

Theoretical Analysis of Effect of LUT Size on Area and Delay of FPGA*

Gao Haixia, Yang Yintang, and Dong Gang

(Microelectronics Institute, Xidian University, Xi'an, 710071, China)

Abstract : Based on architecture analysis of island-style FPGA, area and delay models of LUT FPGA are proposed. The models are used to analyze the effect of LUT size on FPGA area and performance. Results show optimal LUT size obtained by computation models is the same as that from experiments: a LUT size of 4 produces the best area results, and a LUT size of 5 provides the better performance.

Key words : FPGA; LUT; computation models; area; delay

EEACC : 1265A; 1265Z

CLC number : TN791

Document code : A

Article ID : 0253-4177(2005)05-0893-06

1 Introduction

Since their introduction in 1985, field programmable gate arrays (FPGAs) have already shared a considerable IC market and are expected to grow steadily. A maximum gate count of ten million has been reached and the achievable clock frequency has arrived at 300MHz. However, a circuit implemented in an FPGA is about 10 times larger and 3 times slower than the same circuit implemented via a masked programmable gate array (MPGA) in an equivalent process. The larger size of FPGA circuitry makes FPGA implementations more expensive than MPGAs for high volume designs, and the limited speed of FPGAs precludes their use in very high-speed designs. Consequently, there is a compelling motivation to research FPGA architecture including routing architecture and logic block architecture.

In this paper we focus on the logic block architecture. Several studies in the past have examined the effect of logic functionality on the area and per-

formance of FPGAs. The work in Refs. [1, 3] showed that a LUT size of 4 is the most area efficient. In addition, it was demonstrated in Refs. [2, 5] that using a LUT size of 5 to 6 gives the best performance. The work in Ref. [11] has suggested that using a heterogeneous mixture of LUT sizes of 2 and 3 was equivalent in an area efficiency to a LUT size of 4, and hence could be a good choice. Recently, the work in Ref. [8] showed that under deep sub-micron technology a LUT size of 4 to 6 is the most area efficient, and using a LUT size of 6 gives the best performance.

However, prior results^[1~5,8] were based on CAD experiments that need large time. We want to predict the effect of LUT size through theoretical derivations.

2 FPGA architecture assumption

The typical FPGA architecture is an "island-style" structure where logic blocks are surrounded by routing channels as shown in Fig. 1. The number of tracks in all of the routing channels, W , is

* Project supported by National Defense Advanced Foundation of China (No. 41308010205)

Gao Haixia female, was born in 1977, PhD candidate. Her research interest is in design and testing of FPGAs.

Received 25 September 2004, revised manuscript received 22 November 2004

© 2005 Chinese Institute of Electronics

the same ,and the I/O pads are evenly distributed around the perimeter of the FPGA.

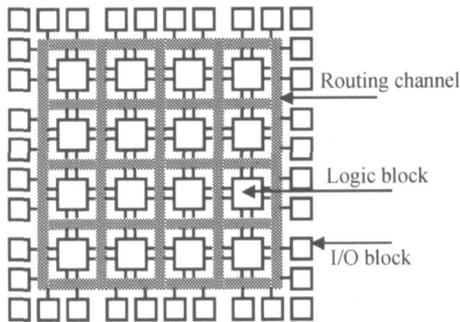


Fig. 1 Island-style FPGA

The structure of the logic block is illustrated in Fig. 2. It consists of a K -input LUT,a DFF and a two-input MUX. A K -LUT can implement any K to 1 combinational logic function ,and the number of implemented function is 2^{2^K} . The LUT style of logic block is chosen because it is easy to vary the functionality of the logic block by changing the number of inputs to the LUT. LUT output is connected to the DFF, which is included because sequential logic is a fundamental component of digital logic. LUT output and registered output are two inputs of the MUX that determines whether the registered or unregistered output is the logic block output.

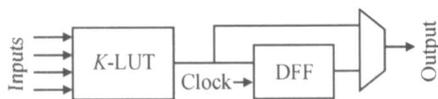


Fig. 2 Structure of logic block

3 Area as a function of K

3.1 Area model

First ,we introduce the concept of “ tile ”:a tile is composed of a logic block and the routing resources at its right and bottom sides as shown in Fig. 3. The area of a tile is presented as

$$A_{tile} = A_l + A_r \tag{1}$$

where A_l and A_r are the area of a logic block and the routing resources in a tile respectively. For a K -LUT FPGA ,three terms in Eq. (1) are replaced by

A_{tile}^K, A_l^K , and A_r^K respectively.

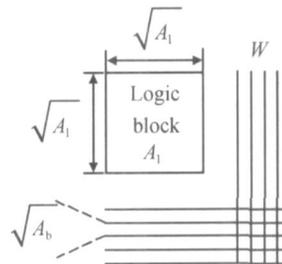


Fig. 3 Area model

The area of a logic block shown in Fig. 2 is the sum of the area of a K -LUT and the area of other fixed logic , A_f , that consist of the routing and circuitry to access K -LUT ,a DFF ,and all other interconnection hardware. A K -LUT requires 2^K bits of information to be stored in a LUT ,because it can implement any K to 1 logic function. If the area required to store one bit is A_b ,the area of a LUT is proportional to 2^K . Using A_b and A_f ,we have the following expression for logic block area

$$A_l^K = 2^K A_b + A_f \tag{2}$$

For given technology ,logic block area is the function of K since A_b and A_f are constants.

To determine routing area ,the area due to each routing track is required including the area of metal wire and that between two neighbor tracks. Each routing track will need at least one bit of information in it to control if a switch is opened or closed. The pitch of a routing track is approximated as the square root of the area required by a bit. Given channel width W_k ,both the area of horizontal and vertical channel in a tile is $A_b W_k^2 + \sqrt{A_b} \times \sqrt{A_l^K} W_k$. Because there is a shared area $A_b W_k^2$ between horizontal and vertical channel ,the expression for the routing area in a tile is

$$A_r^K = A_b W_k^2 + 2 \sqrt{A_b} \sqrt{A_l^K} W_k \tag{3}$$

Combining Eqs. (2) and (3) ,the chip area required to implement any logic on a FPGA is

$$A_{total}^K = A_{tile}^K N_K = (\sqrt{A_b} W_k + \sqrt{A_l^K})^2 N_K \tag{4}$$

where N_K is the block number required. Because A_b is a constant and A_l^K can be found according to Eq. (2) ,remaining terms W_k and N_K need to be determined further.

3.2 Block number for different K

The prediction of logic block number is based on so-called Rent rule^[6], which establishes an empirical relationship between the number of pins T and the number of logic blocks N in a logic design. Nowadays, Rent rule can be widely used in the following areas: layout parameter estimations in electronic design automation, studies of new computer architectures and the generation of synthetic circuit benchmarks^[13]. It shows that a log-log plot between two parameters forms a straight line and empirically yields a relationship

$$T = N^p \tag{5}$$

where \bar{p} is proportionality constant, p is the Rent exponent. \bar{p} is also the average number of pins per block which normally is used to show the complexity of a logic circuit. Chan *et al.*^[7] found p for FPGA devices lies in 0.7 ~ 0.8. We assume $p = 0.75$ throughout the paper.

According to Fig.2, each logic block has $K + 1$ pins. For a given logic circuit, following expressions are true.

$$(K + 1)(N_K)^p = (K + 1 + 1)(N_{K+1})^p \tag{6}$$

$$\frac{N_2}{N_K} = \left(\frac{K+1}{3} \right)^{1/p} \tag{7}$$

Table 1 gives the ratio of N_2 to N_K when K lies in 3 ~ 7. Second row shows theoretical results coming from Eq. (7), and third row gives experimental results in Ref. [8]. Obviously, theoretical results match experimental results well.

Table 1 Ratio of N_2 to N_K

K	3	4	5	6	7
Theoretical results	1.46	1.97	2.50	3.08	3.67
Experimental results ^[8]	1.42	1.96	2.45	3.04	3.40
Difference/ %	2.8	0.5	2	1.3	7.9

When the number of logic blocks required to implement a logic circuit for a given K is known, it is reasonable to determine that for the other K 's using Eq. (6).

3.3 Channel width

Analytical model for channel width is based on

a two dimensional stochastic model by El Gamal^[9]. Although the results in Ref. [9] were developed for master slice circuits, they can also be applied to the FPGAs considered here, since both types of devices are based on a two dimensional array of identical cells. The average channel width is given by

$$W = \bar{R}/2 \tag{8}$$

where \bar{p} is the average number of pins per block, and \bar{R} is average wire-length on a chip. \bar{R} is the estimated using the formula in Ref. [10]

$$\bar{R} = \frac{2}{3} \times \frac{1-p}{p-1/2} \times [(N)^{p-1/2} - 1] \tag{9}$$

where p is Rent exponent, and N is the number of logic blocks.

For a K -LUT FPGA, the channel width W_K can be expressed as

$$W_K = \frac{K+1}{2} \times \frac{2}{3} \times \frac{1-p}{p-1/2} \times [(N_K)^{p-1/2} - 1] \tag{10}$$

3.4 Results

We adopt the same A_b and A_f value as that of Ref. [1] to verify area model through comparing theoretical and experimental results. There are not any detailed circuits employed because we want to know the effect of K on area, but not detailed area value. We assume $N_2 = 10^4$ for $K=2$, which is large enough.

Figure 4 is a plot of the number of logic blocks and block area versus K using theoretical data of Eqs. (7) and (2). The logic block area increases exponentially in K and the number of logic blocks is a decreasing function of K . The reason is that a larger logic block can implement more of the logic circuit, and the number of logic blocks to implement given logic circuit is decreased.

Figure 5 gives a plot of channel width as a function of K using theoretical data of Eqs. (7) and (10). There is an approximately linear increase in channel width as the LUT size K is increased.

Figure 6 shows logic block area and routing area in a tile as a function of K . The data shows

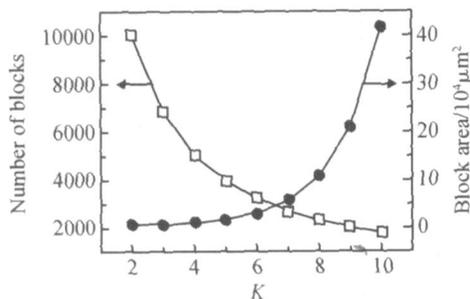


Fig. 4 Number of blocks and block area versus K

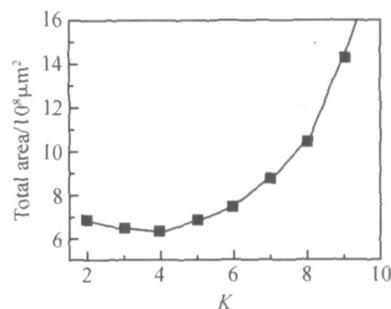


Fig. 7 Total area versus K

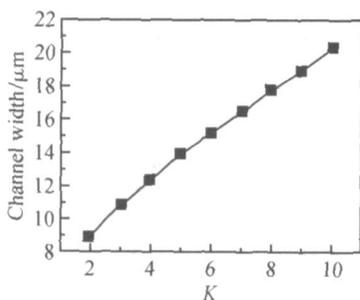


Fig. 5 Channel width versus K

routing area is an increasing function of K as channel width is an increasing function of K . In addition, for all K , routing area is much larger than block area. The minimum of the ratio of routing area to block area is 1.65, and the maximum reaches 10.36. Thus FPGA area is more affected by routing area but not logic block area.

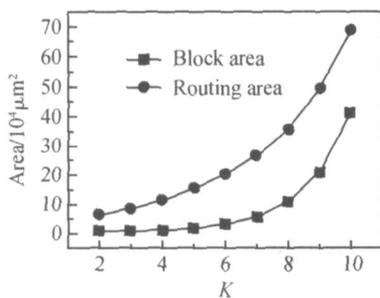


Fig. 6 Block area and routing area versus K

Figure 7 is a plot of total FPGA area versus K . When K is increased, the curve falls first and then hoists with the lowest point at $K = 4$. This means the optimal K for area is 4. This conclusion is the same as that of Refs. [1, 3].

4 Performance as a function of K

4.1 Delay model

We try to determine the effect of K on FPGA performance by comparing the delay of an arbitrary N to 1 logic that has been mapped onto those different K -LUT FPGAs. Since the function is arbitrary, a guaranteed way of mapping is to construct an N -input look-up table on the chip using those K -LUTs. The total number of logic blocks M_K needed to implement an N -LUT on a K -LUT chip and the depth D_K of the implementation will be given in Sec. 4.2.

Assuming the lumped capacitance model is used, the delay time for a logic gate is in proportion to C_l / C_b ^[12], where C_l and C_b are its load and gate capacitance respectively. Extending it to a logic block, the delay T is proportion to $(C_{lk} + \overline{C_{rk}}) / C_g$, where C_{lk} is the logic block load capacitance, $\overline{C_{rk}}$ is the average routing capacitance, and C_g is the gate capacitance of input buffer. The delay T_K of implemented logic on a K -LUT FPGA can be expressed as

$$T_K = \alpha \frac{C_{lk} + \overline{C_{rk}}}{C_g} \times D_K \quad (11)$$

where α is some proportionality constant, and D_K is the depth of the implementation. Because K -LUT has 2^{K-3} times more logic content than that of 3-LUT, we can find

$$C_{lk} / C_l3 = 2^{K-3} \quad (12)$$

Furthermore,

$$\frac{C_{13}}{C_{13}} = \frac{W_K r_K}{W_3 r_3} = \frac{K+1}{2} \times \frac{R_K r_K}{R_3 r_3} = \frac{K+1}{4} \times \frac{(N_K)^{p-1/2} - 1}{(N_3)^{p-1/2} - 1} \times \frac{(M_K)^{p-1/2} - 1}{(M_3)^{p-1/2} - 1} \quad (13)$$

where W_K is channel width, R_K and r_K are average wire length of the chip and average interconnection length of implemented logic respectively, N_K is the number of logic blocks on a chip, and M_K is the number of logic blocks required to complete any N to 1 logic function. Let numerator and denominator divided by C_{13} , and assume $C_{13}/C_{13} = u$, following expression is obtained

$$T_K = \frac{C_{13}}{C_g} \times D_K \times \left(2^{K-3} + u \times \frac{K+1}{4} \times \frac{(N_K)^{p-1/2} - 1}{(N_3)^{p-1/2} - 1} \times \frac{(M_K)^{p-1/2} - 1}{(M_3)^{p-1/2} - 1} \right) \quad (14)$$

According to Eq. (6), following expression is true.

$$N_K = \left(\frac{4}{K+1} \right)^{1/p} \times N_3 \quad (15)$$

For a special technology, delay coefficient C_{13}/C_g and u are constants. To obtain the delay, we only need to further determine the number of logic blocks M_K and the implementation depth D_K needed to implement an N -LUT on a K -LUT chip.

Given $t_K = \frac{T_K}{C_{13}/C_g}$, the change of delay with K can be reflected by that of t_K .

4.2 Estimation of M_K and D_K

In order to obtain better performance, N -LUT is constructed using tree-like structure. Figure 8 shows a schematic diagram of constructing 7-LUT using 4-LUTs. The total number of logic blocks needed to implement N -LUT using 4-LUTs is

$$M_4 = 2^{N-4} \times \left(1 + \frac{1}{4} + \frac{1}{4^2} + \dots \right) = 2^{N-4} \times \frac{1}{1-1/4} \quad (16)$$

and the depth of implementation is

$$D_4 = (N - 4)/2 + 1 \quad (17)$$

Generalizing Eqs. (16) and (17) to a common K ,

$$M_K = 2^{N-K} \times \left(1 + \frac{1}{K} + \frac{1}{K^2} + \dots \right) = 2^{N-K} \times \frac{1}{1-1/K} \quad (18)$$

$$D_K = (N - K)/x(K) + 1 \quad (19)$$

where $x(K)$ is a solution of equation $K = x + 2^{x-1}$.

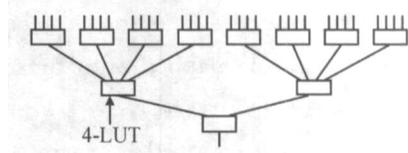


Fig. 8 Constructing 7-LUT with 4-LUTs

4.3 Results

We plot the t_K versus K for different u in Fig. 9 as $N = 20$. The optimal K increases lightly with u . For all u , there are the best performance when K lies between 5 and 6 which is consistent with those of experiments. The increasing u means the advancement of technology. Under the deep sub-micron technologies, routing delay takes the most of total delay. Increasing K reasonably will enhance the functionality of each logic block, and decrease the routing delay, thus improving the speed of the chip. One should notice that delay coefficient in all kinds of technologies is different, and can not draw a conclusion that delay is increased with increasing u from Fig. 9. In fact, delay is becoming low when u is increased.

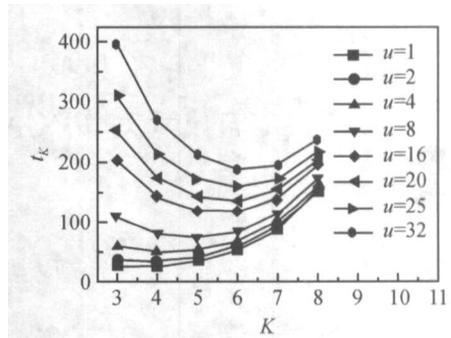


Fig. 9 t_K versus K for different u

5 Conclusion

In this paper we first analyze global architecture of island-style FPGA and internal structure of logic block, then an area model of LUT FPGA and an approach for estimating delay are proposed, which are suitable to symmetric FPGAs. Based on Rent rule, the number of logic blocks required to implement given logic circuit versus K is predic-

ted. Channel width is estimated using El Gamal's result. The effect of K on FPGA performance is evaluated by comparing the delay of an arbitrary N to 1 logic that has been mapped onto those different K -LUT FPGAs. Arbitrary N to 1 logic is implemented by constructing a N -LUT using tree-like structure on a K -LUT FPGA. The results show 4-LUT is the best efficient, and using a LUT size of 5 to 6 gives the best performance. The conclusion of the optimal LUT size consists with experimental conclusion.

References

- [1] Rose J, Francis R J, Lewis D, et al. Architecture of field-programmable gate arrays: The effect of logic functionality on area efficiency. *IEEE J Solid-State Circuits*, 1990, 25(5) : 1217
- [2] Kouloheris J, El Gamal A. FPGA performance versus cell granularity. *Proceedings of Custom Integrated Circuits Conference*, 1991 : 6
- [3] Kouloheris J, El Gamal A. FPGA area versus cell granularity - lookup tables and PLA cells. *First ACM Workshop on FPGAs*, Berkeley, CA, 1992 : 9
- [4] Hill D, Woo N S. The benefits of flexibility in look-up table FPGAs. *Proceedings of Oxford International Workshop on Field Programmable Logic and Applications*, 1991
- [5] Singh S, Rose J, Chow P, et al. The effect of logic block architecture on FPGA performance. *IEEE J Solid-State Circuits*, 1992, 27(3) : 281
- [6] Landman B S, Russo R L. On a pin or block relationship for partitions of logic graphs. *IEEE Trans Comput*, 1971, C-20 : 1469
- [7] Chan P K, Schlag M D F, Zien J Z. On routability prediction for field programmable gate arrays. *Proceedings of 30th ACM/ IEEE Design Automation Conference*, 1993 : 326
- [8] Ahmed E, Rose J. The effect of LUT and cluster size on deep-submicron FPGA performance and density. *IEEE Trans Very Large Scale Integration (VLSI) Systems*, 2004, 12(3) : 288
- [9] El Gamal A. Two-dimensional stochastic model for interconnections in master slice integrated circuits. *IEEE Trans Circuits Syst*, 1981, CAS-28(2) : 127
- [10] Li Wei, Banerji D K. Routability prediction for hierarchical FPGAs. *Proceedings of Great Lakes Symposium on VLSI*, 1999 : 256
- [11] Kaptanoglu S, Bakker G, Kundu A, et al. A new high density and very low cost reprogrammable FPGA architecture. *Proceedings of ACM International Symposium on Field-Programmable Gate Arrays*, 1999 : 3
- [12] Song Renru, Ruan Gang, Liang Qingqing. Comprehensive delay model for CMOS inverters based on velocity saturation. *Chinese Journal of Semiconductors*, 2000, 21(7) : 711
- [13] Christie P, Stroobandt D. The interpretation and application of rent's rule. *IEEE Trans Very Large Scale Integration Systems*, 2000, 8(6) : 639

LUT 尺寸对 FPGA 面积和延迟影响的理论分析*

高海霞 杨银堂 董刚

(西安电子科技大学微电子研究所, 西安 710071)

摘要: 在分析隔离岛式 FPGA 结构的基础上,提出了基于 LUT 的面积和延迟模型,用于分析 LUT 尺寸对 FPGA 面积和性能的影响。结果表明利用计算模型得到的最佳 LUT 尺寸与实验结论一致:4-LUT 获得最好的面积有效性,5-LUT 获得较好的延迟。

关键词: FPGA; LUT; 计算模型; 面积; 延迟

EEACC: 1265A; 1265Z

中图分类号: TN791

文献标识码: A

文章编号: 0253-4177(2005)05-0893-06

* 国防预先研究基金资助项目(批准号:41308010205)

高海霞 女,1977 年出生,博士研究生,主要从事 FPGA 的设计和测试工作。
2004-09-25 收到,2004-11-22 定稿