On the Optimal Sub-routing Structures of 2-D FPGA Greedy Routing Architectures *

Jiaofeng Pan, Yu-Liang Wu, and C. K. Wong
Department of Computer Science and Engineering
The Chinese University of Hong Kong
Shatin, NT, Hong Kong
Tel: 26098411
Fax: 26035024
Email: {jfpan, ylw, wongck}@cse.cuhk.edu.hk

Abstract — For the FPGA Greedy Routing Architectures (GRAs), the optimal mapping problem of the entire chip can be decomposed into a sequence of three kinds of optimal inside predetermined 4-way FPGA mapping problems, where m could be 1, 2, or 3. In this paper, we formulate the graph models of such sub-routing problems and investigate their minimum structures. The results give the lower bounds of routing resources in achieving all such kinds of GRAs and the theoretic models developed could be useful to studies on other FPGA routing problems as well.

1. Introduction

An FPGA is an array of pre-fabricated functional blocks and wire segments with user-programmability of logic and routing resources. The routing architecture discussed here is also described in [2,4,7,10] and a simplified model is shown in Fig.1. In the model, each logic cell (L) can be configured to be a Look-Up-Table (LUT), flip-flop, etc.[10]. We assume there are W prefabricated wire segments running between the cells in both vertical and horizontal channels with each wire length spanning an L block. Each track within a channel is assigned a track id as shown in Fig.1.2. Programmable routing switches are grouped inside the connection (C) boxes and switch (S) boxes. The C boxes contain switches that can connect logic block pins to wire segments. The S boxes contain switches that allow one wire segment to be connected to another. We assume that each logic pin has a switch to connect to any wire inside the C box containing the logic pin. Therefore, the routing property of the architectures investigated is dependent on the number of switches per S box (SpSB) and its structure only.

Routing is crucial in the FPGA design automation process. Conventionally, a global router decides the connection path of each net and the detailed router decides the wire segments and switches that are actually assigned to implement such connection paths. The conventional approach to FPGA routing is a 2-setp global/detailed routing scheme[5,6]. Assume that the channel density of an FPGA chip resulting from a global routing is $W_g$, i.e., the maximum number of global routes which run in parallel in any channel is $W_g$, and $W_d$ is the number of tracks needed to complete the detailed routing. In this research, we will focus on the mappability of architectures which can guarantee a perfect mapping, a routing mapping with $W_d=W_g$. The class of architectures, which can achieve polynomial routing and perfect mapping, are Greedy Routing Architectures (GRAs) [1].

In [8], the concept of GRA was first proposed. A GRA is an architecture with the property that a locally optimum routing of a single switch box can be extended to a globally optimal routing of the whole chip. Since the locally optimal routing needs to be extended to a global solution, a GRA also implies a routing order on its switch boxes. The basic idea is as follows: given the global routing of the entire chip, we start the detailed routing mapping around a prespecified S box. Then a C box with detailed routing just mapped serves as the predetermined side of a neighboring (and unmapped) S box and the mapping process is repeated (or extended) again until the entire chip is mapped. Depending on the linear extending order (e.g. either spiral [8], or snake-like in [7]), the optimal mapping problem of the entire chip can be decomposed into a sequence of three
kinds of optimal m-side predetermined 4-way FPGA mapping problems, where m could be 1, 2, or 3. For example, in a snake-like routing order [7], the majority of the local routing cases would be the two-side predetermined 4-way mapping, while on the chip boundary, the other two kinds would also be used.

An m-side predetermined 4-way FPGA mapping problem is defined to be a routing mapping between a global routing over the 4 C boxes around an S box, with detailed routing of m sides predetermined, to a feasible global routing over the 4 C boxes around an S box, with predetermined 4-way mapping, while on the chip the local routing cases would be the two-side example, in a snake-like routing order [7], the majority of mapping problems, where m could be 1, 2, or 3. For kinds of optimal m-side predetermined 4-way FPGA

If we regard a track as a vertex, and a switch as an edge in a four-way switch box, then we have a corresponding graph, called the S box graph. For a vertex v in graph G, denote the degree of v by d(v), which is the number of edges adjacent to v. N(X) denotes the neighborhood of vertex set X in a graph.

3. S box graph and its properties

3.1. S box graph and its decomposition

There are 6 types of side-to-side switch connection relations in an S box, as shown in Fig.4.

Consider the graph G=(V,E), as shown in Fig.5, in which each track of the S box corresponds to a vertex of G and each switch inside the S box corresponds to an edge of G. Let V_i, V_2, V_3 and V_4 denote the set of vertices (tracks) on the top side, on the left side, on the bottom side and on the right side, respectively. Let E_1, E_2, E_3, E_4, E_5 and E_6 denote the set of edges (switches) between left side and bottom side, bottom side and right side, right side and top side, top side and left side, left side and right side, bottom side and top side, respectively. Then G can be uniquely decomposed into the following 6 subgraphs G_i (i =1,2,3,4,5,6) each of which is a bipartite graph: G_1 = (V_2,V_3; E_1), G_2 = (V_3,V_4; E_2), G_3 = (V_1,V_4; E_3), G_4 = (V_1,V_2; E_4), G_5 = (V_2,V_3; E_5) and G_6 = (V_1,V_3; E_6). Each of them uniquely represents one of the 6 connection relations of an S box.

3.2. On the properties of the S box graph

Consider the case of |V_i|=W, i=1,2,3,4. Let G*= G_1[G_2 as shown in Fig.5. Then G* = (V,V_1[V_2; E_1]∪E_4). Note that there is no edge between V_2 and V_3 in the graph G*, as all the edges between V_2 and V_3 are in E_2 but E_4 ∩ (E_1 ∪ E_2) = φ. For convenience, denote G* = (X*, Y*, E*). Thus, X*= V_2, Y* = V_1[V_2; E_1] and E*= E_1∪E_4. To consider the general case, suppose |Y*=|nW.
Theorem 1 For any subset Z\subseteq Y* with |Z| \leq W, if the induced graph G*,=G*[X*[\cup Z] always has a matching to cover Z, then \(d(x)\geq (n-1)W+1\), \(\forall x\in X*\) in G*; \(d(y)\geq 1\), \(\forall y\in Y*\) in G*.  
Hence, \(|Y*|\leq W\). Therefore, \(|Y*|\leq W\).

Proof. We know that any subgraph of a bipartite graph is still a bipartite graph. G*, is an induced graph of G*; so G*, is a bipartite graph.

Suppose \(\exists x_i \in X*\) in G*, such that \(d(x_i)\leq (n-1)W\), i.e., there exists at least W vertices of Y* which are not adjacent to \(x_i\). Without loss of generality, assume that \(y_1, y_2, ..., y_w\) are not adjacent to \(x_i\). Let \(Z=\{y_1, y_2, ..., y_w\}\). Then, \(x_i \notin N(Z)\). Hence, \(|N(Z)|\leq W-1 < W=|Z|\). According to the Hall Theorem in [3], there is not a matching to cover Z in the induced graph, a contradiction. So the theorem holds.

Corollary 1. \(|E*| \geq W[(n-1)W+1]\).

Proof. From Theorem 1, we have \(\sum d(x)\geq W[(n-1)W+1]\). \(x\in X*\) since \(d(x_i)\geq (n-1)W+1\), \(\forall x_i \in X*\) and \(|X*|=W\.

Since the following equations hold for any bipartite graph G=(X, Y; E):

\[
\sum d(x) = \sum d(y), \quad x \in X, \quad y \in Y \quad (1)
\]

\[
\sum d(x) + \sum d(y) = 2|E| \quad (2)
\]

we have,

\[|E|=\sum d(x), \quad x \in X.\]

For the bipartite graph G*=(X*, Y*; E*),

\[|E*|=\sum d(x), \quad x \in X*.\]

Therefore, \(|E*|\geq W[(n-1)W+1]\).

Remark. For a 4-way FPGA routing, Theorem 1 is a necessary condition that the bipartite graph including n predetermined sides (n=1,2,3) and one nonpredetermined side must satisfy. In fact, the DAD model satisfies the condition Theorem 1 requires.

Consider \(G_0=(V_0, V_3; E_0)\), where \(|V_0|=|V_3|=W\). Let \(R\) be a subset of \(V_0\) and \(|R|<W, R_1=V\setminus R\). Let \(T\) be a subset of \(V_0\) with \(|T|=|R|\).

Theorem 2. For any subset \(R\) of \(V_1\) with \(|R|<W, R_1=V\setminus R\) and any subset \(T\) of \(V_1\) with \(|T|=|R|\), the induced graph \(G[R_1\cup T]\) has a perfect matching if and only if \(d(v)=W\) for all \(v\in V_1\cup N_3\). Here \(G_0=(V_1, V_3; E_0)\) with \(|V_0|=|V_3|=W\).

Proof. (\(\Leftarrow\)) Obviously, the induced graph \(G[R_1\cup T]\) is a bipartite graph.

If \(d(v)=W\) for all \(v\in V_1\cup V_3\), we can prove that \(G[R_1\cup T]\) is a \((W-|T|-|R|)\)-regular bipartite graph, i.e., \(d(u)=W-|R|\) for all \(u\in R_1\cup T\). In fact, we can construct the induced graph in this way:

(1) Remove all vertices of \(R\) from \(G_0\). Then \(d(x)=W\) for \(x\in R_1\) and \(d(y)=W-|R|, \forall y\in V_3\) in the remaining graph.

(2) Remove all vertices of \(V_3\cup T\) from the graph obtained in step one. Then \(d(y)=W-|R|, \forall y\in T\) and \(d(x)=W-|R|, \forall x\in R_1\).

For any non-trivial regular bipartite graph, there must be a perfect matching (corollary of the Hall Theorem). Thus, the necessary condition holds.

(\(\Rightarrow\)) Suppose \(d(v)=W<W\). Without loss of generality, assume \(v\in V_0\), then \(N(v)\subseteq V_1\) and \(|N(v)|<W\). We construct the induced graph in this way:

(1) Take \(R=N(v)\). Then \(|R|<W, R_1=V\setminus N(v), |R_1|\geq 1\). Then for any vertex \(u\in R_1, v\notin N(u)\).

(2) Take \(T\subseteq V_3\), \(|T|=|R_1|, v\subseteq T\).

Now consider the induced graph \(G[R_1\cup T]\). We have \(v\notin N(R_1)\) since \(v
notin N(u)\) for any vertex \(u\in R_1\). Therefore, \(|N(R_1)|\leq |T|-1 \leq |V_3|\). According to the Hall Theorem, there is not a matching to cover \(R_1\). This violates the statement that for any subset \(R\) of \(V_1\) with \(|R|<W, R_1=V\setminus R\) and any subset \(T\) of \(V_3\) with \(|T|=|R_1|\), the induced graph \(G[R_1\cup T]\) has a perfect matching.

Therefore, \(d(v)=W\) for all \(v\in V_1\). 

Corollary 2. If a bipartite graph \(G=(X, Y; E)\) satisfies the condition of Theorem 2, then \(|E|=W^2\), where \(|X|=|Y|=W\).

Proof. Recall that \(|E|=\sum d(x), \quad x \in X\) for a bipartite graph \(G=(X, Y; E)\).

Hence, \(|E|=\sum d(x)=W^2, \quad x \in X\).

Remark. In fact, it is the case that the induced graph results from two predetermined sides in a 4-way S box.

4. Side-to-side graphs and their properties

Let \(H_1, H_2\) and \(H_3\) be the corresponding graph representation of the DD model, AA model, DAD model respectively as shown in Fig.6. We call them the side-to-side graphs.

For the DADA model, it is obvious that \(H_1\) and \(H_2\) correspond to \(G_6\) and \(G_3\) respectively. \(H_3\) corresponds to \(G_4\cup G_5\) and \(G_7\cup G_3\). Also it is obvious that \(H_1, H_2\) and \(H_3\) are bipartite graphs.

Now look at \(H_3\) i.e., the DD graph. Both of the top and bottom sides are predetermined already. The input to the DD model is a vertical net from either of the two sides, and the total number of inputs must be less than or equal to W. Let \(W_t\) and \(W_b\) denote the number of inputs of top side and bottom side, respectively. Then, the following inequality must be satisfied for any feasible routing: \(W_t + W_b \leq W\). To complete the detailed routes from both sides, there must be
two mutually exclusive perfect matchings with size $W_t$ and $W_b$ respectively for any subset with order $W_t$ of top side and any subset with order $W_b$ of bottom side which satisfy $W_t + W_b \leq W$. In fact, it is the restatement of the condition of Theorem 2. Therefore, $H_t$ is a complete graph. Furthermore, $E[H_t] = \mid W^2$. Indeed, we have proved the following

**Lemma 1.** $H_t$ is a $K_{W_1 W_2}$ complete bipartite graph.

Now consider $H_s$ (X,Y;E) which corresponds to the DAD model. X is the vertex set of the nonpredetermined side and $|X|=W$, and Y is the vertex set of the two predetermined sides and $|Y|=2W$. The input to this model is the kind of net “tracks” and their connection must have a complete tracks, the bipartite subgraph including the unused vertices (tracks) and their connection must have a complete matching. We can formulate the problem as follows: $H_t$ = (L,R;E), where $|L|=|R|=W$. For any $S \subseteq L$ and $Z \subseteq R$, and $|S|+|Z|\leq W$, if $G((X,Y) \cap (S \cup Z))$ always has a matching of size $\min\{|L|S|, |R|Z|\}$, then $d(v) \geq \frac{1}{2} W$ for all $v \in V(H_t)$.

Suppose $\exists v \in L$ such that $d(v) < \frac{1}{2} W$, then $N_{G}(v) \subseteq R$ and $|N_{G}(v)| \geq \frac{1}{2} W$. Let $Z=N_{G}(v)$, $S \subseteq L \{v\}$ and $|S|= \frac{1}{2} W$. Then we have, (1) $|LS| = \frac{1}{2} W$; (2) $v \in (XS)$ and $d(v)=0$ in $G((L\cup R) \cap (S \cup Z))$; (3) $|RZ| = |R \cap N_{G}(v)| \geq \frac{1}{2} W$. Hence, from (1) and (3), we have $\min\{|LS|, |R \cap N_{G}(v)|\} = \frac{1}{2} W$. But from (2), we have a maximum matching of $|G((L\cup R) \cap (S \cup Z))| \leq \frac{1}{2} W-1$, a contradiction. So we have the following:

**Lemma 3.** $H_s$ is a bipartite graph with $d(v) \geq \frac{1}{2} W$ for all $v \in V(H_s)$.

5. **Conversion of the two-side predetermined 4-way FPGA routing**

Consider the two-side predetermined 4-way routing at the box level. Without loss of generality, we specify the top side and bottom side are the predetermined sides. So we only need to study the DADA model in our studies of the two-side predetermined 4-way FPGA routing problem.

Now consider the conversion of the DADA model. Obviously, the side-to-side models DD, AA and DAD are components of the DADA model. Correspondingly, the graphs $H_s$, $H_t$ and $H_h$ are components of the S box graph.

Consider the Constraint model as shown in Fig.7. Let the two upper predetermined sides of the Constraint model represent the top side of the DADA model, and the two lower predetermined sides represent the bottom side of the DADA model. Let the two A sides represent the nonpredetermined sides (corresponding to the left and right sides of the DADA model). If there is no confusion, we will also use X, Y, L, R to denote the sides in the order of the leftmost side to the rightmost side. Then, any feasible routing instance in the combination of the DD model and the Constraint model is also feasible in the DADA model.
To prove this property, let us consider the feasible routing instances in details.

There are 11 types of nets entering/exiting the 4-way switch box: two right-bent nets, two left-bent nets, a vertical net, a horizontal net, four 3-way nets and a 4-way net. We regard the 11 types of nets as the input to the DADA model. Any global routing instance is the combination of some of the 11 types of nets. Note that the routing (input) always starts from the predetermined sides. The type of net entering/exiting the DD model is a vertical net, and which entering/exiting the Constraint model is a horizontal net. Any feasible routing instance in the DD model is a set of vertical nets, and which in the Constraint model is a set of horizontal nets. So it is obvious that any feasible routing case in the DD model and the Constraint Model is routable in the DADA model. Therefore, we have the following

**Lemma 4.** Any feasible routing instance in the combination of the DD model and the Constraint model is also feasible in the DADA model.

### 6. The optimal structure of two-side predetermined 2-D FPGA greedy routing

By Lemma 4, in order to explore the structure of the DADA model, we only need to explore the DD model and the Constraint model, respectively. In order to explore the structure of the DADA graph, we only need to explore the structure of the Constraint graph and the DD graph.

From Section 4, we know that the DD graph is $H_i$, and $H_i$ corresponds to $G_6$, $H_2$ corresponds to $G_5$ and $H_3$ corresponds to $G_4$ and $G_3$.

Consider the Constraint graph $CG=(X,L,R,Y;E)$ as shown in Fig.7. Corresponding to the DADA graph, $X$ and $Y$ are the same vertex set, i.e., the vertex set of top side and bottom side. In fact, $CG$ is a four-partite graph and $CG=CG_1\cup CG_2\cup CG_3$, where $CG_1=(X,L;E_1)$, $CG_2=(L,R;E_2)$, $CG_3=(R,Y;E_3)$. Corresponding to the DADA graph, $CG_1=G_1\cup G_2$, $CG_2=G_3$, $CG_3=G_4\cup G_5$. Hence, $CG_1$ and $CG_3$ are $H_1$, $CG_2$ is $H_2$. So the DADA graph is the union of $H_1$ and $CG_1\cup CG_3 \cup CG_2$. From Lemmas 1, 2, and 3, we have the following structure theorem:

**Theorem 3.** (Structure Theorem) The DADA graph $G$ has the following structure:

1. In the subgraph $G_1 \cup G_2$ and $G_3 \cup G_4$, $d(v)\geq W+1$, for any $v \in V_2 \cup V_4$, $d(u)\geq 1$, for any $u \in V_1 \cup V_3$, $|E(G_1 \cup G_2)| = |E(G_3 \cup G_4)| \geq WW+1$.
2. In the subgraph $G_5$, $d(v)=W$, for any $v \in V_1 \cup V_3$, $E(G_5)=W^2$.
3. In the subgraph $G_3$, $d(v)\geq \frac{1}{2} W$, for any $v \in V_2 \cup V_4$, $|E(G_3)| \geq \frac{1}{2} W^2$.

(4) $|E(G)| \geq \frac{1}{2} W^2 + 2W$.

Since the above structure does not depend on any specific Greedy Routing architecture, $\frac{1}{2} W^2 + 2W$ is the lower bound for any GRA used in the two-side predetermined 4-way FPGA routing problem.

The lower bound can be reached. In [1], a specific architecture as shown in Fig.8 is proposed. In Fig.8 and also in Fig.9 and Fig.10, for clarity, we use three simplified connection pattern notations to represent side-to-side connections. $\rightarrow \rightarrow$ represents pattern I, where each track on the left side connects to a track with the same index on the right side. $\rightarrow \longrightarrow$ represents pattern II, where each track $i$ on the left side connects to a track with the same index on the right side. $\rightarrow \longrightarrow \longrightarrow$ represents pattern III, where the connections between the two sides are complete.

The switch box in Fig.8 has an identity mapping between the top side and the right side, the top side and the left side, and a complete matching between the top side and the bottom side, the right side and the bottom side, the left side and the bottom side. Each left track is connected to $R_{i+1}$, $R_{i+2}$, $\ldots$, $R_{i+w/2}$ (subscripts are always mod W). The GRA architecture has exactly $\frac{1}{2} W^2 + 2W$ edges (switches) for any 4-way box. It has been proved that the architecture can solve the two-side predetermined 4-way FPGA problem. Finally, we can conclude that the optimal structure of GRA used to solve the two-side predetermined 4-way FPGA problem is the 4-way box as proposed in [1]. Indeed, we have proved our main result as follows:

**Theorem 4.** The 4-way switch box described in Fig.8 is an optimal structure of GRA used to solve the two-side predetermined 4-way FPGA routing problem.

---

![Fig.8. The 4-way greedy routing S Box with the top and bottom side predetermined.](image-url)
7. Optimal structures for one-side and three-side predetermined 4-way FPGA routings

Due to the page limit, here we briefly discuss the problems and give our main results, with detailed proof skipped (see[9]).

**Theorem 5.** The S box as shown in Fig.9 is the optimal structure of GRA to solve the one-side predetermined 4-way FPGA greedy routing problem, and the S box as shown in Fig.10 is the optimal structure of GRA to solve the three-side predetermined 4-way FPGA greedy routing problem.

The structure of the new switch box shown in Fig.9 is as follows: each track of the top side connects to the track with the same subscript as it in the right side, left side and bottom side, respectively; each left track L_i connects to R_{i+1}, R_{i+2}, ..., R_{i+W/2} (subscripts are always mod W); each bottom track B_i connects to R_{i+1}, R_{i+2}, ..., R_{i+W/2} (subscripts are always mod W); each track of the top side connects to the track with the same subscript on the right side, and connects to each track of the left and bottom sides; each left track connects to each track of the other three sides; each track of bottom side connects to each track of the other three sides. The number of switches in the box is \( \frac{1}{2} W^2 + 3W \).

![Fig 9 The 4-way greedy routing S Box with top side predetermined.](image1)

![Fig 10 The 4-way greedy routing S Box with top, left and bottom sides predetermined.](image2)

Our new S box shown in Fig.9 can detail route any 4-way global routing with one-side. It is easy to see that there are 5 types of nets entering/exiting the model relevant to the predetermined side: a vertical net, a right-bent net, a left-bent net, a 3-way net, and a 4-way net. Any routing instance relevant to the predetermined side is from the predetermined side, say, top side, to the other three nonpredetermined sides. It is easy to check that any combination of the 5 types of nets with size less than or equal to W can be routable. On the other hand, any two nonpredetermined sides and the predetermined side together form a 3-way box described in [1]. Since the 3-way box is the one which has a minimum number of switches while maintaining a perfect mapping property [1], so there is always a complete matching between the unused tracks of any two nonpredetermined sides of the new S box. Therefore, our new S box maintains the perfect mapping property. So any routing instance only relevant to the three nonpredetermined sides is routable. In[9] we show that the S box as shown in Fig.8 is the optimal structure of GRA to solve the one-side predetermined 4-way routing FPGA problem.

The structure of the new switch box shown in Fig.10 is as follows: each track of the top side connects to the track with the same subscript on the right side, and connects to each track of the left and bottom sides; each left track connects to each track of the other three sides; each track of bottom side connects to each track of the other three sides. The number of switches in the box is \( 5W^2 + W \). As in the proof of the optimality of the one-side predetermined 4-way S box, we can prove that the S box as shown in Fig.10 is the optimal structure of GRA to solve a three-side predetermined 4-way FPGA routing problem.

8. Conclusion

All the 2-D GRAs possess certain prespecified extendable linear order on mappings around switch boxes of the chip, and the optimal mapping problem of the entire chip can be decomposed into a sequence of three kinds of optimal m-side predetermined 4-way FPGA mapping problems, where m could be 1, 2, or 3. In this paper, we formulate the graph models of such sub-routing problems and investigate their minimum structures. We showed that the minimum structures of these problems require switch populations of \( \frac{1}{2} W^2 + 3W \), \( \frac{1}{2} W^2 + 2W \), and \( 5W^2 + W \), with m equal to 1, 2, or 3 respectively. It should be pointed out that a shorter alternative proof of these results may be possible. The details will be reported in [9] when completed.

**References**


