V
COMPARISON OF ROUTING RESULTS OF STATISTICAL DENSITY WITH [26]

Circuit	(A) Crosstalk- and performance-driven [26]						(B) Proposed Balanced Density with 100% routability					
	$\#Net_{PEAK_v}$	$\#Net_{PEAK_h}$	$\#Net_{avg_v}$	$\#Net_{avg_h}$	σ_v	σ_h	$\#Net_{PEAK_v}$	$\#Net_{PEAK_h}$	$\#Net_{avg_v}$	$\#Net_{avg_h}$	σ_v	σ_h
Mcc1	75	66	17.96	10.98	6.44	9.46	40	36	9.91	11.33	7.58	7.33
Mcc2	146	172	20.53	28.62	21.64	32.54	80	96	18.74	20.88	17.30	18.54
Compar.	1.84	1.80	1.34	1.23	1.13	1.63	1	1	1	1	1	1
Struct	31	28	4.59	4.85	4.49	5.07	7	7	1.42	1.41	1.07	1.59
Primary1	27	29	2.03	0.94	3.58	3.30	7	15	0.70	0.60	1.20	1.80
Primary2	32	37	2.90	2.65	2.95	5.20	10	25	2.05	1.85	1.56	4.45
S5378	17	13	5.11	3.59	3.62	2.43	15	10	4.40	3.46	3.44	2.10
S9234	16	9	4.79	2.56	3.36	1.77	14	7	3.95	2.56	3.24	1.60
S13207	33	17	9.39	6.35	6.14	3.42	27	15	9.29	5.92	5.23	2.81
S15850	32	15	9.94	7.43	6.33	2.96	26	16	10.31	7.41	5.39	2.91
S38417	29	13	7.41	4.19	4.75	2.23	23	12	7.3	4.27	4.44	2.18
S38584	56	28	10.83	6.45	8.36	3.51	29	16	9.05	5.79	5.43	2.88
Compar.	1.73	1.54	1.18	1.17	1.41	1.34	1	1	1	1	1	1

TABLE VI
NORMALIZED MAXIMUM ILD VARIATION

Circuit	Performance-V [24]	Performance-H [24]	Routability-V [24]	Routability-H [24]	Proposed-V	Proposed-H
Mcc1	4.02	5.14	1.36	1.66	1	1
Mcc2	2.33	2.85	1.69	1.31	1	1
Compar.	3.18	4.00	1.53	1.49	1	1
Struct	4.57	3.71	1.00	1.29	1	1
Primary1	3.57	3.40	1.43	1.13	1	1
Primary2	4.90	3.64	1.00	1.12	1	1
S5378	3.27	3.40	1.13	1.00	1	1
S9234	4.36	5.43	1.07	1.00	1	1
S13207	4.22	3.93	1.11	1.00	1	1
S15850	6.00	6.94	1.15	0.94	1	1
S38417	10.74	22.33	1.17	0.92	1	1
S38584	40.90	124.31	1.07	0.88	1	1
Compar.	9.17	19.68	1.13	1.03	1	1

*V stands for vertical metal layers, while H stands for horizontal metal layers.

that our scheme is more effective for the full-chip benchmarks mcc1 and mcc2. For other intramodule routing, our scheme also improves the results for most cases. Compared with [24], the experimental results show that our router improves the average congestion by about 1.00–4.52×, and improves the balanced congestion (σ_v and σ_h, standard deviation respective for vertical and horizontal tiles) by 1.37–5.55×.

To demonstrate the effectiveness of the proposed algorithm in balancing the routing density, the number of horizontal wires crossing each level-0 tile for benchmark mcc1 is shown in Fig. 10 for the three algorithms. The performance-driven MR results are the least balanced routing; and the peak congestion is 181 ($\#\mathrm{Net}_{\mathrm{PEAK}}$) in mcc1. The routability-driven MR tries to avoid congested area to improve the probability of successful routing, and thus reduces the maximum density; and its peak congestion is still 61. With our proposed algorithm, the maximum density is further reduced to 45; and thus the manufacturability effects, the probability of multiple faults, and crosstalk effects are reduced accordingly.

Mcc1 shows the maximal congestion improvement in our proposed algorithm by 1.36× compared to the routability-driven MR and by 4.02× compared to the performance-driven MR. For mcc1, our proposed algorithm improves the average congestion by 1.01–1.02× compared to the routability-driven MR and by 2.81–2.85× compared to the performance-driven MR. For balanced congestion on mcc1, our proposed algorithm improves the result by 1.38–1.48× compared to the routability-driven MR and by 2.72–3.32× compared to the performance-driven MR. For runtime speedup, our approach improves by 1.06× compared to routability-driven MR and by 3.08× compared to performance-driven MR.

Further, the interconnection congestion, as evident in the intermodule connections in mcc1 and mcc2, demonstrates the respective maximal and average congestion improvements by 1.39–3.23× and 1.03–2.36× with the congestion balance improvement (σ_v and σ_h, standard deviation respective for vertical and horizontal tiles) by 1.37–2.76×.

Table V compares the results of our work with the crosstalk- and performance-driven router presented in [26]. Our router improves the respective maximal and average congestion improvements by 1.54–1.84× and 1.17–1.34× with the congestion balance improvement (σ_v and σ_h, respective standard deviations for vertical and horizontal tiles) by 1.13–1.63×. Please note that we do not compare the maximum (critical path) delay $d_{\max}$, average delay d_{avg} with [26] since our congestion-guided router achieves 100% routing completion, whereas the work [26] does not.

C. Process Variation and Coupling Capacitance

To estimate the reduction in ILD thickness variation in CMP effects, we use (2) to evaluate ILD thickness by

TABLE VII
COMPARISON IN MAXIMUM UNIT-LENGTH COUPLING CAPACITANCE

Circuit	Crosstalk Optimization [26]	Performance-Driven [24]	Routability-Driven [24]	Our Proposed
Mcc1	1.88	4.53	1.53	1
Mcc2	1.79	2.85	1.31	1
Compar.	1.84	3.69	1.42	1
Struct	4.29	4.53	1.29	1
Primary1	1.93	2.85	1.13	1
Primary2	1.48	4.57	1.12	1
S5378	1.13	3.40	1.13	1
S9234	1.14	3.64	1.07	1
S13207	1.22	3.27	1.11	1
S15850	1.23	4.36	1.15	1
S38417	1.26	11.83	1.17	1
S38584	1.93	10.17	1.03	1
Compar.	1.73	5.40	1.13	1

assuming enough polishing time (i.e., $t > (\rho_0 z_1)/K_i$ for the maximum ρ_0). The results are given in Table VI, assuming that vertical and horizontal wires are routed in different metal layers. The maximum ILD variation is calculated in each case, and the result is compared with our method. Since a major cause of process variation is the CMP variation [7], [8], it is essential to achieve CMP process uniformity. A simple and effective way to control the ILD variation due to the CMP process is through balanced density routing. Compared with the performance-driven routing [24], the ILD variations in vertical and horizontal metal layers are improved by 9.17× and 19.68×, respectively. The respective average improvements over the routability-driven router [24] in vertical and horizontal metal layers are 1.13× and 1.03×.

The crosstalk noise is induced by signal transition in the aggressor net and the coupling capacitance between adjacent wires. Thus, the best way to avoid crosstalk noise is to reduce coupling capacitance by increasing wire space, as suggested by (3). In Table VII, we show the worst-case unit-length capacitance in each case, as compared to our method. The results in Table VII show that the proposed balanced-density router achieves the best worst-case unit-length coupling capacitance among all routers. Compared with the routability-driven router [24], our congestion-control router improves the unit-length coupling capacitance by 1.03× to 1.53× with 100% routing completion. Compared with the performance-driven router [24], our congestion-control router improves the unit-length coupling capacitance by 2.85× to 11.83× with also 100% routability. Compared with the multilevel router [26], our congestion-control router improves the unit-length coupling capacitance by 1.13× to 4.29× with also 100% routability. All our experimental results show that reducing congestion is an effective method in reducing coupling capacitance.

V. CONCLUSION

We have shown that the embedded ORT and diagnosis scheme are feasible based on the simulation results with TSMC 0.18-μm process technology. In addition, this ORT scheme achieves 100% fault detection coverage and maximal diagnosability. We have also presented an effective multilevel routing framework that applies a congestion-driven routing algorithm to reduce the multiple-fault probability, CMP- and OPC-induced effects, and crosstalk effects for yield enhancement.

REFERENCES

[1] Semiconductor Industry Association (SIA), International Technology Roadmap for Semiconductors (ITRS), 2004.
[2] *IEEE Standard Testability Method for Embedded Core-based Integrated Circuits*, Test Technology Technical Council of the IEEE Computer Society, IEEE Std. 1500, 2005.
[3] K. S.-M. Li, C. Su, Y.-W. Chang, C.-L. Lee, and J.-E. Chen, "IEEE Standard 1500 compatible interconnect diagnosis for delay and crosstalk faults," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 25, no. 11, pp. 2513–2525, Nov. 2006.
[4] A. B. Kahng, B. Liu, and I. I. Mandoiu, "Non-tree routing for reliability and yield improvement," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Des.*, Nov. 2002, pp. 260–266.
[5] W. Maly, "Moore's law and physical design of ICs," in *Proc. ACM Int. Symp. Phys. Des.*, 1998, p. 36. (special address).
[6] V. K. R. Chiluvuri, "Yield optimization in physical design: A review," in *Proc. 5th ACM/SIGDA Phys. Des. Workshop*, 1996, pp. 198–206.
[7] G. Nanz and L. E. Camilletti, "Modeling of chemical–mechanical polishing: A review," *IEEE Trans. Semicond. Manuf.*, vol. 8, no. 4, pp. 382–389, Nov. 1995.
[8] Y. Chen, A. B. Kahng, G. Robins, and A. Zelikovsky, "Practical iterated fill synthesis for CMP uniformity," in *Proc. ACM/IEEE Des. Autom. Conf.*, 2000, pp. 671–674.
[9] L.-D. Huang and M. D. F. Wong, "Optical proximity correction (OPC)-friendly maze routing," in *Proc. ACM/IEEE Des. Autom. Conf.*, Jun. 2003, pp. 812–817.
[10] K. S.-M. Li, C.-L. Lee, C.-C. Su, and J.-E. Chen, "A unified approach to detecting crosstalk faults of interconnects in deep sub-micron VLSI," in *Proc. Asia Testing Symp.*, Nov. 2004, pp. 145–150.
[11] G. E. Jan, K.-Y. Chang, S. Gao, and I. Parberry, "A 4-geometry maze router and its application on multiterminal nets," *ACM Trans. Design Autom. Electron. Syst.*, vol. 10, no. 1, pp. 116–135, Jan. 2005.
[12] D. Hightower, "A solution to line routing problems on the continuous plane," in *Proc. ACM/IEEE Des. Autom. Conf.*, 1969, pp. 1–24.
[13] C. J. Albrecht, "Global routing by new approximation algorithms for multicommodity flow," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 20, no. 5, pp. 622–632, May 2001.
[14] Y.-W. Chang, K. Zhu, and D. F. Wong, "Timing-driven routing for symmetrical-arraybased FPGAs," *ACM Trans. Design Autom. Electron. Syst.*, vol. 5, no. 3, pp. 433–450, Jul. 2000.
[15] M. Marek-Sadowska, "Router planner for custom chip design," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Des.*, Nov. 1986, pp. 246–249.
[16] J. Cong, J. Fang, and Y. Zhang, "Multilevel approach to full-chip gridless routing," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Des.*, Nov. 2001, pp. 396–403.
[17] C. J. Alpert, J.-H. Huang, and A. B. Kahng, "Multilevel circuit partitioning," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 17, no. 8, pp. 655–667, Aug. 1998.
[18] G. Karypis, R. Aggarwal, V. Kumar, and S. Shekhar, "Multilevel hypergraph partitioning: Application in VLSI domain," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 7, no. 1, pp. 69–79, Mar. 1999.
[19] T. F. Chan, J. Cong, T. Kong, and J. R. Shinnerl, "Multilevel optimization for large-scale circuit placement," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Des.*, Nov. 2000, pp. 171–176.
[20] H.-C. Lee, Y.-W. Chang, J.-M. Hsu, and H. Yang, "Multilevel large-scale module floorplanning/placement using B^*- trees," in *Proc. ACM/IEEE Des. Autom. Conf.*, Jun. 2003, pp. 812–817.
[21] M. Hayashi and S. Tsukiyama, "A hybrid hierarchical global router for multi-layer VLSIs," *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, vol. E78-A, no. 3, pp. 337–344, Mar. 1995.
[22] Y.-L. Lin, Y.-C. Hsu, and F.-S. Tsai, "Hybrid routing," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*., vol. 9, no. 2, pp. 151–157, Feb. 1990.
[23] J. Cong, M. Xie, and Y. Zhang, "An enhanced multilevel routing system," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Des.*, Nov. 2002, pp. 51–58.
[24] S.-P. Lin and Y.-W. Chang, "A novel framework for multilevel routing considering routability and performance," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Des.*, Nov. 2002, pp. 44–50.

[25] R. Kastner, E. Bozorgzadeh, and M. Sarrafzadeh, "Predictable routing," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Des.*, Nov. 2000, pp. 110–114.
[26] T.-Y. Ho, Y.-W. Chang, S.-J. Chen, and D.-T. Lee, "A fast crosstalk- and performance-driven multilevel routing system," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Des.*, Nov. 2003, pp. 382–387.
[27] B. Stine, "A closed-form analytical model for ILD thickness variation in CMP processes," in *Proc. CMP-MIC*, 1997, pp. 266–273.
[28] C. J. Alpert, A. Devgan, and S. T. Quay, "Buffer insertion for noise and delay optimization," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 18, no. 11, pp. 1633–1645, Nov. 1999.

Katherine Shu-Min Li (S'04–M'06) received the B.S. degree from Rutgers University, New Brunswick, NJ, and the M.S. and Ph.D. degrees from the National Chiao Tung University, Hsinchu, Taiwan, R.O.C., in 2001 and 2006, respectively.

She is currently an Assistant Professor in the Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan. Her research interests focus on crosstalk effects, signal integrity, SOC testing, floorplanning and routing for testability and yield enhancement, design for manufacturing, design for yield, transition faults, scan reordering, scan routing, low-power scan techniques, particularly on oscillation ring test scheme, and interconnect optimization in deep submicron and nanotechnology.

Dr. Li is a member of the IEEE Circuits and Systems Society (CAS), the Association for Computing Machinery (ACM), and ACM/Special Interest Group on Design Automation (SIGDA).

Yao-Wen Chang (S'94–M'96) received the B.S. degree from National Taiwan University, Taipei, Taiwan, in 1988, and the M.S. and Ph.D. degrees from the University of Texas at Austin in 1993 and 1996, respectively, all in computer science.

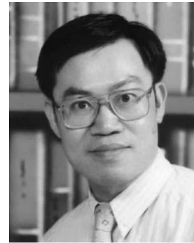
He is a Professor in the Department of Electrical Engineering and the Graduate Institute of Electronics Engineering, National Taiwan University. He is currently also a Visiting Professor at Waseda University, Japan. He was with the IBM T. J. Watson Research Center, Yorktown Heights, NY, in the summer of 1994. From 1996 to 2001, he was on the faculty of National Chiao Tung University, Taiwan. His current research interests lie in VLSI physical design, design for manufacturing, and FPGA. He has been working closely with industry on projects in these areas. He has coauthored one book on routing and over 120 ACM/IEEE conference/journal papers in these areas.

Dr. Chang received an award at the 2006 ACM ISPD Placement Contest, Best Paper Award at ICCD-1995, and nine Best Paper Award Nominations from DAC-2007, ISPD-2007 (two), DAC-2005, 2004 ACM TODAES, ASP-DAC-2003, ICCAD-2002, ICCD-2001, and DAC-2000. He has received many awards for research performance, such as the inaugural First-Class Principal Investigator Awards and the 2004 Mr. Wu Ta You Memorial Award from the National Science Council of Taiwan, the 2004 MXIC Young Chair Professorship from the MXIC Corp, and for excellent teaching from National Taiwan University and National Chiao Tung University. He is an editor of the *Journal of Computer and Information Science*. He has served on the ACM/SIGDA Physical Design Technical Committee and the technical program committees of ASP-DAC (topic chair), DAC, DATE, FPT (program co-chair), GLSVLSI, ICCAD, ICCD, IECON (topic chair), ISPD, SOCC (topic chair), TENCON, and VLSI-DAT (topic chair). He is currently an independent board member of Genesys Logic, Inc, the chair of the Design Automation and Testing (DAT) Consortium of the Ministry of Education, Taiwan, a member of the board of governors of the Taiwan IC Design Society, and a member of the IEEE Circuits and Systems Society, ACM, and ACM/SIGDA.

Chung-Len Lee (S'71–M'81) received the B.S. degree in electrical engineering from the National Taiwan University, Taipei, Taiwan, R.O.C., in 1968, and the M.S. and Ph.D. degrees in electrical engineering from Carnegie Mellon University, Pittsburgh, PA, in 1971 and 1975, respectively.

In 1975, he joined the National Chiao Tung University, Hsinchu, Taiwan, as a Faculty Member, where he was engaged in research on various topics on optoelectronics, semiconductor devices, integrated circuits, and processes. He has published more than 300 papers in these areas.

Dr. Lee is a member of Phi Kappa Phi and Phi Tau Phi.



Chauchin Su (S'80–M'90) received the B.S. and M.S. degrees from the National Chiao Tung University, Hsinchu, Taiwan, R.O.C., in 1979 and 1981, respectively, and the Ph.D. degree from the University of Wisconsin, Madison, in 1990, all in electrical engineering.

In 1990, he joined the Department of Electrical Engineering, National Central University, Jhungli, Taiwan. In 2000, he transferred to the Department of Electrical and Control Engineering, National Chiao Tung University, where he is currently a Professor. His current research interests focus on the design and test of high-speed serial links and mixed signal circuits.

Prof. Su has served as a Technical Program Committee Member of the International Conference on Computer-Aided Design, Asia and South Pacific Design Automation Conference, Asian Test Symposium (ATS), and International Mixed-Signal Test Workshop. He has also served as the Technical Program Cochair of ATS 2000 and the General Cochair of ATS 2004. Presently, he is the Executive Director of the National System-on-Chip Program, Taiwan.

Jwu E. Chen (S'88–M'92) received the B.S., M.S., and Ph.D. degrees in electronics engineering from National Chiao-Tung University, Hsinchu, Taiwan, R.O.C., in 1984, 1986, and 1990, respectively.

He was a Faculty Member in the Department of Electrical Engineering, Chung Hua University, Hsinchu, from 1990 to 2004. He is currently a Faculty Member in the Department of Electrical Engineering, National Central University, Jhungli, Taiwan. His research interests include multiple-valued logic, VLSI testing, reliable computing, yield analysis, and humanoid robotics.

Dr. Chen is a member of the IEEE Computer Society.