Routing Architecture Exploration for Regular Fabrics

V. Kheterpal, A. J. Strojwas, L. Pileggi
{vkheterp, ajs, pileggi}@ece.cmu.edu
Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA - 15213

ABSTRACT
In an effort to control the parameter variations and systematic yield problems that threaten the affordability of application-specific ICs, new forms of design regularity and structure have been proposed. For example, there has been speculation [6] that regular logic fabrics [1] based on regular geometry patterns [2] can offer tighter control of variations and greater control of systematic manufacturing failures. In this paper we describe a routing framework that accommodates arbitrary descriptions of regular and structured routing architectures. We further propose new regular routing architectures and explore the various performance vs. manufacturability trade-offs. Results demonstrate that a more regular, restricted routing architecture can provide a substantial advantage in terms of manufacturability and predictability while incurring a moderate performance penalty.

Categories and Subject Descriptors
B.7.2 [**Placement and routing**]: Design and exploration aids

General Terms
Algorithms, Design, Experimentation

Keywords
BEOL, Regularity

1. INTRODUCTION
As integrated circuit (IC) technologies continue forward into the nano-era, the IC design landscape requires new design styles and circuit fabrics which can leverage the advantages of new technology nodes and provide ASIC performance at an affordable cost. Of particular interest is increased design regularity and structure. One recently proposed regular fabric is a Via Patterned Gate Array (VPGA) [1], which is similar to an FPGA since it is comprised of an array of programmable logic blocks (PLB), but is distinctly different in two ways.

This work was funded in part by C2S2, the MARCO Focus Center for Circuit & System Solution, under MARCO contract 2003-CT-888.

DAC 2004, June 7–11, 2004, San Diego, California, USA

First, the logic configuration is performed by an application-specific via layer, rather than field programmed with SRAM switches. Secondly, using metal mask configuration of "switches," the routing is done on top of the PLBs, rather than beside the PLBs using active transistors. This significantly reduces the die-area overhead in contrast to dedicated routing channels in FPGAs. VPGAs can afford a much higher switch density per switch box, which translates to a much higher flexibility in routing. There are an increasing number of commercial regular fabrics, even hybrids with combinations of SRAM programmable PLBs and via-configurable routing[4] which offer design cost and manufacturability advantages. Various forms of regularity can be used to enhance the printability, hence systematic yield of an IC. Yield losses due to defectivity of metallization layers have been dominant for many generations of IC's. In older technologies with aluminum (Al) BEOL (back end of line) processing, most of the yield losses for mature products were caused by random defects creating shorts between metal wires. Systematic yield loss mechanisms or parametric yield losses were in most cases negligible in high volume production since the new technology development cycle was sufficiently long enough to remove the systematic effects and improve product design robustness. With ever increasing IC complexity and shrinking feature sizes, however, systematic yield loss mechanisms started to dominate for technology generations of 0.18um and below[3]. In particular, to keep the IC yield at economically acceptable levels for the ever increasing number of contact and via holes, their fail rates had to be kept at the levels of few per billion. These problems have been compounded by the introduction of Cu (copper) interconnect in the 0.13 um technology node. In the dual damascene process required to pattern interconnect layers, chemical-mechanical polishing (CMP) is used to remove Cu outside of the patterned trenches and provide a planarized wafer surface. Unfortunately, due to the difference in hardness between Cu and dielectric materials, significant layer height (or topography) variations can occur. As a result, Cu residues may create shorts between metal lines.

It should be also noted that different physical effects have sometimes very different "interaction distances." For example, CMP interaction distance can be on the order of hundreds of microns, while the interaction distance for lithography is only a couple microns for 0.13 um technology.

Enforcement of geometrical regularity in all metallization layers provides an elegant and robust solution to all of the problems listed above Moreover, since line-width uniformity and Cu layer thicknesses will be much more uniform, this will also reduce the variability in the parasitic component values and thus improve timing predictability [5].

In this paper we present an experimental routing framework that explores the trade-offs of regular and structured routing architectures. In particular, we propose two regular routing models and compare them with traditional ASIC routing. We will refer to the first regular routing style as structured routing, whereby all of the metal masks are application specific, but the routing is highly restricted, and fully populates the layers. For our second model we proposed a structured routing that is based on via configurability, whereby only the via masks are application specific, and the metal masks are common for all applications. We further explore the impact of redundant vias, as well as the cost vs. advantage of other manufacturability enhancements.

2. ROUTING MODELS

ASICs employ a fully customized routing model that requires complex design rules and application-specific masks for each routing layer. This routing flexibility is at the cost of complete systematic yield and metal routing density control. Moreover, while total wirelength is minimized as compared to a more restricted routing style, this can often result in a high number of interlayer vias due to the meandering nature of the shortest wiring paths. This problem is becoming worse at nanoscale, where total wiring can often dominate total area, and predicting congestion *a priori* is extremely difficult. Moreover, as via failure rates and parasitic resistances becoming increasingly dominant for shrinking feature sizes, this lack of wiring predictability becomes increasingly problematic for timing closure and design for manufacturability.

2.1 Fully Structured Routing

Figure 1 shows an example of a fully structured routing model as it appears on a gridded template. The router uses resources from the underlying template and is free to "cut" metal lines to desired lengths as required. Although this produces fully customized non-repeating shapes in the routing layer, the router is based on rules which do not allow it to produce undesirable structures. The pieces of metal which are not used for routing remain in the metal mask, thus producing a uniform metal density. Importantly, the gridding restrictions can be used to help ensure that vias are carefully spaced for printability. Referring to the grid template in Figure 1, each grid-point is a "potential via" site. The router is restricted to work on a strict grid, whereby wrong-way wiring and arbitrary spacing and sizing are prohibited. Parameters such as metal overhang at via sites and relative positioning of vias are pre-configured. "Potential vias" at consecutive rows or columns are prohibited in our model to allow for more compact grid-spacing. Clearly, the packing density of the routes, as well as the total interconnect resistance and interconnect capacitance will be worse than those for the ASIC routing model. As a result there is some amount of performance penalty incurred when using this model. This loss in performance, however, is traded-off against the improvement in printability and manufacturability.

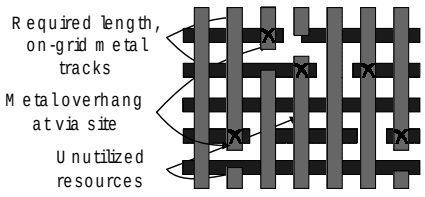

Figure 1: Structured routing on a grid template

2.2 VIA Configurable Routing

By further restricting the structured routing model and not allowing the router to "cut" metal lines at arbitrary locations, we obtain the via configurable routing model. The only freedom that the router has is selection of vias to complete a route. This routing structure clearly incurs additional performance and area penalties, but provides two important advantages: 1) application-specific customization for low volume products requires only a set of customized via masks, or potentially provides for direct write configurability [7]; 2) The number of via unique via patterns that must be printed can be controlled and optimized for printability, significantly better than ASIC or even structured routing.

With fixed metal masks, however, there is a decrease in routing flexibility, along with a penalty of dangling metal capacitance. This extra capacitance results in an additional performance degradation that we will assess in the results section.

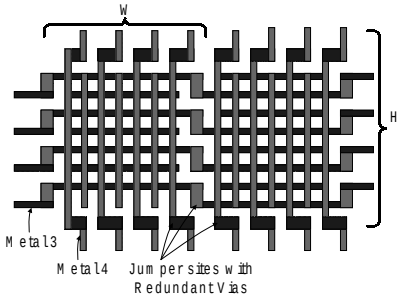

Figure 2: Repeating unit for the Proposed via configurable routing fabric

Figure 2 shows the proposed via configurable routing architecture. The basic repeating unit is an alternating set of short and long metal tracks with a set of jumpers at the end of each track. Each overlap of M3 and M4 tracks is a "potential via" site. M3 tracks also serve as access tracks for pins on each gate which are in M2. Each repeating unit can be viewed as a switch box with tracks entering and leaving. The jumpers provide for paths to enter the switchbox and continue in the same direction. Switching a track orthogonally within the switch box is costly in terms of the dangling capacitance incurred at the source track being switched. The "potential vias" at the jumper sites are more often used during routing than at the other sites. They also form potential failure locations due to long metal lines terminating in vias, thus each jumper site has a redundant via.

The granularity of this structure is determined by the parameters *W* and *H*. The process of choosing these parameters for an optimal performance-yield tradeoff is explained in Section 4. Architecturally, it is better to have this structure with higher granularity in the lower layers (M3 & M4) and have a similar structure with lower granularity in higher metal layers (M5 & M6). Routing that is relatively local can be done in M3 and M4 and longer routes can be done in M5 and M6.

3. ROUTING FRAMEWORK

The proposed structured routing and via configurable routing architecture models lie in-between ASIC and FPGA routing models on the axes of flexibility, manufacturability and performance. We represent both routing models in terms of a resource graph which is similar to the routing resource graphs used for FPGAs. The difference being that the resource graph lies

on top of logic blocks instead of within dedicated routing channels. Moreover, the switch density as well as the packing density of the routes is much higher for via patternable regular fabrics than for FPGAs, and more comparable to that for to ASICs. Therefore, routing methods used in ASICs or FPGAs cannot be directly applied. For this reason we have developed a routing framework which accommodates the proposed routing models and allows us to compare their performance with that for ASICs. The generalized routing architecture (resource graph) is described using an architecture description language.

3.1 Routing Architecture Description

Along with the placement of cells in the design, the router takes in an architecture description file. This file describes the metal structures that are available for routing. The architecture description language has primitives such as *Rectangle*, *Polygon* to define rectangular wires as well as polygon suitable for Manhattan style routing. The *Via* and *PVia* primitives are used to define a via and a "potential via" respectively. *Sections* are used to group various primitives and *instances* of other *Sections*. *Sections* can be parameterized with architectural parameters such as number of wires spanning various lengths, number of redundant vias, etc. This provides a convenient way of changing various parameters and observing the effect on the overall performance. Such an architecture description can be used to define a complex metal structure suitable for the via programmable model, or a simple grid for the structured routing model.

3.2 Routing flow

Figure 3 shows the flow within the proposed routing framework. We use a rip-up reroute scheme to resolve the order in which the nets should be routed. The routing architecture is described using the architecture description language described above. This architecture description is abstracted out into a resource routing graph which uniquely represents the architecture. Metal wires are abstracted as nodes, "potential vias" are abstracted as edges in this graph. Finding a path between any two nodes in this graph corresponds to completing a connection between the corresponding metal wires by selecting a set of "potential vias".

The routing is broken into a global routing and a detailed routing phase. We use the commercial physical synthesis tool "Dolphin" from Monterey Design Systems to perform the array placement and generate the global routing information. The global routes are used to create bounding boxes for each net. Figure 3 shows the bounding box for the global route which is overlaid on the resource graph. Each bounding box represents an area within which the corresponding net where it would preferably appear after detailed routing.

Along with the bounding box, a net-violation cost is defined for each net. This cost represents the penalty incurred by another net if routed using any of the resources used by this net. The initial values of the net-violation costs are same of each net. The router proceeds by finding a min-cost embedding for each net within the bounding box defined for it. For a two point net, this embedding reduces to finding a minimum cost path between the source and the destination and the path must lie within the bounding box. For a multipoint net, a minimum cost path between a node and any of the nodes in a set of nodes has to be found. The factors contributing to this cost function for embedding each net are the total resistance and the capacitance of the embedded net as well as the net-violation penalties of nets it violates. Following the embedding of each net, the number of violations for each net is calculated. Based on this number and the criticality of the net, the bounding box and violation cost is recomputed. Another iteration of embedding the nets is carried out using the newly computed bounding boxes and net-violation costs.

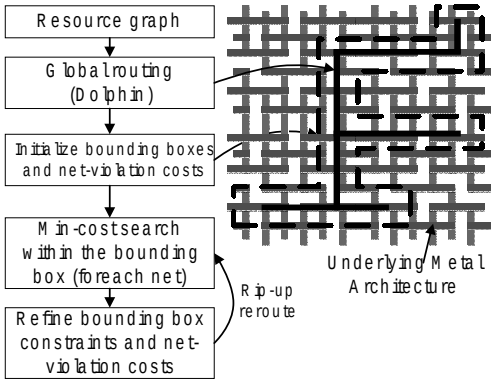

Figure 3: Routing flow

The router stops if all the embedded nets are free of any violations or a maximum number of iterations is reached. The above scheme, when used with the structured routing model, allows the router to create new nodes in the resource routing graph in order to model "slotting" of wires at desired places. This feature of the router is switched off for the via configurable model.

4. EXPERIMENTS/RESULTS

Using the routing flow in Figure 3, we compared the performance of the standard ASIC routing model with structured and via configurable routing. We further explored the advantages of redundant vias at the jumper sites of the via configurable routing fabrics, as shown in Figure 2. We begin with an RTL circuit description and produce the placement in the form of a regular array of logic blocks, as described in [1]. Each logic block consists of a LUT, two NAND gates, a Flip-flop and a couple of buffers.

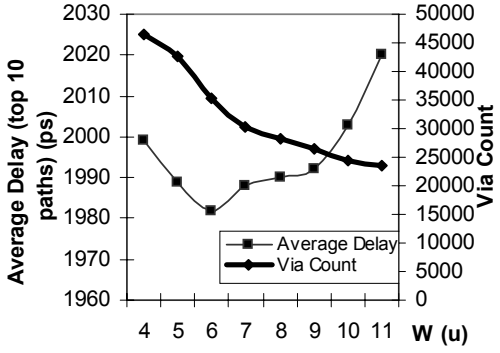

Figure 4: Performance vs. Granularity

Each placed design is routed using the structured model and the via configurable model using our routing framework. The configured metal architectures are incorporated back into

"Dolphin" to facilitate netlist checking and post-routing timing reports. The ASIC routing example is generated using "Dolphin's" standard flow. We used a six-Metal, 130 nanometer commercial process from ST Microelectronics. The structured routing was set up for a uniform grid over the entire routing space in each metal layer. For the via configurable model we used the architecture described in Section 2.2. The *W* and *H* values for the architecture were carefully chosen to provide a good engineering trade-off point. Figure 4 shows the performance plot of a firewire controller which was routed using different values of *W* (*H* was chosen to be equal to *W*). Large values of *W* (low granularity) cause a higher dangling capacitance when switching tracks orthogonally, thus degrading the performance. Lower values of *W* (higher granularity) reduce the dangling capacitance but increase the resistance of the paths due to large number of vias again degrading the performance. Thus an optimal value of *W* exists for minimal delay. As expected, via count monotonically decreases as the value of *W* increases. For this reason we selected a value of *W* which provided a good via count without incurring a substantial delay penalty.

Table 1: ASIC, Structured and Via configurable routing

Design	IC (pF)	IR (KΩ)	Ave Slack (ns)	N-R Via Count	Random Yield
Firewire					
ASIC	27.89	97.13	-1.45	15560	0.999782
Structured	29.43	102.01	-1.456	17631	0.999753
Via cfg- I	38.19	164.31	-1.55	26775	0.999625
Via cfg- II	38.61	144.33	-1.54	16410	0.999770
FPU					
ASIC	311.07	1003.5	-9.08	132K	0.998154
Structured	362.67	1112.9	-9.54	172K	0.997595
Via cfg - I	598.36	2185.4	-11.22	339K	0.995265
Via cfg - II	607.24	1835.6	-11.16	158K	0.997790
Network Switch					
ASIC	1828.9	4854.3	-3.39	564K	0.992135
Structured	2273.4	6389.1	-3.59	767K	0.989319
Via cfg - I	3924.1	144K	-4.11	1384	0.980811
Via cfg - II	4001.3	131K	-4.02	719K	0.989984

Table 1 compares the proposed routing models with the ASIC routing. Each of the benchmarks was routed with the structured model, the proposed via configurable fabric shown in figure 2 with(Via cfg-I) and without (Via cfg –II) redundant vias. IC and IR denote the total interconnect capacitance and the resistance of the routes respectively. We include the average slack of the top ten critical paths in the routed design. The slack is measured with respect to a cycle time of 0.5ns. The "N-R Via Count" is the total non-redundant vias in the design. The last column in the table shows the Random Yield loss due to Via failures. We calculate the yield loss due to via failures using an exponential failure model for which yield loss is modeled by Exp(-α N); where α is the failure rate in parts per billion and N is number of non-redundant vias in design. The performance of regular and structured routed designs is comparable to the ASIC routed designs. The via configurable fabric has higher interconnect capacitance and resistance. The higher capacitance is due to the dangling ends of the metal wires used for routing. The higher resistance is mainly due to the a high via count. On an average, there is a 13.2% degradation in performance for the via configurable routed designs. The FPU suffers from a large performance degradation (5.1% for the Structured and 23.6% for the Via Configurable) due to its long logic depth of the critical path. It should be noted, however, that, we have not considered the performance improvement expected for the more regular routing fabrics due to the ability to more tightly control the interconnect parasitic values, and therefore, require less design margin to cover the uncertainty.

Another important observation is that the simplicity of adding redundant vias to a via configurable routing structure offers significant advantage in terms of lowering the non-redundant via count and decreasing the resistance (Via cfg-II). On average, these vias constitute about 50% of the total vias in the Via cfg-II design examples. The random yield loss penalty is negligible compared to ASICs when the redundant via scheme is employed (Via cfg – II). Moreover, it is expected, that the overall yield will be significantly higher since the systematic and parametric yield components are maximized by using the regular metallization structures. In addition, there will be substantially less variability in the printed feature dimensions and topography as a function of process windows.

5. CONCLUSIONS

In this paper we presented two routing models which address the growing need of regularity in design motivated by the challenges involved with nanoscale IC manufacturing. We presented a routing framework to assess the performance and compare the proposed routing architectures with that for ASICs. Our results show that the performance penalty incurred with the proposed routing models is modest compared with the potential improvement in manufacturability and yield.

6. ACKNOWLEDGEMENTS

We would like to thank P. Gopalakrishnan and V. Rovner for their valuable comments during the course of this work.

7. REFERENCES

[1] L. Pileggi, H. Schmit, A. J. Strojwas et al, "Exploring regular fabrics to optimize the performance-cost trade-off" Proceedings of the ACM/IEEE DAC-2003.

[2] M. Palusinski, A. J. Strojwas and W. Maly, "Regularity in Physical Design", GSRC Workshop, Las Vages, June 2001

[3] A. J. Strojwas, "Process-Design Interaction Modeling Based Design For Manufacturability", Tutorial, DAC-2003.

[4] http://www.easic.com, http://www.lightspeed.com

[5] D.T. Blaauw et al, "Statistical Timing Analysis using Bounds and Selective Enumeration", IEEE Trans. on CAD of Integrated Circuits and Systems, Vol. 22, No.9, Sep 2003

[6] B. Tyrrell et al, "Investigation of the physical and practical limits of dense-only phase shift lithography for circuit feature definition", Society of Photo-Optical Instrumentation Engineers, October-2002.

[7] Gonzales, Anthony J. et al, "Recent results in the application of e-beam direct-write lithography", Proc. SPIE Vol. 1089, p. 37