A Cycle Accurate Power Estimation Tool

Rajat Chaudhry
STI Design Center
IBM Corporation
Austin, TX 78758
Tel : 1 512-838-8794
Fax : 1 512-838-2132
rajat@us.ibm.com

Daniel Stasiak
STI Design Center
IBM Corporation
Austin, TX 78758
Tel : 1 512-838-1952
Fax : 1 512-838-2132
stasiak@us.ibm.com

Stephen Posluszny
STI Design Center
IBM Corporation
Austin, TX 78758
Tel : 1 512-838-6508
Fax : 1 512-838-2131
Stephen.Posluszny@us.ibm.com

Sang Dhong
STI Design Center
IBM Corporation
Austin, TX 78758
Tel : 1 512-838-0106
Fax : 1 512-838-1272
dhong@us.ibm.com

Abstract - Power consumption is one of the major challenges in VLSI Design. Power constrained designs need tools to accurately predict the power consumption and provide feedback to designers on the efficiency of the power management logic. In this paper we present the methodology behind a cycle accurate power estimation tool. This tool was used to estimate the power of a first generation CELL Processor. The tool extracts switching and clock activity from RTL simulations and applies them to transistor level macro power models to calculate the power for every cycle of the simulation trace.

1. Introduction

Power Consumption is one of biggest challenges in VLSI design. In the past, power has mainly been a concern for chips used in battery-powered devices. But due to the continuous increase in frequency and increase in leakage current due to scaling, it is becoming a constraint even in wall socket powered devices. Packaging and thermal cooling costs are the biggest drivers for reducing power in such chips, especially chips that are manufactured in large quantities for price sensitive products.

Due to the tight power constraints, chip designers require power estimation tools that can provide them with accurate power estimates and timely feedback on the efficacy of their power management design. As the margin of error in power budgets reduces, the need for dynamic power estimation increases. Not only is dynamic power estimation required to verify power management logic, but also because static estimates are too pessimistic. It is important to concentrate on realistic high power test cases. Pathologically high power workloads can be handled by hardware based thermal management solutions, by reducing the frequency or shutting down the system.

2. Previous Work

Power estimation has been done at different level of abstraction. The Most accurate estimates are done by running a SPICE like simulator on a transistor level netlist. But this is feasible only for a small circuit. Many authors have worked on building accurate macro power models [1]. Macro power models are not very useful in estimating total chip power unless we have accurate information on the switching and clocking activity of the macros. A lot of work has been also been done at measuring power at the architectural level [2]. Although architectural level solutions are very fast, they are not very accurate and cannot give feedback on clock gating at the RTL level. In ASICs, gate level power estimation methodology has been employed. This does not work well for chips with a lot of custom blocks.

In the first generation CELL Processor (the first generation CELL processor will be referred to as the Processor for the rest of this paper), designers have employed a lot of fine-grained clock gating. We needed a tool that could give accurate power estimates as well as provide information on the power tradeoff of adding clock gating logic. The processor consists of custom blocks and synthesized logic [4]. In our cycle accurate power estimation tool (which will be referred to as CAPET in this paper) we combine the benefit of accurate transistor level macro power models coupled with switching and clocking activity from RTL level simulations. This is essential, because gate level techniques cannot accurately model power for custom designed blocks. We also model the power consumed due to the switching of signal interconnect capacitance.

3. CAPET Approach

The biggest components of the power dissipation of a circuit consists of 1) the leakage current, which depends on the state of the transistors and 2) the switching power which consists of the power dissipated by the switching of node capacitances due to circuit activity. CAPET estimates switching power, so we will limit our discussion to switching power.

The switching power of a circuit can be calculated by the function

$$P = \tfrac{1}{2} CV^2 f. \quad (1)$$

0-7803-9451-8/06/$20.00 ©2006 IEEE.

Where *C* is the total node Capacitance switched, *V* is the power supply voltage and *f* is the frequency of switching. For a circuit design, *V* and *f* are generally fixed.

To reduce the power consumption, designers have to work on reducing the switched capacitance. In a power efficient processor like our Processor, designers employ intensive amount of fine-grained clock gating to reduce power. This reduces the switched capacitance by limiting the amount of switching in the latches and unnecessary switching of combinational logic. Therefore the key determinants of power estimation are monitoring the switching and clocking activity in a design.

The CAPET methodology consists of monitoring the switching and clock activity of each macro in an RTL simulation and then applying that information to a transistor level macro model to estimate the power. This is done for each clock cycle to get a cycle-by-cycle power estimate. The power consumed by switching interconnect capacitance is also estimated by monitoring the switching of each global net. The gate capacitance of buffers on the net is included as part of interconnect capacitance.

The minimum requirements for running CAPET are a functional VHDL model and a chip floorplan. In the early phase of design, macro power models can be roughly estimated using macro area and net capacitance is calculated by Steiner estimates. This way CAPET can be run early in the design for designer feedback and the estimates get more refined as the design progresses as shown in figure 1.

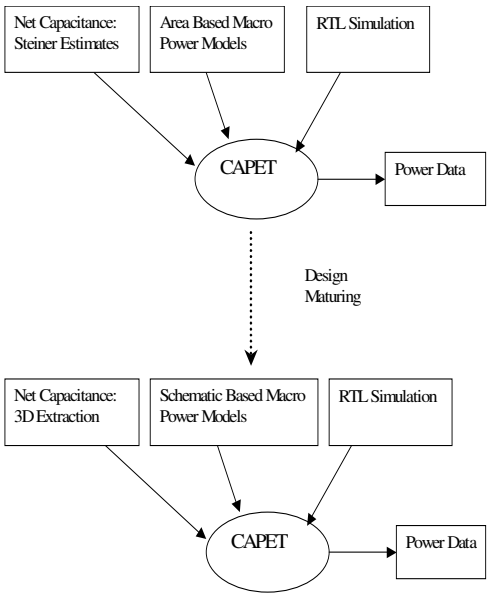

Figure 1: CAPET flow at different stages of design

3.1 Macro Power Models

The node switching in a circuit block in a given cycle is proportional to the percent of inputs switching and the amount of clock activity in the circuit. Switching factor of a circuit is defined as the percent of inputs changing state between two consecutive clock cycles. Clock activity is defined as the percent of capacitive load being driven in a given cycle with respect to the total clock load in the circuit. We build our macro power models at the transistor level using IBM's CPAM tool [3]. CPAM runs random vectors with different Switching Factors on the schematic under two conditions 1) with all local clock buffers turned off and 2) all local clock buffers turned on. The power model assumes that power is linear with switching factor and clock activity. CPAM provides power information at simulated switching factors

If for a given clock cycle, SF is the Switching factor of a circuit block and CLK is the clock activity of the circuit then the Power consumed by the circuit in a given cycle C is

$$P(C) = P_{clk0}(SF) + (P_{clk100}(SF) - P_{clk0}(SF)) * CLK.$$

Where

$P_{clk0}(SF)$ is the power at input switching factor SF when clock activity is 0%

$P_{clk100}(SF)$ is the power at input switching factor SF when clock activity is 100%

The 2 curves P_{clk0} and P_{clk100} are shown in figure 2.

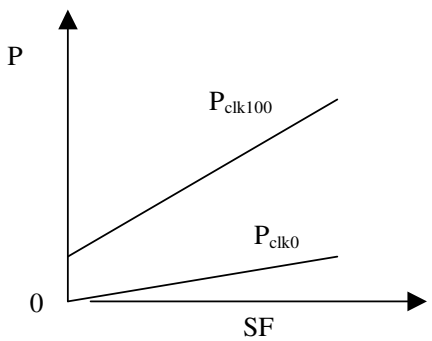

Figure 2: Power dependence on Switching Factor at 100% clock activity and 0% Clock Activity

Currently we simulate macros only at a SF of 0% and 50%; therefore our current models are purely linear with respect to SF. The accuracy of the power model can be improved by simulating at more number of Switching factors. In the early part of the design phase the 2 power curves P_{clk0} and P_{clk100} are estimated using the area of the macro by using a power density derived by scaling from previous technology or from similar macros whose schematics are complete.

3.2 Global Signal Net Power

The macro power models characterize the power switching power inside macros. To estimate the power due to switching of global interconnect capacitance, we calculate the interconnect capacitance using Steiner estimates in the early part of the design phase and use a 3D extraction tool in the later part of the design. The power of switching global nets can be calculated using equation (1). The power of the buffers inserted on the global nets is estimated by adding the gate capacitance of the buffers to the net capacitance. This approach does neglect the shoot through current from VDD to GND while the buffer is switching. Our experiments show that this shoot through current is a negligible part of switching power.

3.3 Measuring Switching Factors and Clock Activity

To calculate the total power, Input switching factors and clock activity for each block instance are monitored in an RTL level simulation of a given workload. For every macro instance the input Switching Factor is calculated by observing the percent of inputs that have changed state from the previous cycle. The Clock Activity is measured by observing the number of clock buffers that are turned on in a given cycle.

Since power is dependent on both Switching factor and clock activity, it needs to be calculated for every cycle of the simulation. The switching factor and clock activity cannot be averaged over the entire simulation.

In a given cycle the total power is calculated as

$$TotalPower(C) = \Sigma BlockPower(SF, CLK) + \tfrac{1}{2} C_{net}(C)V^2 f$$

Where C_{net} (C) is the amount of global net capacitance switched in the given cycle.

The measure of clock activity is dependent on the type of macro. Since each local clock buffer drives different amount of load, they cannot be treated as equal. For custom macros the designers provide a table, which puts relative weights on each local clock buffer. In case of synthesized blocks the relative weights are calculated using the number of latch bits that are driven by each local clock buffer.

4. CAPET Usage

CAPET methodology was used to estimate the power consumption of the first generation CELL processor and for refining and verifying the power management logic of the chip.

4.1 RTL Workloads

Each core or functional unit on the chip was required to run at least 3 different type of workloads, 1) idle, 2) typical and 3) high power. The idle workload is very useful in making sure that when the core is supposed to be at the lowest power state, it shuts off as many clock buffers as possible. Analyzing the results of the idle test case is very useful in catching the most obvious errors. After running CAPET, designers look at the macros that have the highest power and highest clock activity and try to reduce them. Some of the cores also used the tool to verify the power savings by issuing instructions at a slower rate.

4.2 Power Grid and Thermal Analysis

We used the results of the high power test case as a stimulus for Power Grid integrity analysis and thermal analysis. Each core is assigned the average power for the high power workload for IR drop analysis. For *di/dt* analysis on the package the cycle-by-cycle power for the high power workload was used for each core. The ability to have a cycle-by-cycle power for each macro for a realistic workload makes CAPET very useful for package analysis. For thermal analysis the average power for the high power workloads are used to calculate the temperature map for the chip.

4.3 Full Chip Estimates

Since RTL simulation is not fast enough to run a complete program at the chip level, the chip level estimates were calculated by getting utilization information on the different cores from architectural simulation. Based on realistic high power programs, utilization rates for the 3 different work loads, idle, typical and high power were assigned to each core and this was aggregated up to the chip level.

5. Results and Conclusion

In this section we present the results of CAPET on one of the cores of the Processor. Figure 4 shows the power waveforms for the stop, typical and high power test cases. Figure 3 shows the macro power and net power of typical power workload. Figure 5 shows the number of clock buffers active for the 3 different workloads. The runtime for CAPET is similar to the runtime for an RTL simulation. The runtime for 4000 clock cycles on a core with 20.9 million transistors is approximately 30 minutes.

We have correlated the results of CAPET with hardware and the estimates are within plus or minus 10%. Multiple chips were carefully run under the exact same conditions as simulation to correlate the results.

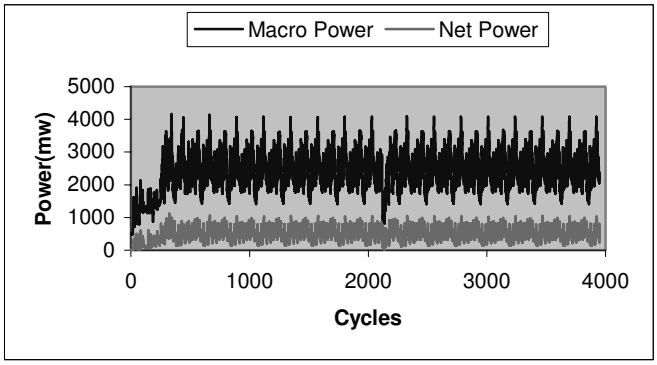

Figure 3: Macro Power and Net Power for typical power workload

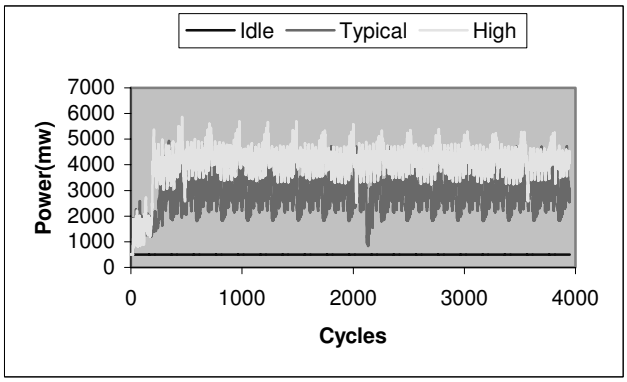

Figure 4: Total Power for Idle, Typical and High power workloads

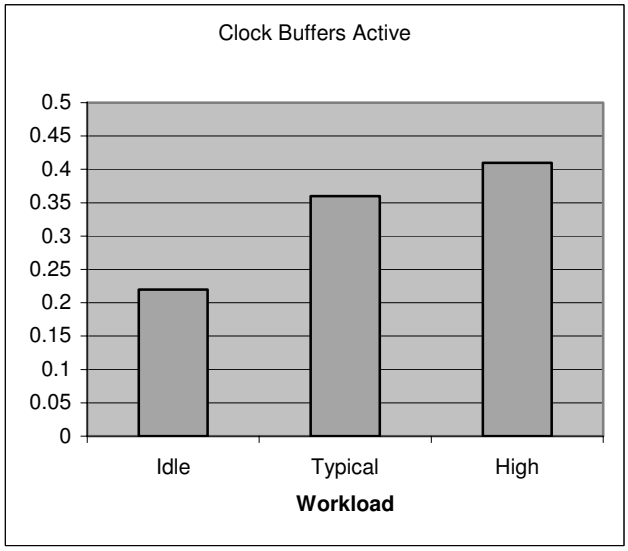

Figure 5: Percent of local clock buffers active for idle, typical and high power workload

We have presented a power estimation methodology that helps in providing accurate power estimates and can provide designers with feedback on the clock gating logic. The hardware measurements show that our methodology is accurate. We are working on improving our macro models by measuring power of macros at multiple switching factors and making power a piece wise linear function with respect to switching factor.

.

6. REFERENCES

[1] Subodh Gupta and Farid N Najim, *Power Modeling for High Level Power Estimation*, IEEE transaction on VLSI, vol 8. No 1, Feb 2000.

[2] Brooks et al., *Power-aware microarchitecture: design and modeling challenges for next-generation microprocessors,* Micro, IEEE, Volume: 20, Issue: 6, Nov.-Dec. 2000 Pages: 26 - 44

[3] Neely, J.S et al., ***CPAM: a common power analysis methodology for high-performance VLSI design,*** Electrical Performance of Electronic Packaging, 2000, IEEE Conference on. , 23-25 Oct. 2000

[4] E Behnen et al, *Ontology of The First Generation CELL Microprocessor*, Submitted to DAC 2005.