Power Modeling for RTL Design with RNN Training on Switching Frequency

Xingyu Li
Department of Electrical Engineering and Computer Science
University of California, Berkeley
xingyuli9961@berkeley.edu

***Abstract*—This paper presents a tentative methodology that fast estimates the switching frequency behaviour of signals in the RTL designs in the software simulation time. First it uses VCD dump to extract the signal activity and train the RNN model and then estimates the switching frequency of different signal for any given sequence of functions. Previously, there are proposed methodology that needs cycle-by-cycle simulation or provides power estimation on a larger scale. On contrary, this project intends to find a detailed signal switching frequency pattern after training and give a convenient estimation during software simulation, without the aid of FPGA. This method is demonstrated with simple designs such as Riscv-mini, and can be also used for (detailed estimation for parts of) more complex designs including the RocketChip and BOOM.**

I. Introduction

Power efficiency has become an important constraint for both systems remaining low power consumption and chips reaching high performance without breaking thermal limitations. As Moore's law is likely to slow down and become less guaranteed in the future, designing circuits that meet the power requirement will impose great challenges. Then, fast and accurate power estimation method in different steps of the design process will be a necessity for designers to modify the circuits timely and achieve their goals.

Previously, there are several micro-architecture-level power analysis tools employed by computer architects, such as [1], [2], [3]. These methods conveniently allow researchers to do analysis and optimization without developing RTL design. However, these methods needs to use existing hardware simulator such as [4] and [5] and, as a result, limits the design to be compatible to the simulators and can hardly support novel, non-traditional designs. Also, they also require long simulation time to collect all the architectural activities.

An alternate option is using commercial CAD tools to get accurate power evaluation for the RTL implementation. In order to cope with long simulation time, several methodology [6], [7] and [8] train macro power models between power and selected signals, and use FPGA emulation as a faster way to collect samples of signal activities.

Inspired by both types of methodology mentioned above, this projects demonstrates a trial to model signal switching frequency, a major parameter for the dynamic power, which can be later used for further power estimation. First, signal activities are extracted from the VCD dump and a design-specific switching frequency model is constructed and trained with respect to the instruction codes with recurrent neural network. After the model is built, switching frequency behaviour can be estimation for any give sequence of instructions.

II. Related Work

In the architectural level, using cycle-accurate hardware simulator like [4] and [5] to collect data and employing methodology such as [1], [2], [3] to build model for power estimation is widely used. The Aladdin [3] is a pre-RTL, power performance simulator for accelerator-centric systems. It takes in algorithms that describe accelerators and uses dynamic data dependence graph to model the signal activities. With the signal flow is correctly analyzed, it can provide a cycle-level power estimation. However, Aladdin and methods alike remain in the architectural level and can not reflect the power difference between different implementation of the RTL, especially for those specially customized circuits.

Another widely used method is to use hardware monitoring counters [9], [10]. For an existing physical system with these counters, it can provide a fast power estimation. However, for new designs, these methods have such a high overhead that made them infeasible.

There are also interesting methods [6], [7] and [8] that automatically builds macro power model on signal activities and gives runtime estimation with the aid of the FPGA emulation. Simmani [8] exploits the observation that signals can be clusters by their switching pattern and the signals in each pattern have similar effects on the dynamic power dissipation, and it cluster signals from the VCD dumps from RTL simulation. Then it trains the selected signals against cycle-accurate power traces from commercial CAD tool to build regression models. It then automatically adds counters in the FPGA design and the software can continuously pull information from the counter and provide a relatively instant power estimation. Simmani focuses more on the general power consumption of a complex design running applications, but it might not be help if more detailed, cycle-level information is needed.

Almost none of the methods provides a fast cycle-level power estimation in the RTL level. This project proposes to build a macro model for the signal switching activity for the RTL design. If this model can give an instant and relatively accurate switching frequency pattern for any given sequence

of instructions, then the designers can speed up their progress. And, this is the motivation for the project.

III. Methodology

There are three steps in the methodology. First, decoding step transfers instructions into simplified and standardized form. Then, the modeling and training step construct a macro model of the switching frequency using the VCD dumps of sufficient workload. Finally, we can test the model with a sequence of instructions and compare the estimated switching frequency with the actual output from the VCD dumps of the RTL simulation.

A. Decoding

For an arbitrary functional circuit block, we can concatenate the meaningful part of all its inputs and form a sequence of standard "instructions". For instructions using standard ISA, the decoding stage might be optional. However, as instructions generally contains useless information, we can automatically transform the original codes into more concise format, which would probably make the training process easier.

B. Modeling and Training

First I compile the RTL design in the Synopsys VCS to further run applications on the design and derive the signal activities. Then, I need to run simulations of micro-benchmarks and some random instruction streams are for initial model training. As the inputs are a temporal sequence, I choose to use RNN in pytorch to train the model.

C. Prediction and Validation

After the model is built, we can use the model to predict the estimated switching times during the executions of the instructions. With the frequency of the clock provided, it can calculate the switching frequencies. Finally, we compare the estimated results with the actually VCD dumps from the RTL simulation, and make some modifications if necessary.

IV. Experimental Setup

A. Riscv-mini

B. Single Core RocketChip

V. Conclusion and future work

Acknowledgment

References

[1] D. Brooks, V. Tiwari, and M. Martonosi. 2000. "Wattch: a framework for architectural-level power analysis and optimizations," SIGARCH Comput. Archit. News 28, 2 (May 2000), 83–94.

[2] S. Li, J. H. Ahn, R. Strong, J. Brockman, D. Tullsen, and N. Jouppi,"McPAT: An integrated power, area, and timing modeling framework for multicore and manycore architectures," in MICRO, 2009.

[3] Y. S. Shao, B. Reagen, G.-Y. Wei, and D. Brooks, "Aladdin: A pre-RTL, power-performance accelerator simulator enabling large design space exploration of customized architectures," in ISCA, 2014.

[4] T. M. Austin Ph.D. and SimpleScalar, LLC, SimpleScalar tool set.

[5] N. Binkert, S. Sardashti, R. Sen, K. Sewell, M. Shoaib, N. Vaish, M. D. Hill, D. A. Wood, B. Beckmann, G. Black, S. K. Reinhardt, A. Saidi, A. Basu, J. Hestness, D. R. Hower, and T. Krishna, "The gem5 simulator," ACM SIGARCH Computer Architecture News, vol. 39, Aug 2011.

[6] D. Kim, V. Iyer, "Automatic Activity-based Power Modeling for Arbitrary RTL"

[7] D. Kim, A. Izraelevitz, C. Celio, H. Kim, B. Zimmer, Y. Lee, J. Bachrach, K. Asanovic, "Strober: Fast and Accurate Sample-Based Energy Simulation for Arbitrary RTL," 2016 ACM/IEEE 43rd Annual International Symposium on Computer Architecture (ISCA), Seoul, 2016, pp. 128-139.

[8] D. Kim, J. Zhao, J. Bachrach, and K. Asanović. 2019. "Simmani: Runtime Power Modeling for Arbitrary RTL with Automatic Signal Selection, " in Proceedings of the 52nd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO '52). Association for Computing Machinery, New York, NY, USA, 1050–1062.

[9] T. Li and L. K. John, "Run-time modeling and estimation of operating system power consumption," in SIGMETRICS, 2003.

[10] T. F. T. Wenisch, R. R. E. R. Wunderlich, M. Ferdman, A. Ailamaki, B. Falsafi, and J. C. J. Hoe, "SimFlex: Statistical Sampling of Computer System Simulation," IEEE Micro, vol. 26, pp. 18–31, Jul 2006.

[11] I. Goodfellow, Y. Bengio, A. Courville, "Deep Learning", MIT Press, 2016