

Macro-models for High Level Area and Power Estimation on FPGAs

Tianyi Jiang, Xiaoyong Tang, Prith Banerjee

Electrical and Computer Engineering, Northwestern University
2145 Sheridan Road, Evanston, IL-60208

Email: {jiang, tang, banerjee}@ece.northwestern.edu

ABSTRACT

As more and more complex applications are implemented on FPGAs, high-level design tools are needed to reduce the design time. A good high-level synthesis tool usually has an automated design space exploration pass to determine the effects of various compiler optimizations on the area and power of the synthesized hardware. Such a pass needs early estimation of area and power. Towards this end, we have developed high-level equation based area and power macro-models for various RTL level operators such as adders, multipliers, and logical operators. The area model is parameterized with the bit width of the device and the power model takes into account input switching activity and input spatial correlation as well as input bit width. These models are derived by actual synthesis of these RTL operators using back-end logic synthesis and place-and-route tools. Compared with the other approaches, our method generated a uniform macro-model for each operator with fewer coefficients and sometimes lower degrees. It is also easier to analyze the power sensitivity to different parameters. Experimental results show that these area and power models are accurate and efficient.

Categories and Subject Descriptors

I.6.5 [Simulation and Modeling]: Model Development – modeling methodologies.

General Terms

Algorithms, Design, Experimentation

Keywords

Model, High-level synthesis, RTL, Area estimation, Power estimation, FPGA.

1. Introduction

While Field-Programmable Gate Arrays (FPGAs) were used in the past for rapid system prototyping due to their short design cycle times, FPGAs are now rapidly approaching ASICs in terms of the complexity of designs that they can support, and their performance. Recent FPGA architectures such as the Xilinx Virtex II [1] and the Altera Stratix [2] are providing built-in DSP functionality such as large numbers of embedded multipliers, DSP blocks and on-chip memories. These FPGAs are capable of implementing high-density, high-performance applications that were once implemented in ASICs.

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, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GLSVLSI'04, April 26-28, 2004, Boston, Massachusetts, USA
Copyright 2004 ACM 1-58113-853-9/04/0004...\$5.00.

As more and more complex applications are implemented on FPGAs, high-level design tools are therefore needed to reduce the design time of complex systems consisting of millions of gates. This paper describes an approach for high-level estimation of area and power within the context of a high-level synthesis tool for FPGA synthesis. This tool implements an automated design space exploration pass which determines the effects of various compiler optimizations on the area and power of the synthesized hardware. Such a pass needs early estimation of area and power. Towards this end, we have developed area models for various RTL level operators such as adders, multipliers, and logical operators, which are parameterized with the bit widths of the devices. These models are derived by actual synthesis of these RTL operators using back-end logic synthesis and place-and-route tools. We have also derived high-level equation based power macro-models for the RTL operators which take into account input switching activities, input spatial correlation and input bit width. The number of the coefficients is less than the other macro-modeling methods without losing accuracy.

The rest of this paper is organized as follows: Section 2 discusses the related work. Section 3 describes the area macro-model. Power macro-modeling approach is presented in Section 4. Experimental results are shown in Section 5 and Section 6 gives the conclusion.

2. Related Work

Vootukuru [3] has presented a method for calculating FPGA CLB resources for all possible functional components and for all possible bit widths, and maintaining a database. Such an approach would result in a huge database and is hence not useful for use with a HLS framework. A fast mapping heuristic method is proposed in [4]. General formulas based on the bit-width and the number of inputs are derived in [5] for area estimation. However, routing and flip-flop costs are not taken into account. Although [6] considers both datapath and control logic, it uses a table to store the operator area for different bits instead of a general function. All of them estimate the logic synthesis area instead of the post layout area after place and route and thus result in larger deviation for the real circuit. In our method, the estimated area is based on the post layout area and can be calculated by a general formula depending on the operator kind and the input bit width.

Macro-modeling for high-level power estimation has attracted considerable attention over the past few years. A look-up table (LUT) model was introduced in [7]. The parameters of that model were the average input signal probability P_{in} , the average input transition density D_{in} and the average output transition density D_{out} . The metric D_{out} was evaluated using zero-delay simulation. The LUT reported estimates for equally spaced discrete values of

the parameters. For input characteristics that did not correspond to any LUT values, estimates were obtained using an interpolation scheme. Several modifications to the LUT method were introduced in [8] to improve its accuracy. The LUT-based approaches are general and can be easily used for different kinds of circuits without any modification of the model itself. However, a large LUT may be required to ensure good accuracy and the LUT schemes do not consider input correlations which have been found to severely influence the overall power consumption [9].

An equation-based macro-modeling technique is proposed in [10]. This model uses an equation to represent the entries of an LUT and thus avoids its potentially large space requirements. In addition to P_{in} and D_{in} , a spatial correlation metric S_{in} and a temporal correlation metric T_{in} are included as the input parameters of the equation-based macro-model. This macro-model achieves better accuracy than the previous ones. However, it requires too many coefficients.

All these approaches discussed above are limited to combinational circuits. Another power macro-modeling approach for both combinational and sequential circuits is introduced in [11]. In addition to P_{in} , D_{in} and D_{out} , this model also uses correlation coefficient S_{in} to account for spatial correlation of the inputs. However, as [7], zero-delay estimation is required during the estimation phase to obtain D_{out} . Moreover, only zero delay output activity (not real delay output activity) is computed.

3. Area Macro-model

Our area estimation methods are based on high-level compile time estimation of the areas of the CDFG nodes. Each CDFG node represents an operator and will be parameterized with the bit-widths of the inputs (such as N-bit adders and multipliers). Since the variables crossing clock boundaries or state boundaries in a finite state machine (FSM) are mapped onto registers, we must also consider the operators with output registered.

To estimate the area of these CDFG nodes, we first take various RTL level arithmetic operators such as +, -, *, logical operators such as OR, AND and comparison operators such as <, <=, and write RTL VHDL code containing only one RTL operation with various bit widths. We then synthesize these RTL operations using a logic synthesis tool and map them onto a Xilinx VirtexII device [1]. Unlike the other area estimation methods, we use the post layout area gained from Xilinx Place and Route tool instead of the one gained by logic synthesis tool. This is because the LUTs utilized by route-through are also counted in the post layout design while they are not counted during logic synthesis. We can even use the number of total equivalent gates as more accurate area value since it also includes registers' area. Finally, a polynomial regression analysis using least squares fit is performed for each operator based on the area values of different bit widths. The general area characterization for each kind of operators can be categorized into three groups where y is the estimated gate-level area as a polynomial function of bit-width N .

Constant: $y = C$

These nodes synthesize as signals and provide an interface with the memory and other devices outside the FPGA chip. Examples include INPUT and OUTPUT nodes.

Linear: $y = p_0 + p_1 * N$

Most of the arithmetic and logical nodes belong to this category. Examples include signed addition and logical AND operations.

Quadratic: $y = p_0 + p_1 * N + p_2 * N^2$

The multiplication operation belongs to this category. Since the Xilinx VirtexII contains built-in 18X18 multipliers, the area of the multiplication operation with input width no more than 18 bits will consume the same area if this built in block is utilized. For all the other situations, it will generate very accurate area estimation.

4. Power Macro-model

Our power macro-modeling method is equation-based and has the following characteristics: (1) Each kind of operator has a uniform template with low number of coefficients; (2) Glitches in the operators are taken into account; (3) It is easy to analyze the impact of different statistic parameters on the power; (4) It can be used for both combinational and sequential circuits.

4.1 Macro-model Metrics

The key challenges in equation-based macro-modeling are the choice of appropriate input parameters (or metrics) and the derivation of the macro-model functions. To obtain accurate and efficient estimation, these metrics should capture the features primarily responsible for the power dissipation and output statistics.

In this paper, we choose the average input signal probability P_{in} , the average input transition density D_{in} and the average input spatial correlation S_{in} as the candidates of input metrics. Input bit width N is also taken into account.

4.2 Macro-model Characterization

The equation-based macro-model used to estimate the average dissipation power can be described as

$$P_{out} = f(P_{in}, D_{in}, S_{in}) \quad (1)$$

where f is a mapping procedure waiting to be determined during the characterization.

In order to get the function f for each kind of the operators, we must have a set of sample input vector streams with different P_{in} , D_{in} and S_{in} and the power dissipation P_{out} of a given circuit based on these input signals. Figure 1 shows the procedure used to get the macro-model characterization.

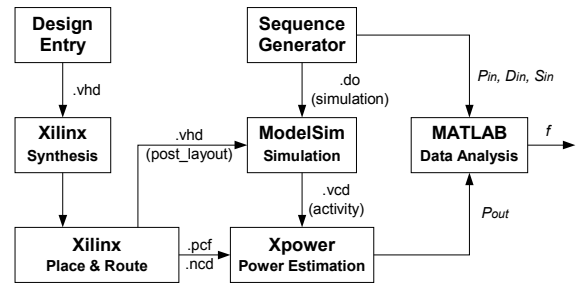


Figure 1: Macro-model Characterization Procedure

First, a VHDL file is written as the design entry containing only the operator to be characterized. Through Xilinx Synthesis and Place & Route tools, the post layout netlist of the operator is generated to obtain the interconnect information. This post layout netlist is simulated using ModelSim simulation tool to get the

switching activities based on the input vector streams. The file containing the switching activities is then sent to the Xilinx Xpower tool along with the placed and routed design to get both the quiescent and dynamic power estimation. Finally, these power estimates together with the corresponding P_{in} , D_{in} and S_{in} are analyzed by MATLAB regression and curve fitting tools to get the estimation function for this operator.

The placed and routed design contains almost all the information of the physical circuit such as the position of the elements, the route of the connections and the delay between different internal signals. Thus the simulation results for such a design are similar with the physical circuit and may contain hazards or glitches. It is definitely much more “real” compared with the other macro-model methods using zero delay switching activity. Moreover, this macro-model can be applied for both combinational and sequential circuits since the information on the state bits is also recorded in the post layout design.

In [10], a random number generator is chosen to generate input sequence. Since there is no constraint for P_{in} , D_{in} and S_{in} , some sequences may have the same P_{in} , D_{in} and S_{in} and some values of P_{in} , D_{in} and S_{in} may never be generated. Moreover, it is very hard to analyze the impact on power of only one sole parameter. In this paper, we use a sequence generator similar as [12]. By given values of P_{in} , D_{in} and S_{in} together with the input width and sequence length, this generator can generate binary input sequences with high accuracy and randomness with more than 99% coverage space. It is much easier for us to analyze the power sensitivity to different parameters by changing one parameter over the entire space while keeping all the other parameters fixed.

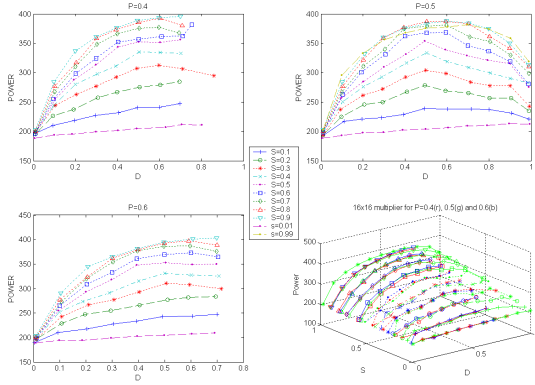


Figure 2: Power of 16x16 bit multiplier on Xilinx Virtex2 FPGA with respect to P_{in} , D_{in} and S_{in}

To make the macro-modeling function efficient and concise, we first analyzed the weight of the impact of different parameters on the average power. With the help of the sequence generator, we generated the input sequences for different input widths with the values of the three parameters in the range of [0.01, 0.99] and with the granularity of 0.1. The length for each input sequence was 4000 and the total number of these sequences was 354 due to the constraints described in [7]. Figure 2 shows the power characterizations for an example 16x16 bit multiplier on the Xilinx VirtexII FPGA (the graph only shows P_{in} with 0.4, 0.5 and 0.6 respectively). From the graph we can see, both D_{in} and S_{in} have almost equal impact on the power, while P_{in} has very little influence. Using a similar method, we also checked the AND and

ADD operators. For the ADD operator, the smaller the value of S_{in} , the more linear was the power with D_{in} . S_{in} has a smaller influence on power compared to D_{in} while P_{in} almost has no impact on power. For the AND operator, the power is linear with D_{in} and both S_{in} and P_{in} do not affect power.

For all the tested designs, P_{in} almost has no impact on power consumption. This is because most of the FPGA devices are CMOS circuits. Only when the input of a CMOS circuit is changed, will there be charge or discharge in the circuit, and will dynamic power be consumed. Although P_{in} is the average input probability of 1 and thus does contain some information about the switching activity (there can not be many switching activities when P_{in} is too small or too large), almost all the information necessary for dynamic power is already included in D_{in} and S_{in} .

Although we can find the same function format for the same kind of operators, we still need to find the values of the coefficients for those with different input bit widths. This is really tedious and inefficient. Our experiments show that we can add input bit width N as one more parameter so that each kind of operators with different input width can have only one uniform power macro-model. Thus, function (1) can be rewritten as

$$P_{out} = f(N, D_{in}, S_{in}) \quad (2)$$

where P_{in} is replaced by N and S_{in} is optional depending on the selection of the operator.

5. Results

In this section, we present the results by applying our area and power macro-models on several operators with different bit widths. The Xilinx ISE 5.2 Synthesis and Place and Route tools were used to get the reference values of area and delay from the post layout of the design. ModelSim simulation tool and Xpower tool were used to get both the quiescent and dynamic power. The target device was the Xilinx VirtexII XC2V250.

5.1 Area and Delay Results

Table 1 shows the accuracy of the area estimation on AND, ADD, and multiplication operators using the general functions described in Section 3. All the operators are signed combinational operators with output registers. Since the area occupied by registers is not included in the LUT numbers, the total equivalent gates count is also taken into account. The first three operators were synthesized using look up tables, while the last one was synthesized using the built-in 18X18 bit multipliers. The input bit widths used for each of the first three operators are 2, 4, 8, 16, 24 and 32 while those for the last one are 19, 20, 24, 28 and 32 since the operators will have the same area when bit width is no more than 18. In our experiments, the Root-Mean-Square (RMS) relative error is no more than 2.13% and the max absolute relative error is at most 4.12%, which means our macro-models for area estimation are quite accurate.

Table 1: Accuracy of Area and Delay Estimation

| Ops | RMS Error | | Max Error | |
|------|-----------|-------|-----------|-------|
| | LUTs | Gates | LUTs | Gates |
| and | 0.00% | 0.00% | 0.00% | 0.00% |
| add | 2.13% | 0.68% | 4.12% | 1.31% |
| mul | 0.22% | 0.19% | 0.36% | 0.32% |
| mul* | 0.44% | 0.00% | 0.72% | 0.00% |

5.2 Power Macro-Modeling Results

Table 2 shows the accuracy results of the equation-based power macro-model (2). The operators in this table are AND, ADD and Multiplier. L, Q and C represent Linear, Quadratic and Cubic functions and N, D and S represent the parameters used in these functions. The max error (ME) between the macro-modeling power P_{mac} estimated by the polynomial function and the measured power P_{out} gained from Xpower is used as reference while the RMS error is considered as the main measurement of the efficiency of these functions. The items with number in italic and bold represent the selected functions.

For multiplication operator, 4.28% RMS relative error and 33.24% max absolute relative error can be achieved by a cubic function with all three parameters. For the ADD operator, a cubic function can get 1.74% RMS error and 7.5% max error with three parameters (20 coefficients) and 3.4% RMS error and 19.84% max error with only two parameters of N and D_{in} (10 coefficients). For the AND operator, a quadratic function with two parameters of N and D_{in} is accurate enough to bring both the RMS and max error below 0.1%. There are at most 20 coefficients for all these functions while 6 coefficients are enough for some of them. This is really concise and efficient compared with those power macro-models reported in other papers which have 35 coefficients.

Table 2: Accuracy of equation-based power macro-model with D_{in} , S_{in} and N as parameter

| Ops | Parameters | RMS Error | | | Max Error | | |
|-----|------------|-----------|--------------|--------------|-----------|--------------|---------------|
| | | L | Q | C | L | Q | C |
| mul | N | 23.48% | 23.42% | 23.42% | 102.04% | 98.08% | 98.08% |
| | D | 24.36% | 23.26% | 23.25% | 128.23% | 123.94% | 124.43% |
| | S | 23.08% | 23.01% | 23.00% | 120.45% | 121.44% | 121.84% |
| | ND | 20.21% | 16.61% | 16.13% | 88.02% | 68.06% | 73.89% |
| | DS | 21.32% | 19.92% | 19.59% | 113.33% | 107.81% | 106.09% |
| | NS | 18.64% | 17.37% | 17.29% | 78.85% | 106.27% | 97.63% |
| | NDS | 15.40% | 7.73% | 4.28% | 66.57% | 48.42% | 33.24% |
| add | N | 24.74% | 24.74% | 24.74% | 90.95% | 90.81% | 90.81% |
| | D | 20.12% | 19.93% | 19.90% | 62.68% | 71.96% | 68.15% |
| | S | 28.73% | 28.73% | 28.54% | 113.22% | 113.22% | 108.95% |
| | ND | 12.92% | 3.68% | 3.40% | 40.98% | 21.97% | 19.84% |
| | DS | 20.10% | 19.84% | 19.64% | 63.50% | 69.58% | 60.84% |
| | NS | 24.53% | 24.47% | 24.33% | 93.08% | 94.67% | 91.68% |
| | NDS | 12.89% | 3.23% | 1.74% | 41.87% | 18.54% | 7.50% |
| and | N | 26.99% | 26.99% | 26.99% | 98.41% | 98.17% | 98.17% |
| | D | 19.67% | 19.67% | 19.66% | 62.57% | 64.06% | 63.32% |
| | S | 30.38% | 30.36% | 29.98% | 118.30% | 119.31% | 114.22% |
| | ND | 13.09% | 0.01% | 0.01% | 41.90% | 0.03% | 0.03% |
| | DS | 19.67% | 19.66% | 19.59% | 62.54% | 63.74% | 60.88% |
| | NS | 26.88% | 26.80% | 26.49% | 99.20% | 100.82% | 97.66% |
| | NDS | 13.09% | 0.01% | 0.01% | 41.91% | 0.03% | 0.03% |

6. Conclusion

In this paper, we present an equation-based macro-modeling technique for high-level area and power estimation on FPGAs. Input bit width N is the only parameter used in our area macro-models which can achieve high accuracy. For our power macro-model, the average input signal probability P_{in} is not selected. Instead, we use the input bit width N , together with the average input transition density D_{in} and the average input spatial correlation S_{in} as parameters. Thus each kind of operator needs only one uniform function for different input widths. To get good estimation results, we can use much less coefficients and sometimes lower degree functions compared with the other macro-models. Glitches are also taken into account and this technique can be applied for both combinational and sequential circuits. By accurate control the input simulation sequence with the parameter values we want, we can easily analyze the power sensitivity to each of these parameters. Experimental results show that our macro-models are quite accurate and efficient. These macro-models are being used in the PACT high-level synthesis tool to perform area and power optimizations.

REFERENCES

- [1] Xilinx, Virtex II Datasheet, www.xilinx.com.
- [2] Altera, Stratix Datasheet, www.altera.com.
- [3] M. Vootukuru, R. Vemuri and N. Kumar, Partitioning of Register Level Designs for Multi-FPGA Synthesis.
- [4] M. Xu and F. J. Kurdahi, Area and Timing Estimation for Lookup Table Based FPGAs, *Technical Report # 9530*, UCI, Aug. 1995.
- [5] D. Kulkarni, W. Najjar, R. Rinker, F. Kurdahi, Fast Area Estimation to Support Compiler Optimizations in FPGA-Based Reconfigurable Systems, *Proc. 10th Annual Symp. On Field Programmable Custom Computing Machines (FCCM 2002)*, Napa, CA, Apr. 2002.
- [6] A. Nayak, M. Haldar, A. Choudhary, and P. Banerjee, Accurate Area and Delay Estimators for FPGAs, *Design Automation & Test in Europe*, Paris, March 2002.
- [7] S. Gupta, F.N. Najm, Power Macromodeling for High Level Power Estimation, *Proc. 34th Design Automation Conf.*, June 1997.
- [8] M. Barocci, L. Benini, A. Bogliolo, B. Ricc'ò, G. De Micheli, Lookup Table Power Macro-Models for Behavioral Library Components, *Proc. IEEE Alessandro Volta Workshop on Low Power Design*, Mar. 1999.
- [9] E. D. Kyriakis-Bitaros, S. Nikolaidis, A. Tatsaki, Accurate Calculation of Bit-Level Transition Activity Using Word-Level Statistics and Entropy Function, *Digest Tech. Papers International Conf. on Computer-Aided Design*, Nov. 1998.
- [10] G. Bernacchia and M. Papaefthymiou, Analytical Macromodeling for High-level Power Estimation, *Proc. IEEE ICCAD*, Nov. 1999.
- [11] S. Gupta, F.N. Najm, Analytical Model for High Level Power Modeling of Combinational and Sequential Circuits, *Proc. IEEE Alessandro Volta Workshop on Low Power Design*, Mar. 1999.
- [12] X. Liu, M. C. Papaefthymiou, A Markov Chain Sequence Generator for Power Macromodeling, *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pp. 404-411, 2002.