

---

School of Electrical, Electronic & Computer Engineering



---

# **The Future Looks Gloomy for FPGA Interconnects**

Terrence Mak

Technical Report Series

NCL-EECE-MSD-TR-2009-145

---

April 2009

Contact:

t.mak@imperial.ac.uk

I would like to thank Alex for his stimulating discussion and Fei for his comments on preparing this report.

NCL-EECE-MSD-TR-2009-145 Copyright © 2009 Newcastle University

School of Electrical, Electronic & Computer Engineering,  
Merz Court,  
Newcastle University,  
Newcastle upon Tyne, NE1 7RU, UK

<http://async.org.uk/>

# The Future Looks Gloomy for FPGA Interconnects

Terrence Mak

April 2009

## Abstract

It is well known that interconnects in FPGA dominate the system performances and power consumption. Especially, long interconnects exhibit a substantial delay and often lead to timing violation and require further optimizations. Here, I developed an analytical model for general FPGA interconnections and performed experiments using VPR (Virtual Place-and-Route) to answer two questions: (i) What does the FPGA interconnect delay look like in the future? (ii) How many long lines are there in an FPGA circuits and how do these long lines contribute to the power dissipation? The answers to these two questions lead to an interesting conclusion that the future of FPGA interconnects looks grim.

## 1 Global Interconnect Modelling

Having an accurate and efficient model for characterizing the interconnection in FPGA is crucial to the circuit implementation. A interconnect model can be applied in FPGA architectural evaluation and predicting interconnect performance for future technology processes. However, interconnections in FPGAs are fairly irregular circuits and idiosyncratic versus conventional wires in ASIC design. The complexity of the interconnection circuit creates a difficulty to generalize and analyze the performance of interconnect signaling. Here, I present a new approach to model long range routings in FPGAs.

Consider an island-style FPGA architecture which comprises a 2D array of logic blocks or slices that can be interconnected via programmable routing. A commercial FPGA, such as Virtex-II style logic clock, comprising four LUTs, each consisting of two 4-input Look-Up-Tables (LUTs), two flip-flops and programming overheads, is assumed. The programmable routing comprises horizontal and vertical routing channels each having sets of different interconnect segments. The segments can be connected to the inputs and outputs of the logic block via connection boxes and to each other via switch boxes.

### 1.1 Multiple-Stage Model

A global interconnection comprises of a combination of programmable interconnects that is physically spanning over a long distance from the source to the sink. Particularly, the interconnection in FPGAs is much more complicated than conventional long line found in ASIC design. Fig. 1(a) shows a typical example of an interconnection. It starts from a source, which is the output of a logic block, and is connected to a short wire segment via interconnect switch. There are a number of ways to realize the interconnect switch. The simplest way is by using a pass transistor [LLTY04, BRM99]. Other approaches such as stages of multiplexer [LLM06] and transmission gates can also achieve the same functionality with a higher speed. Although a pass transistor is shown in the figure, it can be replaced by other types of logics, such

Table 1: Notations used in the derivation

$R_i^d$	Input resistance of the driver at the $i$ -th segment
$C_i^d$	Load and intrinsic capacitances of the driver at the $i$ -th segment
$R_i^s$	Thevenin equivalence resistance of the $i$ -th segment
$C_i^s$	Thevenin equivalence capacitance of the $i$ -th segment
$k_i$	The approximation coefficient for the $i$ -th segment
$\sigma_i$	The time constant for $i$ -th segment
$v_i$	The voltage of the $i$ -th segment
$t$	Time
$\gamma_i$	The discount factor for $i$ -th segment
$T_n$	Propagation delay for the $n$ -segment interconnect
$\Gamma_n$	Throughput for the $n$ -segment interconnect

transmission gates and the closed-form analytical solution remains the same with different resistance and capacitance values. Most modern FPGAs have wire segments of different lengths that can be used to implement interconnections with different requirements. For example, in Xilinx Virtex-4 series FPGAs, there are three types of wire segments, which are with length 1, 3, 6 (Hex-line) and 24 units. Through programmable interconnect switches, wire segments can be connected to form a long range route. Since constructing long range interconnection by aggregating multiple short wires segments increases energy dissipation, long wire segments are frequently being used for realizing global interconnection. Typically these long wire segments will span a long distance (for example 24 tiles in Xilinx Virtex-4 [Xil05]).

In addition, different interconnect routing architectures have been proposed. Examples are the tapped branching lines from long wire, known as early-turns [LLM06]. They are proposed to provide flexible routing to other channels. Also, various short-cut for the interconnect switches are proposed in [LLTY04], such as buffers inserting into the long wire to reduce the delay.

FPGA interconnect is a fairly complex circuit, which comprises of multiple buffers, switches, pass-gates and multiplexers, whereas interconnect in ASIC is rather simple. In order to obtain an analytical model for these long line, the FPGA interconnect structure has to be abstracted and simplified.

The interconnection depicted in Fig. 1(a) can be generalized as a network of resistance and capacitance (RC) pairs, with buffers in between. This is depicted in Fig. 1(b). A simple lumped model can be used to model the short interconnect segment. More complicated models, such as  $\beta$  and  $T$  models [Bak90] can also provide a good approximation with higher accuracy. The switching logics and routing branchings can be modeled by RC circuits and equivalent transistor parasitics can be found by testing circuits. To further generalize the model, I can divide the whole interconnection into  $n$  stages. Each stage is either driven by a driver at each programmable switch or by a buffer in the long wire. By Thevenin's theorem, the combinations of RC pairs at the  $i$ -th stage can be summarized into  $R_i^s$  and  $C_i^s$ , respectively. Following [Bak90, Sak93], I model the driver or buffer by a switch-level RC circuit and, thus, the input resistance and load capacitance of the driver are denoted as  $R_i^d$  and  $C_i^d$  respectively.

The model generalizes the complex interconnection into simple multiple-stage link, which can be characterized by the RC pairs of buffer and interconnects. By resolving the waveform at each of the stages, I can study and evaluate the whole interconnection analytically. In the following subsection, a simple closed form approximation will be reviewed and applied to approximate the waveform at each of the stages.



## 1.2 Waveform Approximation at Each Stage

Since I am interested in the transient behavior at each stage, I can approximate the step response of each stage by considering each stage as a distributed RC line driven by a driver. Therefore, I denote variable  $v_i$  as the fraction of the supply voltage at the far end of the  $i$ -th stage. Thus  $v_i = V_i(t)/V_{DD}$  and it can be expressed in a series expanded with Sakurai's approximation [Sak93]

$$v_i = 1 - \sum_{j=1}^{\infty} k_{i,j} e^{-t_{v_i}/\sigma_{i,j}} \approx 1 - k_{i,1} e^{-t_i/\sigma_{i,1}} \quad (1)$$

To simplify the expression, I drop the index  $j$ , thus I have

$$v_i = 1 - k_i e^{-t_i/\sigma_i} \quad (2)$$

where the coefficient  $k_i$  and time constant  $\sigma_i$  are given

$$\sigma_i = R_i^d C_i^d + R_i^d C_i^s + R_i^s C_i^d + 0.4 R_i^s C_i^s \quad (3)$$

and

$$k_i = 1.01 \frac{R_i^d C_i^s + R_i^s C_i^d + R_i^s C_i^s}{R_i^d C_i^s + R_i^s C_i^d + \frac{\pi}{4} R_i^s C_i^s} \quad (4)$$

I can also provide a more general expression, which is applicable for some stages that have a pass transistor. It is well-known that a pass transistor implemented with a NMOS device is not effective at pulling a node to  $V_{DD}$ . When the pass-transistor pulls a node high, the output only charges up to  $V_{DD} - V_{Tn}$  [RCN03]. Therefore, I let  $\gamma_i$  be the discount factor for the  $i$ -th stage,  $\gamma_i = (V_{DD} - V_{Tn})/V_{DD}$ . Note that using a pass transistor has been abandoned in many commercial FPGA architectures. Instead, multiplexers and transmission gates are used in modern architecture for switch and connection block design. For these cases, I can simply let the discount parameter  $\gamma = 1$ . Then, I can have the general approximation for the rise and fall of  $v_i$  at the  $i$ -th stage as

$$\text{(Rise)} \quad v_i = \gamma_i - \gamma_i k_i e^{-t_i/\sigma_i} \quad (5)$$

$$\text{(Fall)} \quad v_i = \gamma_i k_i e^{-t_i/\sigma_i} \quad (6)$$

The complex interconnection in FPGA has been generalized as a series of buffered segments. Methodologies that were used for ASIC to study global interconnects can be applied here to analyze the overall delay and throughput in an FPGA interconnects. In line with the approach in [DD05], I can derive the delay and throughput for the overall link. However, in contrast to [DD05], interconnect parameters are different at each stage, new expressions of delay and throughput will be derived and presented in the following section.

## 1.3 Delay of a Long Line in FPGAs

Consider a general interconnection that is partitioned into  $n$  stage buffers in the long wire or switching points. The Thevenin's equivalent RC at the  $i$ -th stage are denoted as  $R_i^s$  and  $C_i^s$ , and the input resistance and load capacitance of the driver are denoted as  $R_i^d$  and  $C_i^d$  respectively. The interconnection is modeled

	Technology, in $\mu\text{m}$								
	0.18	0.13	0.1	0.07	0.05	0.035	0.025	0.018	0.013
Aggr. semi-global wires	0.096	0.168	0.260	0.340	0.600	1.224	2.400	4.630	8.876
Aggr. global wires	0.020	0.035	0.054	0.082	0.150	0.306	0.600	1.157	2.219
Cons. semi-global wires	0.096	0.184	0.307	0.627	1.220	2.509	5.014	9.821	19.458
Cons. global wires	0.023	0.044	0.073	0.150	0.292	0.598	1.183	2.298	4.474

Figure 2: The estimation of resistance for different technology processes.

as a multiple-stage link and by resolving the waveform at each of the stage, I can derive the throughput and power of the link analytically.

It is assumed that delay in each stage is given by the 50% rise time (or the fall time)<sup>1</sup> of the interconnect segment. I denote the delay of a driver at the  $i$ -th stage as  $\delta_i$ . Particularly, the delay  $\delta_i$  will also include the delay of logics at the interconnect switch. Thus, the parameter  $\delta_i$  can be unique to each stage as all the interconnect switches can be different, subjected to the routing of the interconnections. The time  $t_i$  for a segment reached  $v_i$  can be derived from Eq. (5) as

$$t_i = \sigma_i \ln \left( \frac{\gamma_i k_i}{\gamma_i - v_i} \right) \quad (7)$$

where  $t_i$  is the time required for the output of the  $i$ -th stage to reach a fraction  $v_i$  of the full-scale voltage. Alternative delay models are available in the literature, such as Elmore Delay. The Elmore delay model is a RC lumped network model, provides a pessimistic estimation. The Bakuglo delay is very similar to the equation I have derived here based on Sakurai's model. The total delay for the voltage at the  $n$ -th stage reached  $v_n$  is given by

$$T_n^{\text{Delay}} = \sigma_n \ln \left( \frac{\gamma_n k_n}{\gamma_n - v_n} \right) + \sum_{i=1}^{n-1} \sigma_i \ln \left( \frac{\gamma_i k_i}{\gamma_i - 0.5} \right) + \sum_{i=1}^n \delta_i \quad (8)$$

The delay-based throughput  $\Gamma_n^{\text{Delay}}$  for an interconnection with  $n$  stages is given by the inverse of the propagation delay as

$$\Gamma_n^{\text{Delay}} = \frac{1}{\sigma_n \ln \left( \frac{\gamma_n k_n}{\gamma_n - v_n} \right) + \sum_{i=1}^{n-1} \sigma_i \ln \left( \frac{\gamma_i k_i}{\gamma_i - 0.5} \right) + \sum_{i=1}^n \delta_i} \quad (9)$$

## 2 Delay Prediction for Long Lines

From the above equations, the line delay has a linear relationship with time constant  $\sigma$ . With technology scaling, coefficient  $k$  converges to 1 and  $\delta$  decreases as the reduction of gate delay. However,  $\sigma$  will increase significantly with the new technology processes, simply because the explosive growing of wire resistance (See Fig. 2<sup>2</sup>).

Using Eq. 8 together with the predictive resistance and capacitance values, the estimation of intercon-

<sup>1</sup>It is common to use rise time to refer to both rise and fall times. This nomenclature will used in the rest of the paper.

<sup>2</sup>R. Ho, "The Future of Wires", Ph.D. thesis, Stanford University

	Technology, in $\mu\text{m}$								
	0.18	0.13	0.1	0.07	0.05	0.035	0.025	0.018	0.013
Aggr. semi-global wires	0.414	0.397	0.374	0.359	0.345	0.315	0.288	0.266	0.247
Aggr. global wires	0.440	0.430	0.403	0.367	0.345	0.315	0.288	0.266	0.247
Cons. semi-global wires	0.414	0.387	0.359	0.333	0.311	0.295	0.287	0.280	0.272
Cons. global wires	0.432	0.403	0.377	0.353	0.332	0.313	0.304	0.296	0.288

Figure 3: The estimation of resistance for different technology processes.

	30 tile	50 tile
	Interconnect Delay (ns)	
180 nm	6.9	11.51
130 nm	7.48	12.47
100 nm	8.1	13.57
70 nm	8.7	14.5
50 nm	10.67	17.8
35 nm	14.9	24.8
25 nm	22.2	37.0
18 nm	35.3	58.9
13 nm	58.9	98.1

Table 2: FPGA interconnect delay predictions

nect delays were computed and listed at Table 2. Here, delay for interconnection with lengths 30 and 50 tiles are presented. Interconnect stage at each tile is assumed to be completely identical and semi-global wires values are used for the computation. It can be observed that the delay increases drastically after 50 nm processes for both the 30 and 50 tiles cases. It is quite clear that interconnect dominates the computational delay. With such prediction, the future FPGA will provide dreadful performance, especially when technology process shrinks into 20 nm or below. The long interconnect delay will also slow down the clocking rate of the system. With a 25 nm process, a circuit implemented on FPGA will have a clock frequency of 45 MHz and the clock rate will be further reduced to 17 MHz for a 13 nm process!

### 3 Experimental Results

Table 3 shows the results of power dissipations of long range routings. These results were obtained from VPR<sup>3</sup> using 20 MCNC benchmark circuits, which includes a large variety of FPGA applications. It is interesting to find out among all the circuits, there are on average 4.5% and 8.6% interconnects have length longer than 50 and 30 tiles, which are regarded as long range routings. In addition, these long interconnections has consumed on average 38.6% and 43.8% overall power of the circuits. Therefore, interconnections with long range routing is a dominant to the circuit power consumption.

Fig. 4 shows the power dissipation of 5 selected benchmark circuits for different interconnection lengths. the average curve the results for the overall average power for the 20 circuits. The full simu-

<sup>3</sup>VPR (Virtual Place and Route) is a software to model different architectures and is able to evaluate the architecture and compute-aided design flow using benchmarking circuits.

	Number of LUTs	% net, $L \geq 50$	% net, $L \geq 30$	% Power ( $L \geq 50$ )	% Power ( $L \geq 30$ )
alu4	1522	2.3	6.9	7.7	15.9
apex2	1878	3.6	7.68	17	23.7
apex4	1262	3.13	9.39	46.5	50
bigkey	1707	0.67	1.15	27.4	28.2
clma	8381	6.3	13.8	46.6	56.8
des	1591	0.56	1.46	8.78	11.08
diffeq	1494	5.1	8.9	29.6	39.1
dsip	1370	0.44	0.87	19.6	20.4
elliptic	3602	5.7	10.2	54	59.6
ex5p	1064	8	10.6	38.78	42
ex1010	4598	1.62	7.98	71.4	75.5
frisc	3539	6.53	13.37	51.2	58.75
misex3	1397	5	10.11	21	33.4
Ppdc	4575	9.65	15.4	63.65	66.78
s298	1930	3.08	3.96	65.46	65.83
s38417	6096	9.65	15.4	63.65	66.78
seq	1750	4.88	8.2	22.9	29.2
spla	3690	7.6	12.97	59.4	63.4
tseng	1046	2.04	5.04	18.07	25
Average		4.52	8.60	38.56	43.76

Table 3: Distribution and power consumption percentages for interconnections with lengths 30 and 50 tiles in 20 MCNC benchmark circuits.

lation results for all circuits are shown in Table 4. These are relatively large circuit with LUTs between 1800 and 3600. The results were obtained using the VPR tool together with the power estimation package. Each data point corresponds to the average power dissipation for interconnect with the particular length. As can be seen from the figure, power grows exponentially with the length. Especially, for interconnect with length longer than 50 tiles, the power dissipation increases drastically. This explains that results from Table 3 that interconnect with length longer than 50 can take up 38.6% of the overall power while there are only 4.5% of these interconnects.

## 4 Conclusion

In this study, I found that FPGA interconnects is poorly scaled. Based on the extrapolation of future device performance, interconnect will become the performance bottleneck, of which the clock rate will be slowed down to 17 MHz in a 13 nm process. In addition, it has been discovered that the interconnect power dissipation increases rapidly with the link length. Typically for links longer than 50 tiles, which consumes on average 38.6% of the overall power, account for 4.5% of all nets. This explains the distribution of power biased to interconnects with longer length.

## References

- [Bak90] H.B. Bakoglu. *Circuits, Interconnections, and Packaging for VLSI*. Addison-Wesley Publishing Company, 1990.

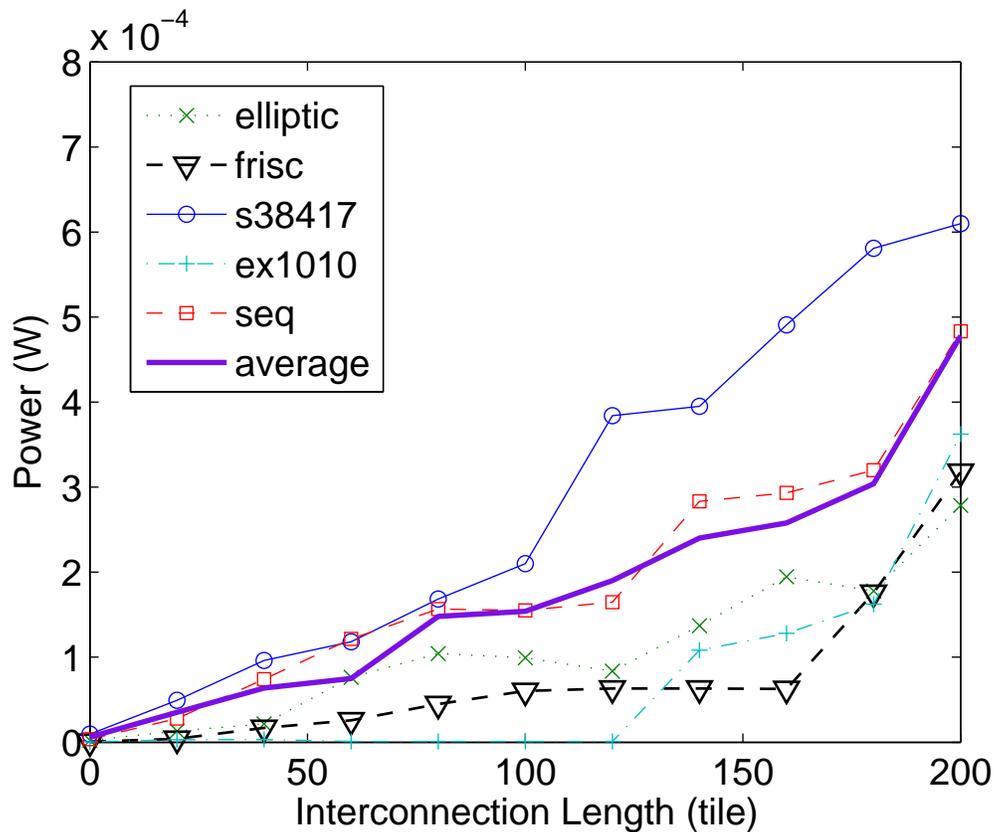


Figure 4: Power dissipation of 5 MCNC benchmark circuits for different interconnection lengths.

- [BRM99] V. Betz, J. Rose, and A. Marquardt. *Architecture and CAD for Deep-Submicron FPGAs*. Kluwer Academic Publishers, 1999.
- [DD05] V. Deodhar and J. Davis. Optimization of Throughput Performance for Low-Power VLSI Interconnects. *IEEE Trans. on VLSI Systems*, 13(3):308–318, 2005.
- [LLM06] E. Lee, G. Lemieux, and S. Mirabbasi. Interconnect Driver Design for Long Wires in Field-Programmable Gate Arrays. In *Proceedings of the International Conference on Field-Programmable Technology*, 2006.
- [LLTY04] G. Lemieux, E. Lee, M. Tom, and A. Yu. Directional and Single-Driver Wires in FPGA Interconnect. In *Proceedings of the International Conference on Field-Programmable Technology*, 2004.
- [RCN03] J.. Rabaey, A. Chandrakasan, and B. Nikolic. *Digital Integrated Circuits: A Design Perspective*. Prentice-Hall, 2003.
- [Sak93] T. Sakurai. Closed-Form Expressions for Interconnection Delay, Coupling, and Crosstalk in VLSI's. *IEEE Trans. on Electron Devices*, 40(1):118–124, 1993.
- [Xil05] Xilinx. Virtex-4 Data Sheets. 2005.

Length (Tile)	1-20	21-40	41-60	61-80	81-100	101-120	121-140	141-160	161-180	181-200	201-220
	Power (W)										
alu4	4.07e-6	2.73e-5	5.80e-5	8.59e-5	1.01e-4	1.27e-4	-	-	-	-	-
apex2	2.79e-6	2.58e-5	7.81e-5	9.43e-5	1.61e-4	2.03e-4	-	-	-	-	-
apex4	2.32e-06	6.85e-06	7.72E-06	6.24E-06	1.65e-4	2.08e-4	2.51e-4	2.37e-4	3.08e-4	5.58e-4	5.93e-4
bigkey	2.92e-5	1.16e-4	1.16e-4	1.16e-4	4.44e-4	-	-	-	-	-	-
clma	1.55e-6	8.32e-6	1.25e-5	2.34e-5	3.79e-5	5.46e-5	7.22e-5	4.04e-5	6.32e-5	6.33e-5	1.55e-4
des	1.26e-5	1.00e-4	1.50e-4	1.50e-4	1.50e-4	3.50e-4	4.50e-4	4.50e-4	8.50e-4	1.02e-3	2.10e-3
diffeq	4.35e-6	3.44e-5	4.56e-5	1.93e-5	3.48e-5	3.33e-4	1.40e-4	3.61e-4	-	-	-
dsip	1.95e-5	1.05e-4	2.01e-4	2.47e-4	-	-	-	-	-	-	-
elliptic	2.27e-6	1.32e-5	2.17e-5	7.63e-5	1.04e-4	9.94e-5	8.35e-5	1.37e-4	1.95e-4	1.78e-4	2.79e-4
ex5p	7.82e-6	2.45e-5	3.57e-5	9.53e-5	1.92e-4	2.85e-4	3.44e-4	4.12e-4	-	-	-
ex1010	4.15e-7	2.98e-6	3.28e-6	1.04e-6	1.02e-6	1.04e-6	1.02e-6	1.08e-4	1.28e-4	1.62e-4	3.62e-4
frisc	5.98e-7	4.09e-6	1.68e-5	2.57e-5	4.49e-5	6.03e-5	6.33e-5	6.33e-5	6.26e-5	1.76e-4	3.18e-4
misex3	4.68e-6	5.63e-5	1.23e-4	1.49e-4	2.07e-4	1.49e-4	4.59e-4	-	-	-	-
pdc	4.53e-7	1.69e-6	4.00e-6	8.63e-6	1.77e-5	1.30e-5	2.23e-5	1.82e-5	3.69e-5	6.66e-5	5.35e-5
s298	1.90e-6	1.39e-5	6.39e-5	6.40e-5	6.39e-5	1.93e-4	2.30e-4	5.51e-4	3.31e-4	7.63e-5	9.63e-5
s38417	9.52e-6	4.93e-5	9.63e-5	1.18e-4	1.68e-4	2.10e-4	3.84e-4	3.95e-4	4.91e-4	5.81e-4	6.10e-4
seq	3.59e-6	2.80e-5	7.41e-5	1.22e-4	1.57e-4	1.55e-4	1.65e-4	2.83e-4	2.93e-4	3.20e-4	4.83e-4
spla	7.21e-7	3.10e-6	9.14e-6	1.72e-5	2.29e-5	2.88e-5	3.10e-5	6.24e-5	7.98e-5	1.43e-4	2.08e-4
tseng	8.52e-6	4.85e-5	9.48e-5	6.03e-6	5.94e-4	-	-	-	-	-	-
Average	6.15e-6	3.52e-5	6.38e-5	7.50e-5	1.48e-4	1.54e-4	1.93e-4	2.40e-4	2.58e-4	3.04e-4	4.78e-4

Table 4: Power consumption (W) for different interconnections of MCNC benchmark circuits. Note that some circuits do not have interconnections of a specific length.