Stability and Scalability in Global Routing

Sung Kyu Han[1], Kwangok Jeong[1], Andrew B. Kahng[1,2] and Jingwei Lu[2]
[1] ECE Department, UC San Diego
[2] CSE Department, UC San Diego
La Jolla, CA 92093, USA
skh004@ucsd.edu, kjeong@ucsd.edu, abk@ucsd.edu, jlu@cs.ucsd.edu

ABSTRACT

As the complexity of physical implementation continues to grow with technology scaling, routability has emerged as a major concern and implementation flow bottleneck. Infeasibility of routing forces a loop back to placement, netlist optimization, or even RTL design and floorplanning. Thus, to maintain convergence and a manageable number of iterations in the physical implementation flow, it is necessary to accurately predict design routability as quickly as possible. Routability estimation during placement typically exploits rough but fast global routers. Fast global routers are integrated with placers and are supposed to provide accurate congestion estimation for each iterative placement optimization, with short turn-around time. Such integrated global routers (as well as congestion estimators without global routers) should give (1) *fast*, and (2) *stably accurate* decisions as to whether a given placement is indeed routable.

In this paper, we evaluate four academic global routers [14] [1] [9] [4] in terms of *stability* and *scalability*. We perturb global routing problem instances in controlled ways, and analyze the sensitivity of routing outcomes and metrics. We observe scaling suboptimality and substantial noise in most of our experiments; this suggests a future need for new global router criteria and metrics.

1. INTRODUCTION

The placement phase of IC physical implementation seeks to determine locations of standard cells and logic units such that subsequent routing will achieve DRC-correct implementation of the netlist connectivity, subject to constraints on routing resources, total power, critical-path delays, etc. Standard-cell blocks often have high area utilization due to the availability of many routing layers. As a result, routability has become a major concern, since an unroutable result is equivalent to a design failure in the fixed-die regime. Turnaround time is a critical factor in competitive IC markets, and routing takes a large portion of the physical implementation turnaround time. Thus, fast and accurate routability prediction is needed at earlier stages of physical implementation.

Early congestion prediction relies on analytical models and/or specific net-shape models. Chen et al. [2] focus on fixed pattern of L- or Z-shape. Some works [6] [7] assign equal probability of route choices for congestion analysis. However, there are discrepancies between the used models and actual routing results. Actual routing shows that paths with fewer bends are dominant. The authors of [5] [13] classify impacts of paths according to number of bends, but this can be testcase-dependent. Another common deficiency of previous congestion estimators is that they target overall prediction accuracy, whereas over-congested areas specifically are the key concern in placers. Other prediction methods, e.g., [15], use Rent's rule for congestion analysis but ignore pin location information from placement that is essential for accuracy.

For more accurate congestion estimation, actual routing can be performed. Congestion estimation during placement can exploit rough but fast global routers. Pan et al. [10] compare placement results from a placer guided by a fast global router with those from the same placer guided by a simple congestion estimator, and show that accurate congestion estimation is essential to obtain routable placements. Fast global routers have been integrated with placers [12] [11] [10] to provide accurate congestion estimation for each iteration of placement optimization, with short turnaround time. Such integrated global routers (as well as congestion estimators without global routers) should give (1) *fast*, and (2) *stably accurate* decisions as to whether the placement results are routable. Otherwise, the resulting design after detailed routing may be unacceptable due to numerous DRC violations or large timing degradation from crosstalk noise and detouring on congested routes. Such routing problems, when observed in late design stages, require long and costly feedback loops to earlier design steps.

In this paper, we examine global routers' behavior in the presence of routing instance *offset noise*, *resource noise* and *scaling*. We discover that there are limits to the usability of global routers for stable routability evaluation. The remainder of this paper is organized as follows. Section 2 presents the metrics to quantify global routing solutions. Section 3 shows the performance evaluation of academic global routers[1] with ISPD benchmarks, in terms of stability and scalability. Finally, Section 4 gives discussions and conclusions.

[1]Due to the inability to access and control internal data models used in commercial global routers, our work is limited to academic global routers.

2. GLOBAL ROUTING INSTANCE PARAMETERS AND ROUTABILITY METRICS

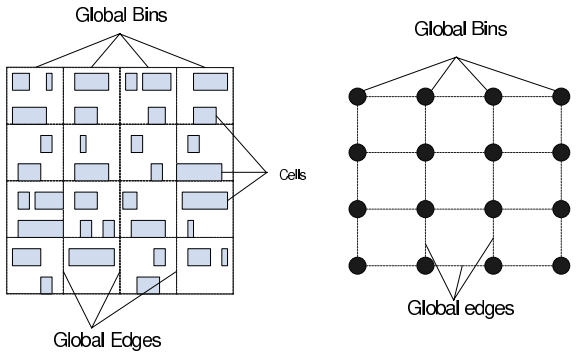


Figure 1: An example of design decomposed into global cells (gcells). Adjacent gcells are connected by global edges (gedges).

In our discussion, we use the following notation and terminology. The design area is decomposed into rectangular global cells, or *gcells*. The boundary between two adjacent gcells is a global edge, or *gedge*. An example global routing instance is shown in Figure 1. The global routing instance can be described using the following parameters.

- G: a global routing instance
- N: number of nets
- t_i: number of terminals of net i
- w_i: minimum width of net i
- g (or $g_{i,j}$): global routing cell (=*gcell*), (or gcell at i^{th}-row and j^{th} column of an array of gcells)
- e: edges of a gcell (=*gedge*); e_l, e_r, e_t, and e_b denote the left, right, top, and bottom edge, respectively
- w_g, and h_g: width and height of a gcell
- n_r, and n_c: number of rows (n_r) and columns (n_c) of gcells in G
- $o(G) = (o_x, o_y)$: origin, i.e., bottom-left corner of the bottom-left gcell of G
- L: maximum number of routing layers, not including via layers
- $d(l) \in \{h, v\}$: preferred routing direction of a routing layer l, i.e., horizontal (h) or vertical (v)
- c: capacity, maximum allowed number of routing segments per gedge; $c_v(l)$ or $c_h(l)$) is capacity of a gedge for a routing layer l in vertical or horizontal direction, respectively
- $w_{min}(l)$: default minimum width of layer l
- $s_{min}(l)$: minimum spacing of layer l
- $B = \{b_i(j) \mid i = 1, ..., N_B$ and $j = 1, ..., L\}$: set of routing blockages $b_i(j)$ defined between two adjacent gcells, for routing layer j

In general, routing congestion is measured by utilization (u) of gedges; this is calculated as the number of tracks (r) used on a gedge by routing segments and routing blockages, divided by the total track resources (c) of the gedges. Various routing metrics can be applied to define the gedge utilization.

- TOF: Total overflow. The overflow of a gedge is the utilization by routing segments beyond the gedge capacity. Total overflow is the sum of overflow over all gedges in the global routing instance.
- MOF: Maximum overflow. The maximum overflow of a gedge.
- $WCI(A)$: Worst congestion index. The number of nets crossing any gedge(s) whose utilization is more than $A\%$.
- $U(A)$: Average net-score of the top-$A\%$ highest net-score nets. The *net-score* of a net is the maximum utilization among all gedges through which segments of the net pass.
- TWL: Total wirelength.

TOF, MOF and TWL are specified in the ISPD 2008 Global Routing Contest [8] as quality metrics of global routers. $WCI(A)$ and $U(A)$ are industry metrics currently used in a production ASIC flow [16]. Additionally, the congestion map can be used as a (human-interpretable) quality metric.

3. STABILITY AND SCALABILITY OF GLOBAL ROUTERS

The perceived solution quality of any global router will vary as a function of routing parameters such as number of tracks on gedges, size of gcells, number of routing layers and blockages, technology rules, minimum width/space, etc., and various router control parameters. While various metrics are used and reported in commercial and academic routers, these metrics do not give a clear '0 or 1' decision for routability, since there are no golden threshold values for these metrics.

In this section, we show inherent instabilities or inconsistencies of current global routers and metrics. We use the routability metrics above to evaluate *stability* and *scalability* of *FastRoute-4.1 [14]*, *NTHU-2.0 [1]*, *FGR-1.2 [9]*, and *NTUgr-1.1 [4]*, with the ISPD 2008 global routing benchmark suite summarized in Table 1.

3.1 Stability to Offset Noise

A *grid offset* ($o(G)$) shifts the entire gcell array relative to the fixed pin locations of the placement. By applying a grid offset, the set of pins covered by a gcell may be changed. Hence, the shift can also change the global routing instance, unless the shift is an integer multiple of the gcell size. In practice, the effect of offset is commonly ignored,[2] since most pin-to-gcell assignments remain unchanged by small offset changes. Intuitively such small changes should not significantly affect routing solutions.

We evaluate the stability of global routers to offset changes, as follows.

(1) Shift the entire gcell array in the bottom-left direction as much as possible, as long as all pins are covered by the gcell array. Let the coordinate of the origin (i.e., the bottom-left corner of the bottom-left gcell) of the array be (x_{min}, y_{min}).

[2]Gcell merging at hard-macro boundaries is a known technique when variable gcell size is permitted.

Table 1: Benchmarks for ISPD global routing contest.

Benchmarks	#nets	G dimensions		Orig. offset		Gcell size	
		n_c	n_r	o_x	o_y	w_g	h_g
adaptec1	219794	324	324	136	136	35	35
adaptec2	260159	424	424	216	216	35	35
adaptec3	466295	774	779	21	43	30	30
adaptec4	515304	774	779	21	43	30	30
adaptec5	867441	465	468	11	33	50	50
newblue1	331663	399	399	218	218	30	30
newblue2	463213	557	463	7	7	50	50
newblue3	551667	973	1256	16	19	40	40
newblue4	636195	455	458	4	12	40	40
newblue5	1257555	637	640	16	47	40	40
newblue6	1286452	463	464	6	46	60	60
newblue7	2635625	488	490	0	9	80	80
bigblue1	282974	227	227	128	128	50	50
bigblue2	576816	468	471	16	56	40	40
bigblue3	1122340	555	557	11	51	50	50
bigblue4	2228903	403	405	0	18	80	80

(2) Shift the entire gcell array in the top-right direction as much as possible, as long as all pins are covered by the gcell array. Let the coordinate of the origin (i.e., the bottom-left corner of the bottom-left gcell) of the array be (x_{max}, y_{max}).

(3) For $o_x = x_{min}$ to x_{max}
  For $o_y = y_{min}$ to y_{max}
    Run global routing with offset (o_x, o_y)
    Record the routability metrics of Section 2

Figures 2(a), 2(b), 2(c) and 2(d) show the changes of congestion metrics (i.e., TOF, MOF, $WCI(100)$, and $U(20)$ - plotted on y-axis) with respect to (o_x,o_y) offset values (given in units of tracks) for the two global routers *FastRoute* [14] *NTHU* [1], using the ISPD 2008 'bigblue4' benchmark. We observe large fluctuations in the metric values due to offset changes. For example, the number of nets passing over 100% congested gcells ($WCI(100)$) can change from 426 to 1611 in *FastRoute*. Another interesting observation from these figures is that we cannot determine which router is better than the other, since there are intersections between the two curves. For some cases, *FastRoute* shows better quality, but for the others, *NTHU* shows better quality. In general, *NTHU* performs more stably than *FastRoute*. From these observations, we can conclude that global routers are not robust for small input changes, and further, that the quality rank was not stably determined in the ISPD 2008 contest.

3.2 Stability to Resource Noise

Effective capacity can be defined as capacity minus blockage. It is the maximum number of routing segments that can traverse a given gedge, and is an indicator of available routing resources. Despite varying capacity and blockages, the global routing instance can be regarded as unchanged if the effective capacity is unchanged. We have evaluated another stability aspect of global routers, i.e., *resource noise*. Here we measure the quality metrics based on varied blockage units but constant effective capacity of every gedge. For example, we increase gedge capacities from c to $c + d$, but add d blockages to each gedge, so that the original effective capacity is maintained. Figure 3 illustrates an example of adding resource noise.

Our experimental procedure is described as follows.

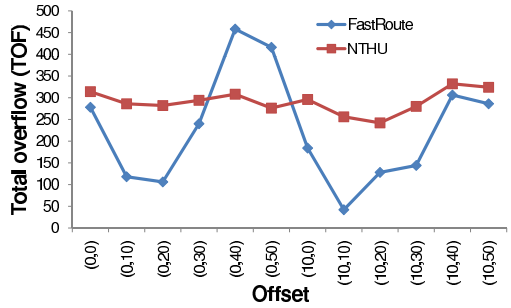
(1) Remove all blockages from the benchmarks
(2) Foreach $d \in \{0, 1, 5, 20\}$
  Add d blockages to all gedges
  Replace capacity c of all gedges by $c' = c + d$
  Run global routing with new gedge capacities c'
  Record the metrics of Section 2

Figures 4(a), 4(b), 4(c), and 4(d) show variation of quality metrics with respect to resource noise d for four global routers [14] [1] [4] and [9], using the ISPD 2008 'newblue1' benchmark.

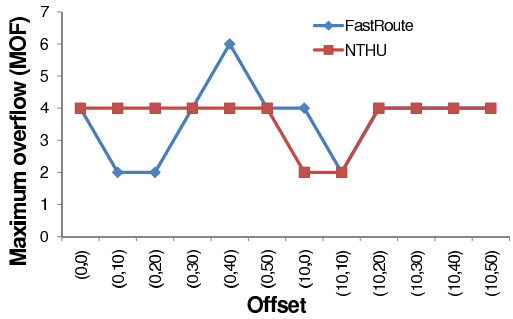
The interesting observation is again that there can be non-trivial fluctuations with respect to the introduced noise, e.g., TOF of *NTHU* can change from 132 to 160. This being said, the tested global routers show comparative stability with respect to resource noise. There is no crossover between curves, e.g., *FastRoute* always shows best quality, and *FGR* always shows worst quality, for all metrics and for all resource noise d values.

3.3 Scalability

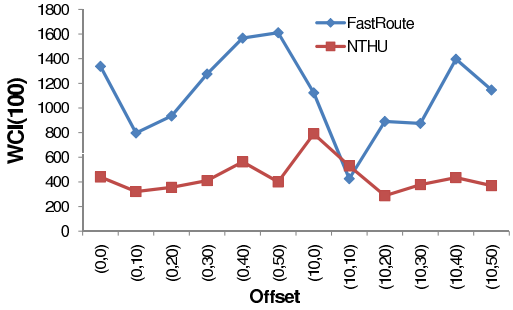
Our third experiment evaluates the *scaling behavior* of global routers. We generate twice ($X2$), three times ($X3$) and four times ($X4$) larger benchmarks from the given benchmarks. For example, to generate $X2$ benchmarks, we duplicate all pins and nets of the original benchmarks, and we also enlarge gedge capacity (and hence gcell dimension) to twice their original values, as shown in Figure 5. By doing so, the difficulty of routing, i.e., the number of nets per unit routing resource, will not be changed. Further, as noted in [3], an *upper bound* for the optimal solution of the $X2$ instance



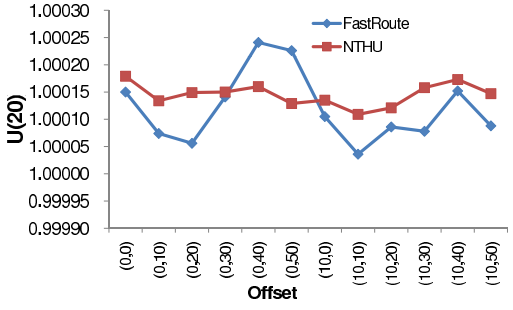
(a) TOF variation



(b) MOF variation



(c) $WCI(100)$ variation



(d) $U(20)$ variation

Figure 2: Variation of routing quality due to offset noise.

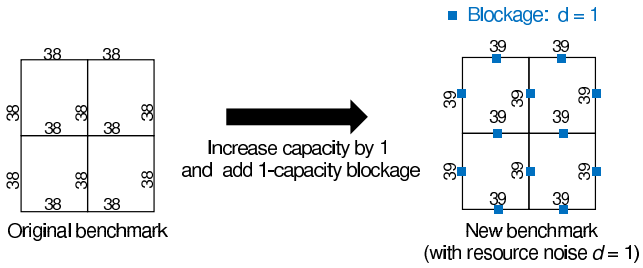


Figure 3: Resource changes that maintain effective capacity. Numbers on edges show specified capacities.

would simply be two copies of the optimal solution for the $X1$ instance. For the scaled benchmarks, we again assess currently-used routability metrics TOF, MOF, $WCI(100)$, and $U(20)$.

Our experimental procedure is as follows.

(1) Remove all blockages from the benchmarks
(2) Foreach $S \in \{1, 2, 3, 4\}$
  Increase gcell dimensions w_g and h_g by S times
  Increase capacity c by S times
  Move all pins (x_i, y_i) to $(x_i \cdot S, y_i \cdot S)$
  Duplicate all nets by S times with new net IDs
  Run global routing with S-scaled instance
  Record the metrics of Section 2

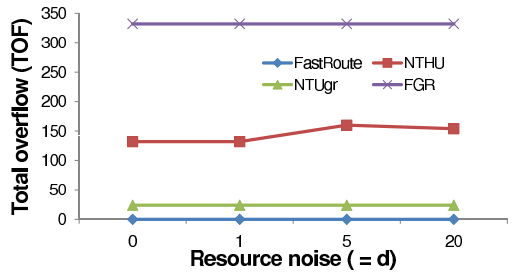
Figures 6(a), 6(b), 6(c) and 6(d) show the variation of quality metrics with respect to the scaling factor (x-axis) for four global routers and the 'bigblue4' benchmark. Each metric value of scaled benchmarks is normalized to that of $X1$. The dashed line in each figure indicates the ideal value of each metric over the scaled benchmarks.

During benchmark scaling, effective capacity is maintained by increasing both supply (i.e., capacity) and demand (i.e., the number of nets to route) by the same factor. As a result, *TOF*, *MOF* and *WCI(100)* should scale linearly, and *U(20)* should remain unchanged, as the scaling factor increases. However, we see that all metric values can worsen rapidly for some routers, compared to the ideal outcome. *FastRoute* has the worst scalability with respect to *TOF* and *WCI(100)*. *NTUgr* and *FGR* scale badly with respect to *MOF* and *U(20)*. *NTHU* has relatively better scalability compared to other routers. The overall poor scaling may present an obstacle for use of constructive, global router-based congestion prediction in placement tools.

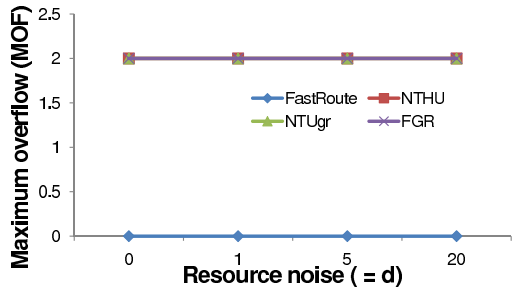
Figure 6(e) shows the runtime scaling of the global routers, again normalized to the value of $X1$. Ideally, runtime should increase linearly with the number of nets to route, as denoted by the dashed line. However, *FastRoute* and *NTUgr* scale very badly compared to the other two global routers.

4. DISCUSSION AND CONCLUSION

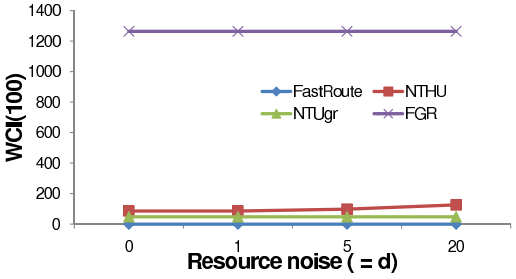
Global routing-based congestion estimation is a standard technique in modern design flows. Effective routability estimation should be stable for global routers over a wide range



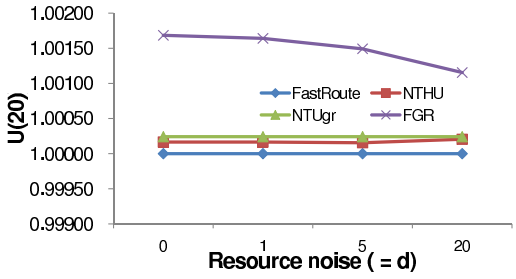
(a) TOF variation



(b) MOF variation



(c) $WCI(100)$ variation



(d) $U(20)$ variation

Figure 4: Variation of routing quality due to resource noise.

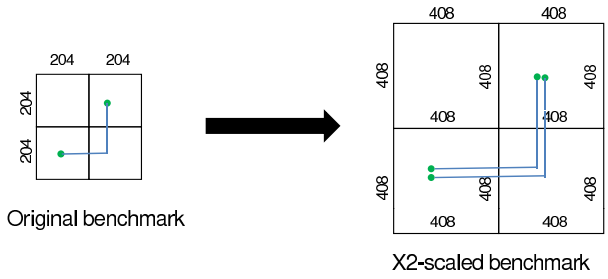


Figure 5: An example of scaling benchmarks. Numbers on edges show specified capacities. Dots and lines across gedges represent pins and nets.

of routing instances. In this paper, we have empirically studied four modern academic global routers with respect to *instance offset noise*, *resource noise* and *scaling*. According to our experimental results, all the routers show room for improvement according to these criteria.

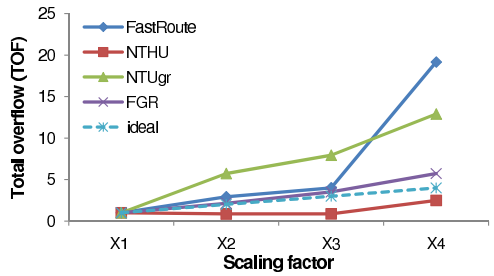
From the examination of router source codes, we may speculate that some instability may result from tuning of routers to specific benchmark instances, i.e., instance-specific tuning via many knobs may lose its advantages (or even cause negative effects) on other benchmarks. Another potential reason of instability may be the over-reduction of congestion in academic routers. Moderate congestion is usually tolerable in early-stage estimation at early stages with less than 10% impact on wiring. However, academic routers heavily focus on congestion minimization due to competition metrics, resulting in many detours and extra vias. This complicates the optimization and enhances the sensitivity towards noise.

Separately from the routers themselves, there are also open challenges related to unstable metrics for global routing performance evaluation. All of the metrics (*TOF, MOF, WCI(100), U(20)*) vary significantly over different gcell definitions, which reflects inherent limits (i.e., bounds) on stability. New metrics with better stability may be needed to facilitate future global routing research.

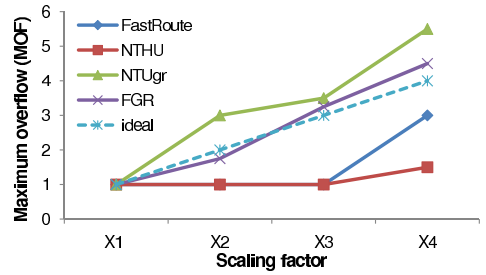
Although we have not provided clear reasons for the noisy behaviors of routers nor specific techniques to mitigate such instability, we believe that our results can provide some guidance for future developments of global routers and congestion-driven placers, as well as their integrations. Our ongoing work seeks to develop stable routability metrics, and fast and accurate routability estimation techniques. With such techniques in place, our ultimate goal will be to integrate fast routability feedback within congestion-driven placers. Finally, we hope to study industrial global routing tools to determine whether they show similar behaviors with respect to noise and scalability.

5. ACKNOWLEDGMENTS

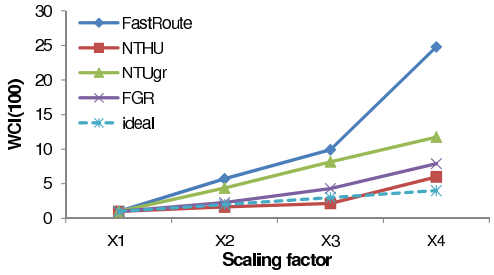
We thank Dr. Charles Alpert, Dr. Zhuo Li, Dr. Cliff Sze, and Dr. Natarajan Viswanathan of IBM Austin Research Laboratory for motivating discussions and feedback on these studies.



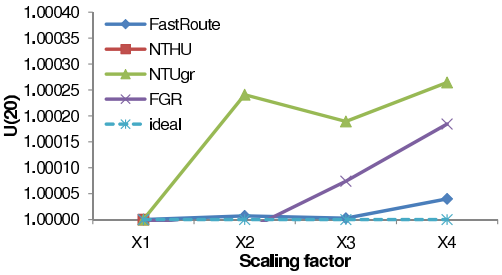
(a) Normalized TOF variation



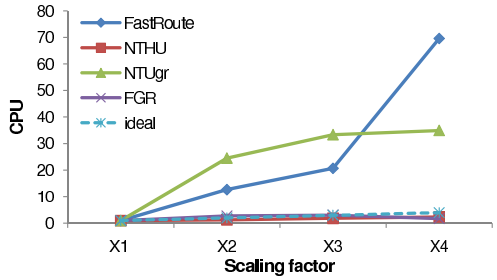
(b) Normalized MOF variation



(c) Normalized $WCI(100)$ variation



(d) Normalized $U(20)$ variation



(e) Normalized CPU time variation

Figure 6: Variation of routing quality due to scaling noise.

6. REFERENCES

[1] Y.-J. Chang, Y.-T. Lee and T.-C. Wang, "NTHU-Route 2.0: A Fast and Stable Global Router", *Proc. IEEE/ACM International Conference on Computer Aided Design*, 2008, pp. 338-343.

[2] H.-M. Chen, H. Zhou, F. Y. Young, D. F. Wong, H. H. Yang and N. Sherwani, "Integrated Floorplanning and Interconnect Planning", *Proc. IEEE/ACM International Conference on Computer Aided Design*, 1999, pp. 354-357.

[3] L. Hagen, J. H. Huang and A. B. Kahng, "Quantified Suboptimality of VLSI Layout Heuristics", *Proc. ACM/IEEE Design Automation Conference*, 1995, pp. 216-221.

[4] C.-H. Hsu, H.-Y. Chen and Y.-W. Chang, "High-Performance Global Routing with Fast Overflow Reduction", *Proc. IEEE/ACM Asia and South Pacific Design Automation Conference*, 2009, pp. 582-587.

[5] A. B. Kahng and X. Xu, "Accurate Pseudo-Constructive Wirelength and Congestion Estimation", *Proc. International Workshop on System-level Interconnect Prediction*, 2003, pp. 61-68.

[6] Kusnadi and J. D. Carothers, "A Method of Measuring Nets Routability for MCM's General Area Routing Problems", *Proc. International Symposium on Physical Design*, 1999, pp. 186-194.

[7] J. Lou, S. Thakur, S. Krishnamoorthy and H. S. Sheng, "Estimating Routing Congestion Using Probabilistic Analysis", *IEEE Transactions on Computer-Aided Design* 21(1) (2002), pp. 32-41.

[8] G.-J. Nam, C. Sze and M. Yildiz, "The ISPD Global Routing Benchmark Suite", *Proc. International Symposium on Physical Design*, 2008, pp. 156-159.

[9] J. A. Roy and I. L. Markov, "High-Performance Routing at the Nanometer Scale", *Proc. IEEE/ACM International Conference on Computer Aided Design*, 2007, pp. 496-502.

[10] M. Pan and C. Chu, "IPR: An Integrated Placement and Routing Algorithm", *Proc. ACM/IEEE Design Automation Conference*, 2007, pp. 59-62.

[11] M. Wang and M. Sarrafzadeh, "Modeling and Minimization of Routing Congestion", *Proc. ACM/IEEE Design Automation Conference*, 2000, pp. 185-190.

[12] M. Wang, X. Yang and M. Sarrafzadeh, "Congestion Minimization During Placement", *IEEE Transactions on Computer-Aided Design* 21(1) (2002), pp. 71-80.

[13] J. Westra, C. Bartels and P. Groeneveld, "Probabilistic Congestion Prediction", *Proc. International Symposium on Physical Design*, 2004, pp. 204-209.

[14] Y. Xu, Y. Zhang and C. Chu, "FastRoute 4.0: Global Router with Efficient Via Minimization", *Proc. IEEE/ACM Asia and South Pacific Design Automation Conference*, 2009, pp. 576-581.

[15] X. Yang, R. Kastner and M. Sarrafzadeh, "Congestion Prediction During Top-Down Placement", *IEEE Transactions on Computer-Aided Design* 21(1) (2002), pp. 72-80.

[16] C. Alpert et al., *personal communication*, 2010.