Power Optimization Techniques for High Speed Processor Core in Sub 14nm Technology Node

Kolusu Siva Mounica, Prashant K. Shah

[1]*SVNIT, Surat, Gujarat, India*

[2]*Associate Professor, ECED-SVNIT*

Abstract

Optimization of power can be done at different levels of abstraction e.g., system level, RTL level, Circuit level, Layout level. With continuous scaling of the technology node optimization of power and overall power management on SoC are the key challenges in addition to meeting the performance requirements. This paper gives an idea of various techniques at circuit level to reduce power consumption without affecting the performance of the chip.

Keywords —*Scaling, Technology node, SoC (Silicon on Chip), Layout*

I. INTRODUCTION

Power was not a major concern due to the low device density and low operating frequency. Consumer wants smaller and sleeker designs which require high levels of integration on silicon chip in advanced processes [1], which increases device density and operating frequency. There is a need of aggressive scaling to meet these demands. Continuous scaling enables high level of integration but rapidly increases power-density more than expected [2]. Overall power consumption on SoC is dramatically increasing. If the semiconductor integration continues to follow this trend of scaling,, power density inside the chips will reach far higher than the rocket nozzle [3].

In designing of server chips, with the requirements of increased performance the power density of a chip has grown down to the point where is no longer possible to increase the clock speed as technology shrinks. Power density not only presents packaging and cooling costs but also causes reliability problems, since the mean time to failure Decreases exponentially with increase in temperature. In addition with temperature, performance degrades leakage increases. Fig. 1 shows main sources of power consumption in CMOS circuits.

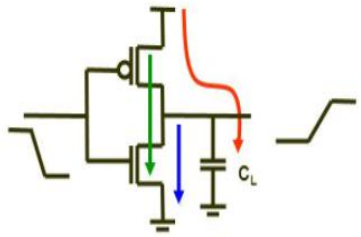

Fig. 1: Sources of Power consumption In CMOS circuits [4]

Though leakage power is increasing enormously with scaling [5], when it comes to full chip power dynamic power dominates. Dynamic power contributes typically 80-85% of full chip power.

Dynamic power consumption depends on supply voltage (V_{DD}), load capacitance (C_L), activity factor (α) and operating frequency (f) [6]. The relation can be expressed as follows

$$P_{Dyn} = C_L V_{DD}^2 \alpha f$$

Reducing V_{DD} has a quadratic effect on P_{Dy} [7-9]. This is one of the reasons that lower supply voltages are becoming more and more attractive. For a design at circuit level as the voltage and frequency are constant, product of load capacitance and activity factor is considered as a measure of dynamic power.

Before concentrating on power reduction techniques, meeting the required performance is the primary concern [10]. Power optimization means reducing the power consumption without affecting the performance of the system.

II. CURRENT DESIGN

Full chip core is divided into smaller blocks to reduce the complexity while designing and easy to analyse the timing, power and layout aspects, designer can easily work to meet the requirements. The leaf cell in hierarchy is called as Functional Block (FB) [3]. At FB level not every circuit has similar type of optimization techniques. Designer

should understand the whole design, what are the power hungry spots in the design.

From observations, at core level 60% of the time 64bits of data is replicated to 192 bits of data and is transferred to the left and right side stacks of the core to perform the required arithmetical and logical operations. Data is being replicated at a distance of 1100um away from where it is actually used. Unnecessarily 192bits are toggling all the time, which contributes to considerable amount of dynamic power.

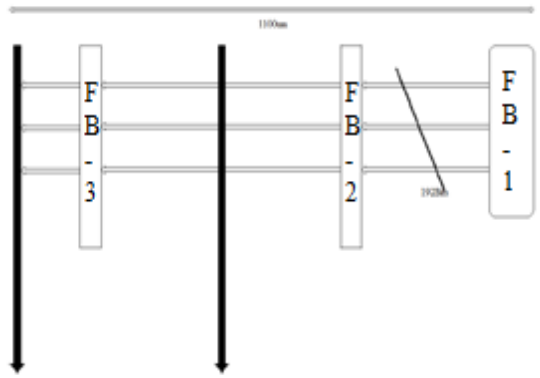

Fig. 2: Replication of 64bits to 192bits

FB-2 is the interface block which is receiving 192bits of data and transfers it to right and left side stacks of the core. Fig. 3 shows the current design of FB-2.

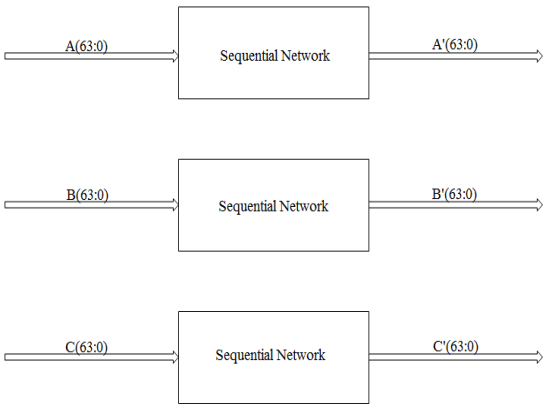

Fig. 3: FB-2 Current Design

To avoid this unnecessary toggling of data, we have proposed a new implementation in which data replication is done at FB-2.

III. PROPOSED DESIGN TO REDUCE THE AF

Selection line of the multiplexer decides whether to replicate the 64bits data to 192bits or directly 192bits of data is to be transferred. If the selection input is 1 then 64bits of data is copied to remaining 128bits of data and processed.

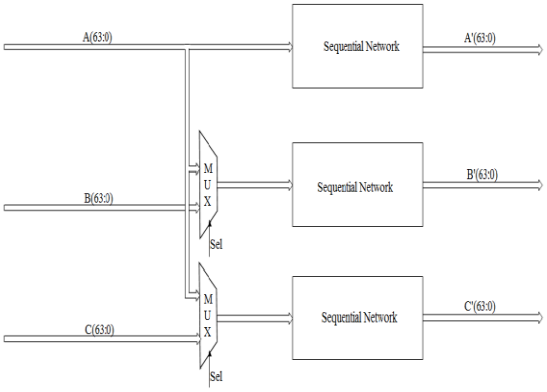

Fig. 4: Proposed Implementation of FB-2

Comparison of power numbers before and after implementation is given in the below table.

Table 1: Power Numbers Before And After The Implementation

Para meter	Before Implementation	After Implementation
AF	0.11	0.11(1-0.68) =0.035
Capacitance	0.14pF	0.14pF
Dynamic Power consumed by 128bits	0.11x0.14x128 =1.971pF	0.035x0.14x128= 0.627pF
Power consumed by mux(Cap =0.004pF /Each)	-	0.11x0.004x128= 0.056pF
Power for 128bits	1.971pF	0.683pF
Total FB Power	2.38+1.971 =4.35pF	2.38+0.683 =3.063pF

29.58% of power save is expected from the proposed implementation.

IV. TECHNIQUES TO REDUCE CAPACITANCE

A .Redundant Buffer Removal:

Though buffers don't serve any logic functionality, they are added in the design to act as repeaters, if the routing length is too large. There

exist some cases where the driver and receiver are adjacent to the buffer. Removing the buffer in such cases helps in reducing the overall capacitance of the FB.

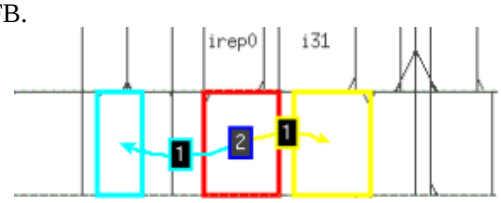

Fig 5: Redundant Buffer Removal

Blue and yellow colour blocks in Fig. 5 represent the driver and receiver of the buffer (Red colour). Removing the buffer in this case helps.

B. Placement and Routing Optimization:

Interconnects are used in the design to connect the devices. Each metal layer has a specific value of resistance and capacitance based on its dimensions. In IC designing, high ROF (R_{line}/R_{eff}) is a quality issue that is to be monitored. To address this problem resistance of interconnect should be reduced, so it is advisable to use higher metal layers for larger route lengths. Observations show that usage of higher metal layers offers less resistance and more capacitance when compared to lower metal layers. Optimizing the placement helps in reducing the routing length thereby limiting the usage of higher metal layers.

Fig. 6: Placement and Routing Optimization

Blue colour in Fig. 6 represents the driver and yellow colour represents receiver, in case 1 of the above shown figure the device is not optimally placed, usage of higher metal layer is optimum to maintain low ROF. Changing the placement helps in reducing the overall capacitance.

C .Merging of Latches:

Merging two single latches into a single dual latch reduces the gate capacitance of the previous stage. In each latch if we observe, at transistor level clock is first given to an inverter and then fed to the pass gate. Merging 2 single latches to a single dual latch reduce one inverter there by reducing the gate cap at the previous stage.

V. TECHNIQUES TO REDUCE THE PRODUCT OF AF AND CAPACITANCE

A. Clock Buffer Merging:

Merging of clock buffers reduces the gate capacitance. As the activity factor of a free clock net is 1, reducing a small amount of capacitance will translates to considerable amount of reduction in dynamic power [11].

In addition to this there might be some cases where free clock is running for larger distance and clock gating is implemented in later stages. Swapping of stages helps here. Swapping of inverter with clock buffer in Fig. 7 reduces the product of activity factor and capacitance value. In addition also helps in meeting the frequency requirements as inverter has better driving capability than clock buffer.

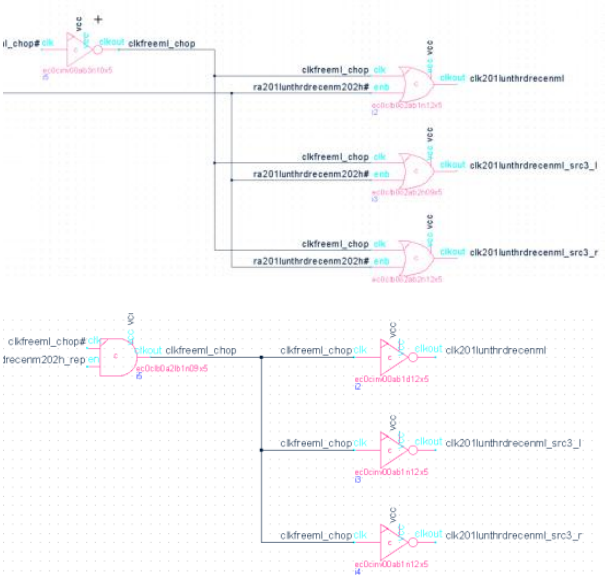

Fig. 7: Implementation clock gating at initial stage

B. Clock Cell Duplication:

Reducing the duplication of clock cells by taking cells out of the template and driving by a single cell outside of template helps in reduction of number of clock cells. To perform similar type of functionality on different data, it is advised to use template based design, which reduces manual effort of designing. Exchanging the buffer with inverter in Fig. 8 reduces a single stage in each template.

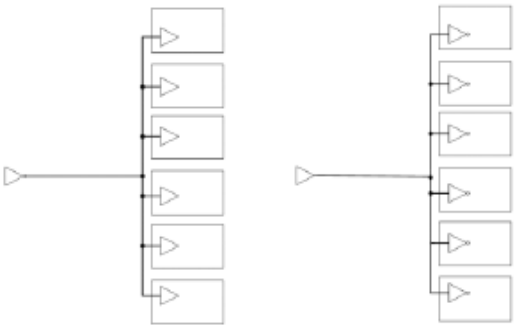

Fig8: Clock Cell Duplication

VI. TECHNIQUES TO REDUCE SHORT CIRCUIT POWER

Short circuit power is the one that flows from VDD to ground [12]. Path between VDD and ground comes when the standard cell is driven with a signal having bad slope. Bad slope means the driver takes long time to charge the load capacitance, this can happen if the drive strength of driver is low or there is a high resistive interconnect.

It is important to either increase the drive strength of driver or to improve the interconnect. Interconnect can be improved by adding strapping's or by upgrading the low level metal layer to high level metal layer, thus reducing the resistance of the interconnect thereby improving the signal slope.

VII. TECHNIQUES TO REDUCE STATIC POWER

In industries many libraries today offer two or three versions of cells: Low VT, Standard VT, and High VT [13-14]. The implementation tools can take advantage of these libraries to optimize timing and power simultaneously. Using High VT cells wherever performance goals allow and low VT cells where necessary to meet timing. Higher the threshold voltage less is the leakage, lower will be the rate of charging and discharging of capacitors. Circuits operate at slower speeds. While in case of low VT cells leakage is more and higher the rate of charging and discharging of capacitors leading to faster operation of circuits.

There is always a trade-off between speed and power consumption of a logic circuit. Usually there is a minimum performance which must be met before optimizing power. In real scenario first the design is synthesized for high performance, high leakage library and then relaxing back any cells not on the critical path by swapping them for their lower performing, lower leakage equivalents. If minimizing leakage is more important than achieving the minimum performance then this process can be done the other way around.

VIII. CHALLENGES

We have modelled the multiplexers of 0.004pF capacitance for the proposed implementation. When it comes to the physical implementation due to the addition of extra stage, performance got degraded by 27.27%. To recover the performance degradation, increased the width of multiplexer's there by increasing the driving capability. Capacitance of each multiplexer increased to 0.007pF from 0.004Pf, power save became 28.62% from an expected value of 29.58%in total power. Comparison of power numbers for different optimization techniques is as shown in Fig. 9

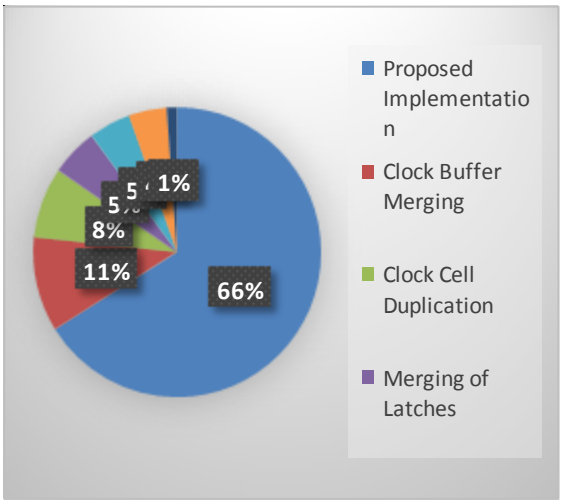

Fig. 9: Comparison of Power Numbers for Different Optimization Techniques

IX. CONCLUSION

The need for lower power systems is being driven by many market segments. Unfortunately designing for low power adds another dimension to the already complex design problem and the design has to be optimized for power as well as performance and area. Before focusing on power optimization, our primary focus is to converge timing to meet the frequency requirements. On one of the functional unit block the methods that are discussed in this paper are applied and results have been analysed. Among all the methods discussed reducing activity factor of data gave the better results. Designer should take care of performance targets while applying the power optimization techniques. All the mentioned optimization techniques are applicable at circuit level of abstraction.

REFERENCES

[1] Jin-Fu Li, "Low Power VLSI Design," [Online]. Availablehttp://www.ee.ncu.edu.tw/~jfli/vlsi21/lecture/ch04.pdf

[2] "Low power vlsi design" [Online].Available: http://www.eeherald.com/section/design-guide/ Low-Power-VLSI-Design.html.

[3] M. law, "Intelprocessors,"http://www.intel.com/museum/ archives/historydocs/ mooreslaw.htm.

[4] "Short Circuit Power" [Online] Available: https://www.google.co.in/search?q=short +circuit+power+Diss ipation&espv=2&biw=1366&bih=638&source=lnms&tbm=isch&sa=X&ved=0ahUKEwiYtdGZsNfQAhVF wI8KHfHvC_gQ_AUIBigB

[5] "Sub Threshold Leakage in MOSFET" [Online].Available: https://www.google.co.in/searchq=sub+threshold+leakage+in+mosfet&biw=1366&bih=638&source=lnms&tbm=isch&sa=X&ved=0ahUKEwjF89OrdfQAhXDK4 8KHW2OAcQQ_AUI Big

[6] J. M. Rabaey, A. P. Chandrakasan, and B. Nikolic, "Digital integratedcircuits. Prentice hall Englewood Cliffs ", vol. 2, 2002.

[7] Kanika Kaur and Arti Noor, "Strategies & Methodologies For Low Power Vlsi Designs: A Review," IJAET, May 2011

[8] Michael Keating, David Flynn, Robert Aitken, Ala Gibsons and Kaijian Shi, "Low Power Methodology Manual for System on Chip Design", Springer Publications, New York, 2007.

[9] Creating Low-Power Digital Integrated Circuits The Implementation Phase, Cadence, 2007.

[10] I. library, "Intel design methodology," http://www.intelpedia. intel.com.

[11] Peiyi Zhao, Zhongfeng Wang and Guoqiang Hang, "Power Optimization for VLSI Circuits and systems," 978-1-4244-5798-4/10/$26.00 IEEE 2010

[12] "Short Circuit Power" [Online] Available: https://www.google.co.in/search?q=short +circuit+Power+Dissi pation&espv=2&biw=1366&bih=638&source=lnms&tbm=isch&sa=X&ved=0ahUKEwiYtdGZsNfQAhVFw I8KHfHvC_gQ_AUIBigB

[13] Lecture, "CMOS Power Consumption" [Online]. Available: http://shodhganga.infli bnet.ac.in/bitstream/10603/6521/8/08_cha pter%203. Pdf

[14] S. Narendra and A. Chandrakasan, "Leakage in nanometer CMOS technologies," Springer, 2006.