Analytical Framework for Switch Block Design

Guy G. Lemieux and David M. Lewis

Edward S. Rogers Sr. Dept. of Electrical and Computer Engineering
University of Toronto, Toronto, Ontario, Canada
{Lemieux, Lewis}@eecg.utoronto.ca

Abstract. One popular FPGA interconnection network is based on the island-style model, where rows and columns of logic blocks are separated by channels containing routing wires. Switch blocks are placed at the intersections of the horizontal and vertical channels to allow the wires to be connected together. Previous switch block design has focused on the analysis of individual switch blocks or the use of ad hoc design with experimental evaluation. This paper presents an analytical framework which considers the design of a continuous fabric of switch blocks containing wire segments of any length. The framework is used to design new switch blocks which are experimentally shown to be as effective as the best ones known to date. With this framework, we hope to inspire new ways of looking at switch block design.

1 Introduction

Over the past several years, a number of different switch block designs have been proposed such as those shown in Figure 1. FPGAs such as the Xilinx XC4000-series [1] use a switch block style known as *disjoint*. Some alternatives to this style, known as *universal* [2] and *Wilton* [3], require fewer routing tracks and use less transistor area with interconnect of single-length wires. However, with longer wire segments they use more switches per track and often require more transistor area overall [4]. The *Imran* block [5] addresses this overhead by modifying the *Wilton* pattern to use the same number of switches as the *disjoint* pattern.

These switch blocks are designed using different methodologies. The *universal* switch block is analytically designed to be independently routable for all two-point nets. Recently, the *hyperuniversal* switch block [6] extends this for multi-point nets. These blocks rely on reordering nets at every switch block, so their local optimality does not extend to the entire routing fabric. In comparison, the *Wilton* and *Imran* switch blocks are examples of ad hoc design with experimental validation. The *Wilton* block changes the track number assigned to a net as it turns. This way, two different global routes may reach two different tracks at the same destination channel. This forms two disjoint paths, a feature we call the *diversity* of a network. The *Wilton* and *Imran* designs introduce the notion that a switch block must consider its role as part of a larger switching fabric.

The above methods have produced switch blocks that perform well, but there is no formal method to design a switch block while considering the overall routing fabric. In pursuit of this goal, this paper introduces an analytical framework which considers

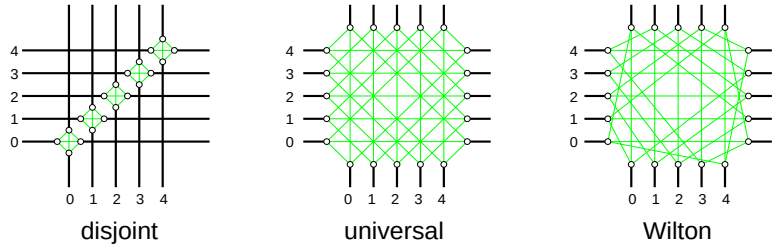

Fig. 1. Different switch block styles.

both long wire segments and the interaction of many switch blocks connected together. This framework includes a restricted switch block model which allows us to analyse the diversity of the network. The framework is used to design an ad hoc switch block named *shifty* and two analytic ones named *diverse* and *diverse-clique*. These new switch blocks are very diverse, and routing experiments show they are as effective as the others.

2 Design Framework

This section describes the switch block framework being composed of a switch block model, permutation mapping functions, and simplifying assumptions and properties.

2.1 Switch Block Model

The traditional model of a switch block draws a large box around the intersection of a horizontal and vertical routing channel. Within the box, switches connect a wire on one side to any wires on the other three sides. Long wire segments pass straight across the switch block, but some track shifting is necessary to implement fixed length wires with one layout tile. Figure 2a) presents this model in a new way by partitioning the switch block into three subblocks: *endpoint* (f_e), *midpoint* (f_m), and *midpoint-endpoint* (f_{me}) subblocks. The endpoint (midpoint) subblock is the region where the ends (midpoints) of wire segments connect to the ends (midpoints) of other wire segments. The f_{me} subblock connects the middle regions of some wires to the ends of others. A switch placed between two sides always falls into one of these subblocks.

The traditional model in Figure 2a) is too general for simple diversity analysis, so we propose restricting the permissible switch locations. One restriction is to prohibit f_{me} switches; this was done in the *Imran* block [5]. We propose to further constrain the f_m switch locations to lie within smaller subblocks called $f_{m,i}$, as shown in Figure 2b) for length-four wires. This *track group model* is a key component to the framework.

The track group model partitions wires into *track groups* according to their wire length and starting points. The midpoint subblocks are labeled $f_{m,i}$, where i is a position between 1 and $L-1$ along a wire of length L. This model is somewhat restrictive, but it can still represent many switch blocks, *e.g.*, *Imran*, and we will show that it performs well. As well, early experiments we conducted without the $f_{m,i}$ subblock restric-

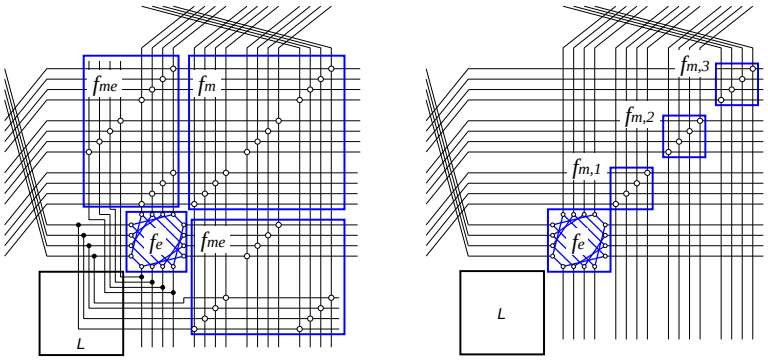

Fig. 2. Switch block models with subblocks, a) traditional and b) track group model.

tions did not produce better results. However, the $f_{m,i}$ subblocks explicitly force track groups to be in separate routing domains so each group can be treated independently.

2.2 Permutation Mapping Functions

Previous work suggests only a small number of switches need to be placed within a switch block. Early work [7] defined switch block flexibility, F_s, as the number of other wires connecting to each wire in this block. They found that $F_s = 3$ is the lowest that is routable with single-length wire segments. Other work [4, 5] has used $F_s = 3$ at wire endpoints and $F_s = 1$ at wire midpoints when long wire segments are used. As well, our experience with $F_s < 3$ is that a few more tracks but less transistor area is needed [8]. This suggests $6W$ and W are reasonable upper bounds for the number of switches in endpoint and midpoint subblocks, respectively.

Given these upper bounds, switch locations can be represented by a *permutation mapping function* between each pair of sides. The different mapping functions and their implied forward direction are shown in Figure 3. In this figure, $f_{e,i}(t)$, or simply $f_{e,i}$, represents an endpoint turn of type i. A switch connects the wire originating at track t to track $f_{e,i}(t)$ on the destination side. Turns in the reverse direction to those indicated are represented as $f_{e,i}^{-1}$ such that $f^{-1}(f(t)) = t$.

Similarly, $f_{m,i}$ is a mapping function for a midpoint turn at position i along the length of a wire, with the most South/West endpoint being the origin at position $i = 0$. Figure 3b) illustrates the different midpoint subblocks in a fabric of 2×4 logic blocks (L) for a single track group. The other three track groups are independent, but they would be similar and have staggered starting locations. There are no connections between the track groups.

Examples of mapping functions for various switch blocks are shown in Table 1. Each of these functions are modulo W, where W is the track group width. Also, note that it is common for connections straight across a switch block (E–W or N–S) to stay in the same track, so it is usually assumed that $f_{e,5} = f_{e,6} = t$.

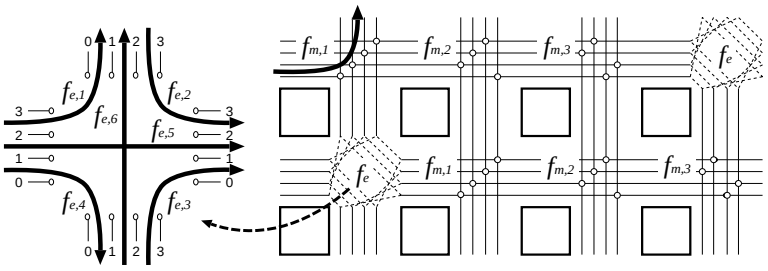

Fig. 3. Mapping functions for a) endpoint and b) midpoint subblock turns.

2.3 Additional Assumptions

In addition to the explicit assumptions above, there are a few implicit ones being made as well. It is assumed that the subblocks are square with W tracks on each side and there is a one-to-one correspondence between the originating track and the destination track. Since $f^{-1}(f(t)) = t$, it is also presumed that each switch is bidirectional. Additionally, we assume a track group contains only one wire length and switch type.

2.4 Commutative Switch Blocks

The mapping functions of the *universal* and *Imran* switch blocks involve *twists* where the function is of the form $f(t) = W - t + c$. Unfortunately, these functions are difficult to analyse because the twist is not commutative. Using commutative functions simplifies the model because the order in which turns are made becomes unimportant. Paths with an arbitrary number or sequence of turns can be reduced to a canonical permutation which uniquely determines the destination track. Later in Section 3.2, this will allow us to significantly reduce the search space. We define a switch block to be *commutative* if all of its mapping functions are commutative.

Consider the example shown in Figure 4, where two paths are compared in two different architectures. The left architecture uses commutative switch blocks, but the right one does not. The destination track of the upper path is $f_{e,2}(f_{e,4}(f_{e,3}(f_{e,1}(t))))$, while the lower path is $f_{e,3}(f_{e,1}(f_{e,2}(f_{e,4}(t))))$. In a commutative architecture, both paths can be rewritten as $f_{e,1}(f_{e,2}(f_{e,3}(f_{e,4}(t))))$. These necessarily reach the same track. In a non-commutative architecture, the operations cannot be reordered and the paths may reach different tracks. This example suggests that commutative architectures are less diverse. However, results will demonstrate that commutative switch blocks are very diverse and as routable as the non-commutative *Imran* block.

3 Framework Applications

To illustrate the use of the new framework, two approaches will be used to determine a set of permutation mapping functions. The first, named *shifty*, is an ad hoc commutative switch block. The second creates two switch blocks, named *diverse* and

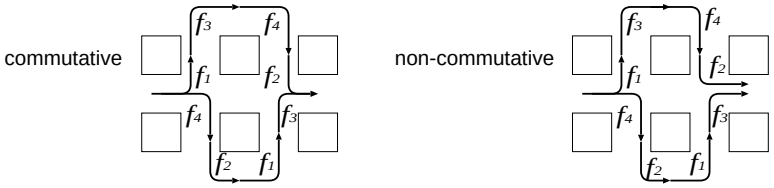

Fig. 4. Turn order is not important in commutative switch blocks.

diverse-clique, by optimizing diversity. Both of these approaches assume length-four interconnect wires. As well, they will assume that two separate layout tiles can be used in a checkered fashion to further increase diversity.

3.1 Application: *shifty* and *universal-TG* Designs

The first application of the new framework is the design of a commutative switch block similar to *Imran* but without the non-commutative twists. The following mapping functions describe the new switch block: $f_{e,1} = t-1$, $f_{e,2} = t-3$, $f_{e,3} = t-2$, $f_{e,4} = t-4$, and $f_{m,i} = t \pmod W$. This block is named *shifty* because each turn involves a shift from one track number to another by a constant amount. The constant values are chosen to be small because the arithmetic is always done modulo W. This avoids $f_{e,1}$ from being equivalent to $f_{e,4}$, for example, except with certain small W values.

Other switch blocks can also be adopted within this framework. For example, the *disjoint* and *Imran* switch blocks naturally conform to the track group model already. As well, suppose the *universal* pattern is applied only at endpoint subblocks and the identity mapping $f_{m,i} = t$ is used at midpoint subblocks. This new pattern, *universal-TG*, is similar to the original in that *each subblock* can connect any set of two-point nets that obey basic bandwidth constraints. It also requires less transistor area with long wire segments in the same way that *Imran* improves *Wilton* by reducing the number of switches per track.

To create additional diversity, it is possible to use two different switch block designs arranged in a checkerboard pattern. If the above switch blocks are assigned to the white square locations, a modified one can be used on the black square locations. These black switch blocks are characterized by their own mapping functions, g. Ad hoc designs for various g switch blocks, which are chosen to be slightly different from their f counterparts, are shown in Table 1. In choosing the specific g_e functions for the *disjoint* and *universal-TG* blocks, care is taken to preserve their layout structures by merely re-ordering the horizontal tracks.

3.2 Application: *diverse* and *diverse-clique* Designs

This section will use the design framework to develop commutative switch blocks that are maximally diverse for all possible two-turn paths. Two different switch blocks will be designed, *diverse* and *diverse-clique*. The latter design is more restricted because its endpoint subblock uses the 4-wire clique layout structure of the *disjoint* switch block.

Table 1. Complete switch block mappings used for white (f) and black (g) squares

	White Square Switch Block					Black Square Switch Block			
Turn	*disjoint*	*universal-TG*	*Imran*	*shifty*	Turn	*disjoint*	*universal-TG*	*Imran*	*shifty*
$f_{e,1}$	t	$W-t-1$	$W-t$	$t-1$	$g_{e,1}$	$t-1$	$W-t-2$	$W-t+3$	$t-8$
$f_{e,2}$	t	t	$t+1$	$t-3$	$g_{e,2}$	$t+1$	$t+1$	$t+3$	$t-7$
$f_{e,3}$	t	$W-t-1$	$W-t-2$	$t-2$	$g_{e,3}$	$t+1$	$W-t$	$W-t+2$	$t-9$
$f_{e,4}$	t	t	$t-1$	$t-4$	$g_{e,4}$	$t-1$	$t-1$	$t+1$	$t-6$
$f_{m,i}$	t	t	t	t	$g_{m,i}$	$t+1$	$t+1$	$t+1$	$t+1$

This design approach is repeated for an architecture containing two layout tiles, f and g, arranged in a checkered pattern.

Design Space Let each switch block mapping function be represented by the equations $f_i(t) = t + a_i \bmod W$ or $g_i(t) = t + b_i \bmod W$, where i represents one of the endpoint or midpoint turn types. The a_i and b_i values are constants which can be summarized in vector form as:

$$\boldsymbol{x}_W = \left[\, a_{e,1}\ a_{e,2}\ a_{e,3}\ a_{e,4}\ a_{m,1}\ a_{m,2}\ a_{m,3}\ b_{e,1}\ b_{e,2}\ b_{e,3}\ b_{e,4}\ b_{m,1}\ b_{m,2}\ b_{m,3} \,\right]^T .$$

Note that a solution $\boldsymbol{x}_W$ is only valid for a specific value of W. Constraining f and g in this way explores only a portion the design space. However, this is sufficient to develop very diverse switch blocks.

Enumerating the Path-Pairs Before counting diversity, we enumerate all paths containing two turns and the pairs of these paths that should be diverse.

The six basic two-turn paths created by $\binom{4}{2} = 6$ pairs of single turns are: ENE, ESE, ENW, WNE, NES and SEN, where N, S, E, or W refer to compass directions. For example, two different ENE paths, using columns A and B to reach row *out1*, are shown in Figure 5. In general, the commutative property allows all ENE paths (or ESE paths, etc.) of an infinite routing fabric to be enumerated using the 8×8 grid or *supertile* in Figure 5. The size of the supertile arises from the length-four wires and two (checkerboard) layout tiles. Within it, each subblock is labeled with the mapping functions from one track group.

A number of isomorphic paths can be eliminated using the supertile and commutative property. Longer horizontal or vertical distances would reach another supertile and turn at a switch block equivalent to one in this supertile. Similarly, other input rows can be disregarded. Since NEN and SES paths are commutatively equivalent to ENE and ESE paths, they are also ignored.

For maximum diversity, each *pair of paths* that reach the same output row must reach different tracks. With 8 possible routes (columns A–H), there are $\binom{8}{2} = 28$ pairs of paths to be compared. Hence, for all turn types and all output rows, there are $6 \times 7 \times 28 = 1176$ *path-pairs* to be compared.

Counting Diversity To detect diversity between a pair of paths, first compute the difference between the two permutation mappings, $y = f_{pathA} - f_{pathB}$. The path-pair is diverse if y is non-zero. This can be written in matrix form as $\boldsymbol{y} = \boldsymbol{A} \cdot \boldsymbol{x}_W$ where each

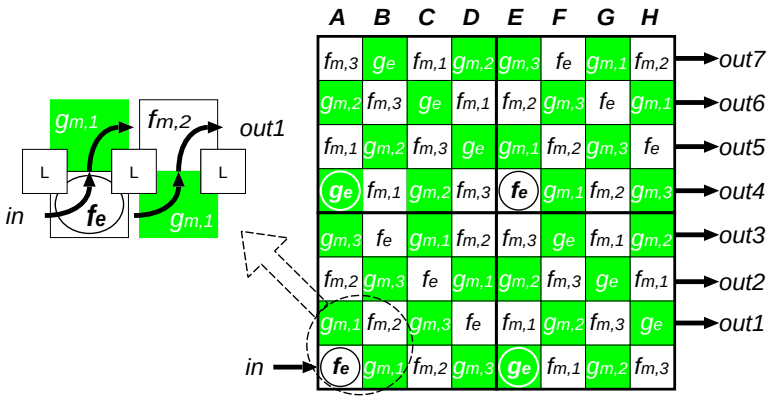

Fig. 5. An 8 × 8 grid or supertile used for enumerating all two-turn paths.

row in $\boldsymbol{A}$ is predetermined based on the path-pair being considered, and each row in $\boldsymbol{y}$ is the corresponding diversity test result.The number of rows has been considerably reduced by the large number of equivalent paths eliminated earlier. Additional types of path-pairs (not only those with two turns) can be represented by adding more rows to $\boldsymbol{A}$ and $\boldsymbol{y}$.

Diversity of a given switch block $\boldsymbol{x}_W$ is measured by counting the number of non-zero entries in $\boldsymbol{y}$. For our architecture, the maximum diversity is 1176.

Searching Design Space Rather than solve large matrix equations to maximize the number of non-zero values in $\boldsymbol{y}$, we performed various random and brute-force searches of $\boldsymbol{x}_W$ for W ranging from 2 to 18. Typically, an exhaustive search produced the best results in about one CPU-day (1 GHz Pentium) even though it wasn't allowed to run to completion.

Switch Blocks Created Using the above procedure, switch blocks named *diverse* are designed for a variety of track group widths, $W \leq 18$. For each W, a solution set $\boldsymbol{x}_W$ is found. Similarly, we designed a *diverse-clique* switch block which preserves the 4-wire clique structure at endpoint subblocks. A layout strategy for these cliques is given in [9]. The precise solution sets obtained for these two switch blocks can be found in [8].

4 Results

The new switch blocks are evaluated below by counting diversity and computing the minimum channel width and area from numerous routing experiments.

Diversity Results The diversity of various switch blocks is shown in Figure 7. The *disjoint* switch block has no diversity but its checkered version has considerably more. The *shifty* switch block and its checkered version provide even more diversity. However, the *diverse* and *diverse-clique* checkered switch blocks reach the highest levels

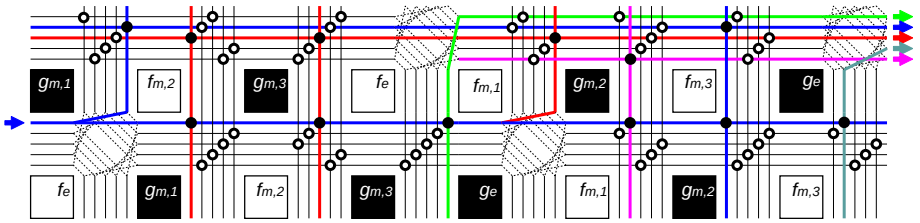

Fig. 6. Checkered layout showing diversity of a $W = 5$ *diverse-clique* switch block.

of diversity. For $W > 10$, these are within 99% of the maximum possible. However, note that it is impossible to attain maximum diversity when $W < 8$ because some of the 8 global routes necessarily map to the same track. Considering this, the *diverse* and *diverse-clique* switch blocks perform very well at being diverse.

Routing Results The experimental environment is similar to the one used in [4]. Benchmark circuits are mapped into 4, 5, and 6-input LUTs, clustered into groups of 6, and placed once. The new switch blocks are evaluated using a modified version [8] of the VPR router [4]. Routing experiments use only length four wires in the interconnect. Half of all tracks use pass transistors 16x minimum width, and the other half use buffers of size 6x minimum [10]. Although not shown, similar results are obtained if all wiring tracks contain buffers.

The routability performance of the new switch blocks is presented in Figures 8 and 9. The former plots the minimum number of tracks required to route, W_{min}, while the latter plots the transistor area of the FPGA at the low-stress point of $1.2 \times W_{min}$ tracks. The graphs on the left compare *shifty* to the older switch blocks, and the graphs on the right compare *disjoint* to the newer switch blocks. A number of different curves are drawn in the graphs, corresponding to different LUT sizes (as labeled) and whether one layout tile is used (**bold** curves) or two tiles are checkered (thin curves). Delay results have been omitted because there is no apparent correlation with switch block style.

The area and W_{min} results exhibit only small variations across the designs, so conclusions might be sensitive to noise and must be carefully drawn. Each data point is an arithmetic average of the twenty largest MCNC circuits, so large variations should not be expected unless many circuits are affected. To mitigate the influence of noise, it is important to identify trends in the results, *e.g.*, across all of the different LUT sizes.

Analysis One clear trend in the routing results is that the plain *disjoint* switch block performs worse than any perturbation of it (including its own checkered version). Beyond this, the ranking of specific switch blocks is difficult. It appears that *shifty* is the best, followed by *universal-TG* and *Imran*, then *disjoint*. The diversity-optimized switch blocks are better than *disjoint*, but worse than *shifty*.

In addition to *shifty*, a variety of other ad hoc switch blocks (both commutative and not) were explored. The *shifty* design gives better results, but the differences are small.

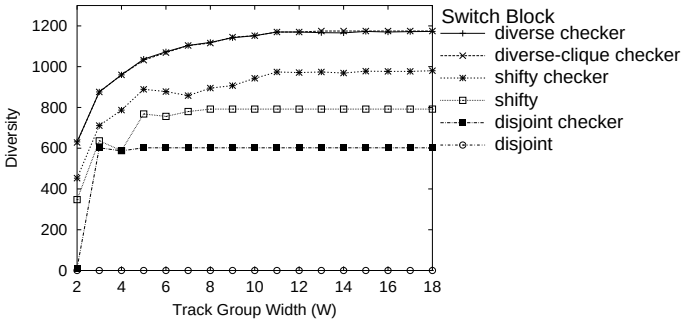

Fig. 7. Diversity of various commutative switch blocks.

These experiments did not clearly suggest that one particular design is significantly better. The effectiveness of *shifty* demonstrates that the twist or non-commutative features of the *universal-TG* and *Imran* blocks is not likely the key factor to their good performance. However, it makes us ask why is track shifting effective? Is it because of increased diversity?

The diverse switch blocks always require fewer routing tracks than the *disjoint* baseline. However, *shifty* always outperforms the diversity-optimized ones. This suggests that it is **not** the diverse property that makes *shifty* effective. It also counters the belief that the *Imran* and *Wilton* switch blocks are effective because they add some diversity.

Why do the diversity-optimized switch blocks not perform as well as anticipated? One conjecture is that negotiated-congestion type CAD tools, like the VPR router, might have difficulty with diversity. This seems plausible because a local re-routing near the source of a net would force most downstream connections to use a new track, even if they continue using the same routing channels. With less diversity, it may be easier for a net to resume using the previous routing tracks. This difficulty might increase the number of router iterations, but our experiments show little increase. Diversity adds a new degree of freedom to routing, but CAD tools must be able to efficiently utilise it.

5 Conclusions

This paper presents an analytical framework for the design of switch blocks. It is the first known framework to consider the switch block as part of an infinite switching fabric that easily works with long wire segments. The most fundamental component of this framework is the new track group switch block model.By separating wiring tracks into independent groups, each can be considered separately. Using permutation mapping functions to model switch block turns adds a mathematical representation to the framework. With commutative switch blocks, the order in which a net executes turns becomes unimportant and the network is easier to analyse. This framework can design diverse switch blocks, but it is not clear the router is utilising this diversity.

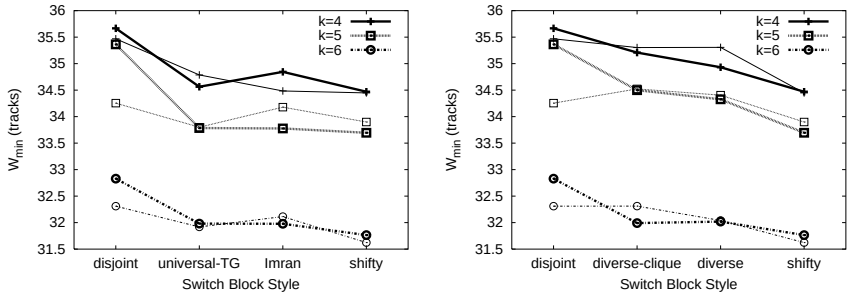

Fig. 8. Minimum channel width results using the new switch blocks.

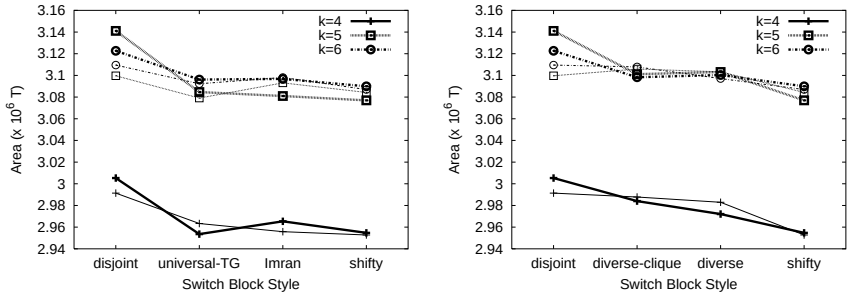

Fig. 9. Area results using the new switch blocks.

References

1. Xilinx, San Jose, CA, *Online Databook*, 2002. `www.xilinx.com/partinfo/databook.htm`.
2. Y.-W. Chang, D. F. Wong, and C. K. Wong, "Universal switch-module design for symmetric-array-based FPGAs," *ACM Transactions on Design Automation of Electronic Systems*, vol. 1, pp. 80–101, January 1996.
3. S. J. Wilton, *Architectures and Algorithms for Field-Programmable Gate Arrays with Embedded Memories*. PhD thesis, Dept. of ECE, Univ. of Toronto, 1997.
4. V. Betz, J. Rose, and A. Marquardt, *Architecture and CAD for Deep-Submicron FPGAs*. Boston: Kluwer Academic Publishers, 1999.
5. M. I. Masud and S. Wilton, "A new switch block for segmented FPGAs," in *Field Programmable Logic*, August 1999.
6. H. Fan, J. Liu, Y.-L. Wu, and C.-C. Cheung, "On optimum switch box designs for 2-D FPGAs," in *ACM/IEEE Design Automation Conference*, June 2001.
7. J. Rose and S. Brown, "Flexibility of interconnection structures in fi eld-programmable gate arrays," *IEEE J of Solid State Ccts*, vol. 26, pp. 277–282, March 1991.
8. G. Lemieux, *In preparation*. PhD thesis, Dept. of ECE, Univ. of Toronto, 2002.
9. H. Schmit and V. Chandra, "FPGA switch block layout and evaluation," in *Int. Symp. on FPGAs*, (Monterey, CA), pp. 11–18, February 2002.
10. G. Lemieux and D. Lewis, "Circuit design of routing switches," in *Int. Symp. on FPGAs*, (Monterey, CA), pp. 19–28, Feb. 2002.