

New Timing and Routability Driven Placement Algorithms for FPGA Synthesis

Yue Zhuo and Hao Li
Department of Computer Science and
Engineering
University of North Texas
Denton, Texas, USA

Qiang Zhou, Yici Cai and Xianlong Hong
Department of Computer Science and
Technology
Tsinghua University
Beijing, China

ABSTRACT

We present new timing and congestion driven FPGA placement algorithms with minimal runtime overhead. By predicting the post-routing critical edges and estimating congestion accurately, our algorithms simultaneously reduce the critical path delay and the minimum number of routing tracks. The core of our algorithm consists of a criticality-history record of connection edges and a congestion map. This approach is applied to the 20 largest MCNC benchmark circuits. Experimental results show that compared with VPR [2], our algorithms yield an average of 8.1% reduction (maximum 30.5%) in the critical path delay and 5% reduction in channel width. Meanwhile, the average runtime of our algorithms is only 2.3X as of VPR's.

Categories and Subject Descriptors

B.7.2 [Hardware, Integrated Circuit, Design Aids]: Placement and Routing

General Terms

Algorithms, Design, Performance, Congestion

Keywords

Timing Driven Placement, Congestion Driven Placement, Physical Synthesis, Net Weight

1. INTRODUCTION

Two important metrics of a field programmable gate array (FPGA) circuit are its size and speed. The chip area of an FPGA consists of routing area and logic area. For an FPGA, 80%-90% of the chip area is dedicated to the routing resources [7]. Thus, reducing routing tracks in FPGA will significantly reduce the area. Also, interconnect delay assumes the major proportion of the total FPGA delay and hence becomes crucial in optimizing device speed.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GLSVLSI'07, March 11–13, 2007, Stresa-Lago Maggiore, Italy.
Copyright 2007 ACM 978-1-59593-605-9/07/0003 ...\$5.00.

The major contribution of this work is new placement algorithms which optimize both timing and congestion. The rest of this paper is organized as follows: In Section 2, we present the background and discuss related research works. In Section 3, we briefly review Versatile Place and Route (VPR) since it is the framework into which our algorithms are integrated. In Section 4, we propose the details of our placement algorithms. In Section 5, we present the experimental results. In Section 6, we draw the conclusions and discuss future research.

2. RELATED WORKS

Current research in FPGA placement is rich in optimizations of performance, routability and runtime.

2.1 Timing-Driven Approaches

In placement, timing-driven algorithms can be broadly divided into two classes: path-based and net-based. Path-based algorithms try to compute the delay of all paths and directly minimize the longest path delay [9, 10, 14]. Net-based algorithms, on the contrary, do not handle path-based constraints directly [8, 11, 16]. Instead, they usually transform timing constraints on paths into either net-length or net-weight constraints. A popular approach is called *net-weighting* method. In such approaches, static timing analysis is applied at intermediate phases and nets are assigned criticality weights; higher weights are assigned to nets which are more timing critical.

To perform net-weighting analysis, a directed graph $G(V, E)$ representing the circuit is constructed. Each wire or logic block pin becomes a node in the graph. Each switch becomes a directed edge or a pair of directed edges between two appropriate nodes. Every edge is annotated with a physical delay between the nodes. A *source* is the pin of an input pad or a register output, while a *sink* is the pin of an output pad or a register input.

Given a node j , its *arrival time*, $Arr(j)$, can be computed from the following equation:

$$Arr(j) = \begin{cases} 0, & j \in sources \\ \max\{Arr(i) + Delay(i, j)\}, & (i, j) \in E \end{cases} \quad (1)$$

where $Delay(i, j)$ is the delay value of the edge connecting node i and node j . The maximum arrival time of all nodes in the circuit, D_{max} , can be computed as:

$$D_{max} = \max\{Arr(j)\}, j \in sinks. \quad (2)$$

For node i , the *required arrival time*, $Req(i)$, can be computed as follows:

$$Req(i) = \begin{cases} D_{max}, & i \in sinks \\ \min\{Req(j) - Delay(i, j)\}, & (i, j) \in E \end{cases} \quad (3)$$

Finally, we can determine the *slack* of an edge (i, j) by:

$$Slack(i, j) = Req(j) - Arr(i) - Delay(i, j) \quad (4)$$

Under this framework, Kong proposed an efficient all-path counting algorithm called PATH [11]. Wang *et al.* tried to improve timing by using linear programming (LP) relaxation [16]. The timing-driven part of our proposed algorithm in this paper is based on the net-weighting analysis.

Differing from the above approaches which work in placement phase, Lin *et al.* proposed an algorithm named SMAC recently optimizing timing during mapping and packing [12]. But this algorithm introduces a high area overhead.

2.2 Routability-Driven Approaches

The problem of routability optimization is often taken care of in the phase of packing [4] or placement [3, 18]. Bozorgzadeh *et al.* presented a routability-driven clustering method called t-RPack [4]. In VPR, there is an optional *non-linear congestion* method [3], but it is only used to be compared with VPR's *linear congestion* method since it requires too much runtime. We *et al.* proposed a congestion-map based method to reduce channel width [18], which will be called *cMap* throughout. In this paper, we extended *cMap* to reduce congestion.

2.3 Runtime

The runtime of a simulated annealing (SA) based algorithm is usually longer than an *analytical placement* algorithm, such as *force-directed* based approaches. To overcome this drawback, many methods were tried. One is to enable SA only when the temperature is low, like Frontier, a system developed by Tessier to achieve highly routable and high-performance layouts [15]. Another one is utilizing hardware to assist SA, like the systolic structure proposed by Wrighton *et al.* in [17].

3. PLACEMENT IN VPR

VPR uses the following cost function [13] in its placement:

$$\Delta C = \lambda \cdot \frac{\Delta C_T}{Previous_C_T} + (1 - \lambda) \cdot \frac{\Delta C_W}{Previous_C_W} \quad (5)$$

where C_T is the timing cost, C_W is the wiring cost, and λ is a constant between 0 and 1 which trades off between timing cost and wiring cost. $Previous_C_T$ and $Previous_C_W$ are updated at the beginning of every temperature and used by all moves at this temperature. The equations to compute C_T are given below [13]:

$$C_T = \sum_{\forall i, j \in circuit} C_T(i, j) \quad (6)$$

$$C_T(i, j) = Delay(i, j) \cdot Crit(i, j)^\beta \quad (7)$$

$$Crit(i, j) = 1 - \frac{Slack(i, j)}{D_{max}} \quad (8)$$

where $C_T(i, j)$ is the timing cost for each *edge* (i, j) . The purpose of including an exponent β on the *Crit* in timing cost function is to heavily weight connections that are critical, while giving less weight to connections that are non-critical. In the process of SA, β increases slowly from 1 to a user-defined maximal value which is 8 by default. The wiring cost is defined as [13]:

$$C_W = \sum_{i=1}^{N_{nets}} q[i](bb_x(i) + bb_y(i)) \quad (9)$$

where N_{nets} is the total number of nets in the circuit. For each net i , $bb_x(i)$ is its horizontal span, and $bb_y(i)$ is its vertical span. The values of $q(i)$ are obtained from [6].

4. OUR PROPOSED APPROACH

To compare with VPR, our algorithm is also based on SA. We use the same top-level cost function in Equation (5) and the top-level timing cost function in Equation (6). The formulas to compute the timing cost for each edge and the wiring cost are modified and shown below:

$$C_T(i, j) = History(i, j) \cdot Delay(i, j) \cdot Crit(i, j)^\beta \quad (10)$$

$$C_W = Congestion \cdot \sum_{i=1}^{N_{nets}} q[i](bb_x(i) + bb_y(i)) \quad (11)$$

Compared with Equation (7) and (9), the new elements are the *History* (i, j) factor and the *Congestion* factor. We will explain the details of *History* (i, j) in Section 4.1 and *Congestion* in Section 4.2.

4.1 Our Proposed Timing Driven Approach

Since the cost to compute the criticality for each edge from a timing-analysis graph is very high, general timing analysis algorithms can only afford executing this computation once every temperature. As a result, the accuracy becomes less reliable because there are tens of thousands of moves per temperature. The delay of an edge may differ a lot from the original value when the graph was annotated.

Our approach addresses the accuracy problem without paying much run time penalty. We observed the following two facts: 1) The criticality of a given edge varies in multiple rounds of timing analysis. 2) Some edges are almost always among the most critical edges. In other words, no matter how the CLBs and IO pads are placed, the criticality value of some edges are always high.

Our approach takes advantage of the second phenomenon. In our algorithm, the criticality value of an edge (i, j) is accumulated into a variable *CritStat* (i, j) every time it is computed. An edge which is often timing-critical will get a high accumulation value. We then favor those historically timing-critical edges by using this statistical data. The accumulation value *CritStat* (i, j) is computed as:

$$CritStat(i, j) = \left(\sum_{k=1}^{N_t} Crit_k(i, j) \cdot \alpha^{N_t-k} \right) / \left(\sum_{k=1}^{N_t} \alpha^{N_t-k} \right) \quad (12)$$

where N_t is the number of different temperatures since the beginning of the SA. $Crit_k(i, j)$ is the value of *Crit* (i, j) at the k th temperature. α is a decay constant and we use 0.96 based on our experiments. As we can see, *CritStat* (i, j)

is a weighted mean over all $Crit(i, j)$ in the history. The weight for the k th $Crit(i, j)$ is $\alpha^{N_t - k}$. In other words, the criticalities at different temperatures are not treated equally. We assign more weight to a recent criticality value since it is more reliable. Utilizing the whole criticality history, we avoid the randomness encountered by general timing-analysis methods and hence increase the accuracy.

The $History(i, j)$ factor, used in Equation (10), is computed from $CritStat(i, j)$ as defined in Equation (12). In order to compute $History(i, j)$, we divide the SA process into two phases based on the swapping radius. In VPR, the maximal radius to swap two blocks decrease gradually as the temperature cools down. As long as the radius to swap two blocks is greater than 1.0, we define the SA to be in $Phase_0$, otherwise it is in $Phase_1$.

In $Phase_0$, we update $CritStat(i, j)$ at every temperature, but do not use it to compute $History(i, j)$ because the statistical data are not enough as guidance. So, in $Phase_0$, $History(i, j)$ is always 1.0 and does not affect timing costs.

In $Phase_1$, we keep updating $CritStat(i, j)$ and use it to compute $History(i, j)$. Let N_C denote the number of edges considered most timing-critical in history. At the beginning of each temperature, we recompute $CritStat(i, j)$ for every edge and select N_C edges with the highest $CritStat$ values. We call these edges potential critical edges (*PCE*). Also we set the variable $CritThres$ to the $(N_C + 1)$ th highest $CritStat$. Then we compute $History(i, j)$ by the following formula:

$$History(i, j) = \max(CritStat(i, j) - CritThres + 1, 1) \quad (13)$$

It can be seen that for the N_C edges with the highest criticalities in the history, their $History(i, j)$ values are greater than 1. Assume that there are N_{edges} edges in the timing-analysis graph, then the values of $History(i, j)$ for all the remaining $(N_{edges} - N_C)$ edges are 1. We only favor the first N_C edges in the criticality history. And the more critical an edge is in the history, the more it is favored.

In our algorithm, N_C is a constant for a particular circuit. Intuitively, the value of N_C should be as small as possible. Since only the edges with the potentials to appear in the post-routing longest path should be favored. On the other hand, it should not be too small, otherwise the probability of missing a final critical edge is high. We use the following formula to compute N_C :

$$N_C = \max(\sqrt{1.54 \cdot N_{edges}} \cdot (1 - 4 \cdot empty_rate), 64) \quad (14)$$

where $empty_rate$ is the percentage of unoccupied CLBs on the chip. First, N_C grows as the total number of edges increases. This is natural because we need to consider more edges as the circuit becomes larger and more complex. Second, N_C increases as $empty_rate$ decreases. The reason is that when $empty_rate$ is low, the layout is compact and relatively more congested. In this case, favoring a critical edge may easily force other non-critical edges to take detours and become critical. So we need to consider more edges simultaneously. The constant 1.54 and 64 are acquired from experiment data.

Two techniques are used to further improve our proposed algorithm. First, this algorithm works better in a less congested environment because it will be easier to favor the

Table 1: Temperature Update Schedule

Fraction of moves accepted (R_{accept})	γ
$R_{accept} > 0.96$	0.65
$0.8 < R_{accept} \leq 0.96$	0.976
$0.15 < R_{accept} \leq 0.8$	0.996
$R_{accept} \leq 0.15$	0.93

PCEs without influencing other edges. As a result, we set $\lambda = 0.3$ in Equation (5) when $empty_rate < 2\%$ to alleviate congestion. The value of λ for each circuit is shown in Table 2.

The other technique used is to get more criticality data. Our algorithm is based on the statistics of criticalities and works better when it gathers more history information. Therefore, we need to modify the annealing schedule. The goal is to keep the total number of moves about the same as in VPR but increase the number of different temperatures. The number of moves evaluated at each temperature is $(N_{blocks})^{1.33}$ in our algorithm, about 1/10 as in VPR. And the number of different temperature is about 10 times as in VPR. A new temperature is computed as $T_{new} = \gamma T_{old}$, where γ depends on the fraction of attempted moves that were accepted (R_{accept}) at T_{old} , as shown in Table 1.

4.2 Our Proposed Routability Driven Approach

A metric characterizing the congestion status of a placement is introduced in cMap [18]. The essence of this approach is straightforward. First, every net fits within a corresponding bounding box. Second, each net probably needs some routing tracks around the CLBs within its bounding box. So, we can estimate the routing resource requirement around a CLB by counting the number of bounding boxes covering it. The congestion factor used in Equation (11) is computed by the following formula [18]:

$$Congestion = \left(\frac{\sum_{x,y} U_{x,y}^2}{nx \cdot ny} / \left(\frac{\sum_{x,y} U_{x,y}}{nx \cdot ny} \right)^2 \right)^k, \quad 1 \leq x \leq nx, 1 \leq y \leq ny \quad (15)$$

where $U_{x,y}$ is the number of bounding boxes covering $CLB_{x,y}$, k is a small positive integer, and the whole chip consists of nx by ny CLBs. This algorithm is better explained in Fig.2. Fig.1(a) shows a placement for a circuit consisting of three nets. We assume the target FPGA chip consists of 6x6 CLBs. Fig.1(b) shows the corresponding U array for this placement. The number in each CLB indicates how many bounding boxes are covering this CLB at this moment. A CLB without a label is not covered by any bounding box. For example, the value of $U_{3,3}$ is 3 since it is inside the bounding box of all the 3 nets, and the value of $U_{1,1}$ is assigned 0 because it is not inside any bounding box. In Fig.1(b), only the CLBs used by the circuit are shaded. The value of $Congestion$ shows how congested the whole placement is expected to be. If it is small, the placement is balanced, otherwise the placement is congested.

In FPGA based designs, we can hardly utilize all CLBs and hence those unused CLBs are generally neglected and wasted. However, they are valuable since they do not consume any routing resources. Instead, routing tracks around them can be used by other CLBs. So, unused CLBs should

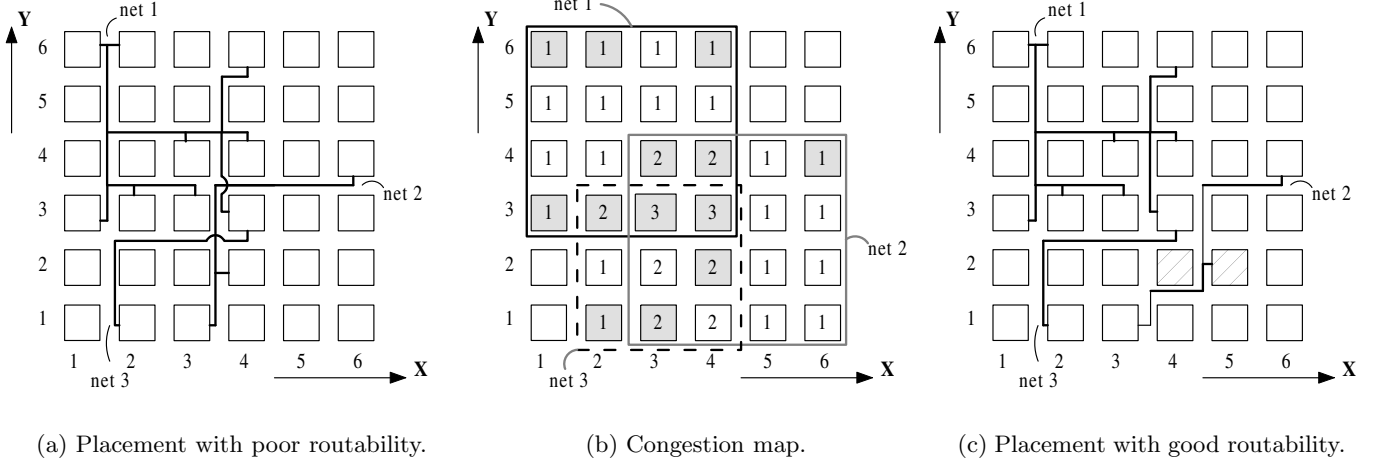


Figure 1: A circuit with three nets: our goal is to achieve (c).

be placed in the most congested regions as long as this placement does not degrade other metrics too much. Fig.1(c) is a placement after swapping $CLB_{4,2}$ and $CLB_{5,2}$ (striped) in Fig.1(a). This swap will probably reduce channel width since the traffic at the center of the chip is reduced. But this swap will not be favored by cMap since these two placements have the same U array as shown in Fig.1(b), and thus the same value of *Congestion*. Similarly, it can not be favored by VPR’s wiring cost function because the dimension and position of all the three nets do not change at all.

Our approach solves this problem by considering unoccupied CLBs and IO pads. The objective is to make the factor *Congestion* smaller when more unoccupied CLBs are placed in congested regions. To quantify the alleviation of congestion brought by an unused block, we define a factor $Sat_{x,y}$ indicating the effective percentage of $U_{x,y}$. That is, for $CLB_{x,y}$, $(1 - Sat_{x,y})$ fraction of $U_{x,y}$ is alleviated. The following formulas show how to compute $EU_{x,y}$, the effective percentage of $U_{x,y}$,

$$EU_{x,y} = U_{x,y} \cdot Sat_{x,y} \quad (16)$$

$$Sat_{x,y} = \begin{cases} 1 - N_{un} \cdot allev, & CLB_{x,y} \text{ is used} \\ 1 - N_{un} \cdot allev/2, & \text{otherwise} \end{cases} \quad (17)$$

$$allev = \min(\beta/empty_rate, 6\%) \quad (18)$$

where N_{un} is the number of unused CLBs or IO pads adjacent to $CLB_{x,y}$. The maximum value of N_{un} can be 4. We need consider the number of empty blocks adjacent to each CLB. Preferably, we would like to compute $Sat_{x,y}$ in such a way that the more empty blocks $CLB_{x,y}$ is adjacent to, the smaller $Sat_{x,y}$ becomes. This is taken into account by introducing N_{un} in Equation (17). On the other hand, we want to avoid placing a large number of unused blocks in the same region in order to reduce congestion. This is why we use $allev/2$ in Equation (17) when $CLB_{x,y}$ is not occupied. Several facts can be inferred from these equations. First, if $CLB_{x,y}$ is not adjacent to any unused blocks, its N_{un} is 0 and $Sat_{x,y}$ is 1, so $EU_{x,y}$ equals $U_{x,y}$. Second, if $CLB_{x,y}$ is adjacent to some unused block(s), its $EU_{x,y}$ is

less than $U_{x,y}$. Third, when adjacent to the same N_{un} number of empty blocks, $U_{x,y}$ of an unused CLB, $CLB_{x,y}$, is not alleviated as much as a used CLB.

It is clear that the value of *allev* is very important. If it is too small, the effect will be negligible. If it is too large, the locations of empty blocks will become a primary factor when computing the *Congestion* factor in Equation (19). As a result, all empty blocks will be converged to the center of the chip, which is not favorable. Based on our experiments, our algorithm performs well when the *allev* factor is smaller than 6%. In general, we want to restrict the value of *allev* between 2% to 6%. Since most circuits have an *empty_rate* also between 2% to 6%, the value of β is set to $2\% \cdot 6\% = 0.12\%$ in our approach. Also we put an upper bound of 6% to *allev* in Equation (18) in case some circuit has an extremely small *empty_rate*.

We use the following *modified* congestion function in our approach:

$$Congestion = \left(\frac{\sum_{x,y} EU_{x,y}^2}{nx \cdot ny} / \left(\frac{\sum_{x,y} U_{x,y}}{nx \cdot ny} \right)^2 \right)^k, \quad 1 \leq x \leq nx, 1 \leq y \leq ny \quad (19)$$

Note, we only use EU in the numerator. When the U array remains constant, the denominator remains constant. And the numerator will be minimized if unoccupied blocks are moved to the neighborhood of CLBs with highest $U_{x,y}$ values. So, utilizing the approach above has the effect to move unused CLBs to congested regions.

Based on the discussion above, we can draw two reasonable conclusions. First, if there are plenty of empty blocks, the routability will be good by nature. So, we turn off this optimization when *empty_rate* is greater than 25%. On the other hand, if all empty blocks are IO pads instead of CLBs, this algorithm will not work well since IO pads can only be moved along the borders. According to VPR’s layout method, a small value of *empty_rate* means there are a lot of unused IO pads but very few empty CLBs. So, we turn off this optimization when *empty_rate* is less than 0.6%. The

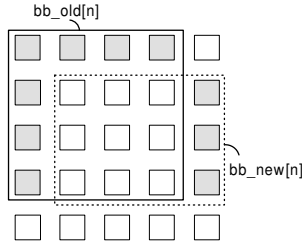


Figure 2: Incremental computation of the U array.

exponent attenuation scheme of k used in cMap is too complicated. We use a much simpler version which is shown in Table 2. The value of k is a constant for a particular circuit and its value can be either 2 or 1. The criterion decide the value of k is based on *empty_rate*, i.e., k is set to 2 for a circuit with a low *empty_rate* ($< 4\%$) and set to 1 otherwise.

4.3 Runtime Minimization

Although the timing-driven part of our algorithm does not cause much runtime overhead, the congestion-driven part actually does. So our main task is to reduce the complexity in computing the *Congestion* factor.

During the process of SA, each swap causes a set of nets to change their bounding boxes. Every time a net’s bounding box is changed, we need to update the U array. Assume n is the index of an affected net, bb_old and bb_new are its bounding boxes before and after the swap. A naive implementation needs $bb_old[n].width \cdot bb_old[n].height$ number of subtractions and $bb_new[n].width \cdot bb_new[n].height$ number of additions for each affected net. In another word, the complexity depends on the area of the bounding boxes. This requires much more computation than VPR’s linear congestion method.

To reduce the computation complexity, we use two techniques. First, not all elements inside the $bb_old[n]$ or $bb_new[n]$ should be updated. Figure 2 shows the bounding boxes of a net before and after a swap. We can see, only the shaded elements need to be updated. The overlap part does not change at all. In general, $bb_old[n]$ always intersects $bb_new[n]$. This is because a net connects at least two blocks and a swap can at most change the positions of two blocks. So we can always use this optimization to save runtime.

The other technique is to enable the congestion optimization only after the swap radius is reduced to 1.0. This technique reduces the runtime considerably in that: 1) The length of period when the swap radius equals 1.0 is only about 1/3 of the entire placement period. 2) After the swap radius reaches 1.0, the SA process is about to end. At this moment, the dimension of almost every net (the area of every net’s bounding box) has shrunk a lot. As the complexity of updating the U array is directly related to the area of a net’s bounding box, this optimization technique is able to reduce computation complexity effectively. According to our experience, factor 2 is more important than factor 1.

5. EXPERIMENTAL RESULTS

We have implemented and integrated our proposed algorithm in the framework of VPR. The experiments were carried out on a Pentium4 2.8GHz PC with 1GB memory

Table 3: Comparison of VPR, t-RPack, cMap, SMAC and ours.

algorithm	delay	channel width	runtime
VPR	1	1	1
t-RPack _p	1	0.941	NA
t-RPack _{up}	0.95	0.973	NA
cMap	0.993	0.93	5.12X
SMAC	0.88	1.22 ^a	100x ^b
ours	0.919	0.950	2.3x

^aIn terms of CLB counts.

^bIt is compared to DAOmap[5] + T-VPack.

running the CentOS Linux system. The MCNC benchmark circuits and the VPR source code (version 4.3) were downloaded from [1]. As we made a lot of modifications to VPR, those single-precision floating-point variables cannot provide enough precision any more. So we changed all “float” variables to “double”. To compare with VPR fairly, we also changed all “float” variables in VPR’s source code to “double”. Note, this is the only change that we made to the VPR’s source code.

Table 2 shows the results of VPR and our algorithm. The circuits are sorted in the ascending order of *empty_rate*. Compared with VPR, our algorithm reduces the critical path delay by 8.1% and the channel width by 5.0% at the same time. The “ N_C/N_{edges} ” column shows the fraction of edges that are favored in our timing cost computation. Let PCE_{last} be the set of potential critical edges that are picked up prior to the final temperature in the annealing phase. The “hit” column shows the percentage of the post-routing critical edges that belong to PCE_{last} . It can be seen that although we only pick up an average of 1.2% edges, the hit rate is above 50%. This explains why our approach reduces the longest path delay so effectively.

The placement runtime overhead of our algorithm is reasonable, only 2.3X as of VPR’s. What is more important is that the runtime ratio between ours and VPR’s does not increase with the increasing circuit size. It can be seen that the ratio for the smallest circuit “ex5p” is 2.03, and for the largest circuit “clma” is 2.41. This indicates that our approach has very good scalability.

Table 3 compares several algorithms which are based on VPR. There are two lines for t-RPack where t-RPack_p means it runs in the population mode, while t-RPack_{up} means it runs without population. Still, our approach outperforms all of them if we consider timing and channel width together. Also, the runtime of our algorithm is the most efficient among all algorithms whose runtime are available.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we presented timing and congestion driven placement algorithms for FPGAs. By utilizing the criticality history and disperses unoccupied CLBs/IO pads, our algorithms improves both timing and routability. Besides, as a whole, they need only 2.3X runtime as of VPR’s. Optimizing multiple metrics simultaneously is important. In our future research, we would like to investigate the relationship and trade-off among multiple metrics.

Table 2: Experiment Results: VPR vs. Ours

Circuit			Critical Path Delay			Routing Tracks			Parameters		Statistics		runtime
name	<i>empty_rate</i>	<i>N_{edges}</i>	VPR	ours	ratio	VPR	ours	ratio	<i>k</i>	λ	<i>N_C/N_{edges}</i>	hit	ratio
s298	0.258%	6951	148.2	136.9	0.924	8	8	1	2	0.3	1.5%	100%	1.76
ex1010	0.562%	16078	174.5	176.0	1.009	12	11	0.917	2	0.3	1.0%	25.0%	1.69
seq	0.794%	6193	109.5	101.6	0.928	12	12	1	2	0.3	1.5%	71.4%	2.11
spla	0.833%	13808	165.6	167.3	1.01	15	16	1.06	2	0.3	1.0%	33.3%	2.41
clma	0.957%	30462	206.5	187.2	0.907	15	13	0.867	2	0.3	0.7%	40.0%	2.41
pdc	1.060%	17193	194.8	206.7	1.061	18	18	1	2	0.3	0.9%	40.0%	2.40
frisc	1.222%	12772	132.5	133.8	1.01	14	13	0.929	2	0.3	1.0%	100%	2.33
diffeq	1.578%	5296	63.72	60.45	0.949	9	8	0.889	2	0.3	1.6%	100%	2.11
s38584.1	1.738%	20840	112.3	82.95	0.739	9	8	0.899	2	0.3	0.8%	100%	2.25
ex5p	2.296%	4002	77.21	79.75	1.033	15	14	0.933	2	0.5	1.8%	0%	2.03
s38417	2.362%	21344	103.3	83.00	0.803	9	9	1	2	0.5	0.8%	63.6%	3.14
apex4	2.623%	4479	103.1	75.43	0.732	14	14	1	2	0.5	1.7%	14.3%	2.56
apex2	2.996%	6692	100.8	94.60	0.939	12	12	1	2	0.5	1.3%	100%	2.76
elliptic	3.144%	12634	136.6	126.0	0.922	13	11	0.846	2	0.5	1.0%	0%	2.84
misex3	3.252%	4968	82.57	88.58	1.073	12	12	1	2	0.5	1.5%	28.6%	2.66
tseng	3.857%	3760	57.96	53.24	0.918	8	7	0.875	2	0.5	1.7%	0%	2.31
alu4	4.875%	5408	90.58	94.76	1.046	11	10	0.909	1	0.5	1.3%	25.0%	3.06
bigkey	41.46%	6313	72.56	55.27	0.762	7	7	1	1	0.5	1.0%	66.7%	1.63
dsip	53.02%	5645	74.37	51.69	0.695	8	7	0.875	1	0.5	1.1%	100%	1.54
des	59.91%	6110	87.21	80.79	0.926	8	8	1	1	0.5	1.0%	0%	1.72
Ave					0.919			0.950			1.2%	50.4%	2.29

7. REFERENCES

- [1] The FPGA place-and-route challenge. <http://www.eecg.toronto.edu/~vaughn/challenge/challenge.html>.
- [2] V. Betz and J. Rose. VPR: A new packing, placement and routing tool for FPGA research. In *Proceedings of the International Workshop on Field-Programmable Logic and Applications*, pages 213–222, 1997.
- [3] V. Betz, J. Rose, and A. Marquardt. *Architecture and CAD for Deep-Submicron FPGAs*. Kluwer Academic Publishers, 1999.
- [4] E. Bozorgzadeh, S. Ogrenic-Memik, X. Yang, and M. Sarrafzadeh. Routability-driven packing: Metrics and algorithms for cluster-based FPGAs. *Journal of Circuits Systems and Computers*, 13(1):77–100, 2004.
- [5] D. Chen and J. Cong. Daomap: a depth-optimal area optimization mapping algorithm for FPGA designs. In *Proceedings of the International conference on Computer-aided design*, pages 752–759, 2004.
- [6] C. Cheng. RISA: Accurate and efficient placement routability modeling. In *Proceedings of the International Conference on Computer-Aided Design*, pages 690–695, 1994.
- [7] A. DeHon. Balancing interconnect and computation in a reconfigurable computing array (or, why you don’t really want 100% lut utilization). In *Proceedings of the International Symposium on Field Programmable Gate Arrays*, pages 69–78, 1999.
- [8] A. E. Dunlop, V. D. Agrawal, D. N. Deutsch, M. F. Jukl, P. Kozak, and M. Wiesel. Chip layout optimization using critical path weighting. In *Proceedings of the conference on Design automation*, pages 133–136, 1984.
- [9] T. Hamada, C.-K. Cheng, and P. M. Chau. Prime: a timing-driven placement tool using a piecewise linear resistive network approach. In *Proceedings of the international conference on Design automation*, pages 531–536, 1993.
- [10] M. Jackson and E. S. Kuh. Performance-driven placement of cell based ic’s. In *Proceedings of the conference on Design automation*, pages 370–375, 1989.
- [11] T. Kong. A novel net weighting algorithm for timing-driven placement. In *Proceedings of the International Conference on Computer-Aided Design*, pages 172–176, 2002.
- [12] J. Y. Lin, D. Chen, and J. Cong. Optimal simultaneous mapping and clustering for fpga delay optimization. In *Proceedings of the conference on design automation*, pages 472–477, 2006.
- [13] A. Marquardt, V. Betz, and J. Rose. Timing-driven placement for FPGAs. In *Proceedings of the International Symposium on Field Programmable Gate Arrays*, pages 203–213, 2000.
- [14] A. Srinivasan, K. Chaudhary, and E. S. Kuh. Ritual: A performance driven placement algorithm for small cell ic’s. In *International Conference on Computer-Aided Design*, pages 45–51, 1991.
- [15] R. Tessier. Fast placement approaches for FPGAs. *ACM Transactions on Design Automation of Electronic Systems*, 7(2):284–305, April 2002.
- [16] Q. Wang, J. Lillis, and S. Sanyal. An LP-based methodology for improved timing-driven placement. In *Proceedings of the Conference on Asia South Pacific Design Automation*, pages 1139–1147, 2005.
- [17] M. G. Wrighton and A. M. DeHon. Hardware-assisted simulated annealing with application for fast fpga placement. In *Proceedings of the international symposium on Field programmable gate arrays*, pages 33–42, 2003.
- [18] Y. Zhuo, H. Li, and S. P. Mohanty. A congestion driven placement algorithm for FPGA synthesis. In *Proceedings of the International Conference on Field Programmable Logic and Applications*, 2006.