CLOCK-GATING OF STREAMING APPLICATIONS FOR ENERGY EFFICIENT IMPLEMENTATIONS ON FPGAS

[1]Dr. John Paul Pulipati, Principal, [2]Dr.Purushotham Naik, [3]P.Venkatapathi,
[4]D. Rajendra Prasad, [5]K. Rajeswar
[1]principal@mrce. in,Malla Reddy College of Engineering
[2]Professor, Dept. of ECE,Malla Reddy College of Engineering
[3]Assistant Professor, Dept. of ECE,Malla Reddy College of Engineering
[4]Assistant Professor, Dept. of ECE,Malla Reddy College of Engineering
[5]Assistant Professor, Dept. of ECE, Malla Reddy College of Engineering

Abstract— This project deals with reducing the dynamic power in streaming applications by using clock-gating. Streaming applications consists of broad class of computing algorithms used in areas such as cryptography, digital media coding ,communications, signal processing, video analytics , network routing etc.,. The power saving can be gained by selectively turning off or switching off parts of the circuit when they are temporarily inactive. So, that the switching state power can be reduced. In this the clock gating methodology is introduced in dataflow designs that are automatically included in the synthesis stage of the high level dataflow design flow. This concept describes an approach for developing energy optimized run-time reconfigurable design which is a benefit from clock gating. For this, FPGAs are highly desirable due to its reconfigurable (or) re-programmable nature, flexibility. It uses clock gating strategy for reducing dynamic power. The main aim of this project is to reduce dynamic power dissipation without much affecting performance and the experimental results also show that the applications synthesized on FPGA platforms shows that power reductions can be achieved with no loss in data throughput. By using this approach we can achieve low power , delay, area and low noise levels.

Index Terms— Clock-gating, data flow , dynamic power, high level synthesis

I. INTRODUCTION

In general, power consumption is one of the major challenges in VLSI design performance. For a silicon device there are two components of power dissipation. They are static power dissipation and dynamic power dissipation. Static power dissipation is due to the leakage current produced within the transistor and by the ambient temperature. Dynamic power consumption can contribute upto 50% of the total power consumption. Inorder to reduce unwanted power dissipation additional circuit is added into the design that effectively clocks the gate. Power dissipation increases linearly with frequency due to highly influence of parasitic capacitances. To counteract this effect, ASIC designers have employed clock gating from past few years[1],[2],[3]. Clock gating is nothing but a power saving feature used in semiconductor devices which enables switching off or turning off circuits. Many silicon devices use clock gating to switch off controllers, parts of processor, bridges and buses inorder to reduces the power dissipation. power dissipation is due to switching of transistors and by losses of charges being moved along wires.

Reducing power has also another benefits. They are it needs less stringent to cool the device, improved life of device or battery and low power costs. Due to these reasons, the power also frequently affects the choice of computing platform right at outset. For instance, Field-Programmable Gate Array

(FPGA) imply higher power dissipation per logic unit by comparing with ASIC. This paper shows the impact of chosen technology on the architecture, but do not describe how to reduce the power at abstraction level of the design.

The major difference between FPGA and ASIC is ASIC is fixed implementation, that means these are pre-defined for a specific task, where as FPGA can be reprogrammed ON- site. FPGA eliminates non- recurring Engineering costs and thereby reduces time to market. Therefore FPGAs are highly desirable to implement digital systems due to their programmability , low end product life cycle, flexibility and all these makes ideal for small and medium applications. FPGAs are slower and less efficient due to added circuitry that is used to make flexible when compared to fixed implementation(ASIC).

Dynamic dataflow designs, for instance the RVC- CAL language possess interesting properties that are used for reducing the power without affecting the behavioral characteristics and the construction of the application. In RVC-CAL language, every actor can execute the processing tasks, executions may be disabled by the input blocking reads and communication between the actors can occur only by the order of storing lossless queue. As a result, an actor may be stopped for a period of time if its processing tasks are idle(or) output buffers (queues) are full without affecting overall throughput and semantical behavior of the design.

II. RELATED WORK & METHODOLOGY

In this paper, the work and methodology described below are based on Xilinx FPGA, which can be used to support the other architectures. A system which works or process on continues stream of bits is said to be streaming application. Cryptography, digital media coding, network routing, video analytics etc., are the examples. In clock gating scheme(CG), the clock is given to those modules that are working at that instant. This clock-gating support the adding of additional logic to the existing synchronous circuit inorder to prune the clock tree, thereby disabling the portions of the circuit that are not in use. In an architecture, the additional device is called clocking circuit which is inserted before the data path unit which provides the clock inputs for working or active modules only. Hence it reduces dynamic power.

Related to our work, S.C Brunet, E. Bezati, C. Alberti M. Mattavelli, E. Amaldi, and J.W.Janneck proposed a multiple clock, domain-design methodology inorder to reduce the power consumption of dataflow programs. Their design motive was to optimize the mapping of the application along with meeting the desired design requirements. The optimization is gained by assigning optimized clock frequency to reduce power consumption.

Present FPGA support various clock-gating strategies and every manufacturer creates its own IP [5] for managing these approaches. The methodology explained here is based on primitives used specific to Xilinx FPGA architectures. Anyway these are modified inorder to support other FPGA vendor primitives.

The execution of a dataflow program contains a series of action firings [2],[6], which can be correlated to one another in a graph-based representation using an approach called Execution Trace Graphing (ETG). In this graph each node represents an action firing and directed arcs represent data (or) control dependency between two various action firings. ETG is a directed and acyclic graph.

By using weighted ETG more optimized buffer size can be gained [8],[9],[10]. In this graph, each firing action is represented with its timing information, hence transform it into weighted graph. It consists of two parts

A. Clock-gating strategy
B. Clock enabling circuit

A. Clock – gating strategy:

The following fig.1 shows the clock gating scheme or strategy. The blocks it contains

ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, VOLUME-4, ISSUE-12, 2017
DOI:10.21276/ijcesr.2017.4.12.30

mainly queue, clock enabling circuit and actor. Actor will be chosen based on the application taken. In this the clock enabling circuit controls the clock of the actor. Whenever the output buffer of the block becomes full then the clock should be turned off inorder to show that the block is in idle state. Switching off the block does not show any effect on the throughput because it is in idle state, there by reduces the power dissipation.

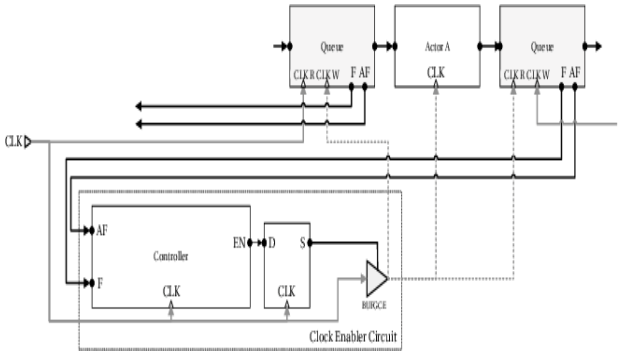


Figure.1 Clock Gating Strategy

Generally RVC-CAL dataflow designs are used for the behavior description, that can be applied to systems that represent execution of process that communicate with asynchronous FIFO buffers. Here the queue blocks should have lossless communication when the actor is clock gated with asynchronous buffers and the design will have different input clock domains.

In the figure shown each queue will have two clocks as inputs, CLKW and CLKR. CLKW is for taking data and CLKR is for producing data. And there are two output ports for the queue. They are Almost Full(AF) and Full (F). From the queue1, the data is given as input to the actor and from the actor the output data is collected by queue2, from this queue2 is readout. When the output of the actor is full, then queue1 will be stopped. To the queue1 the enabling circuit will provide the clock, which will turn off the actor when it is full, hence saving the power. From the fig.1 shown above, the input to the actor is connected to the clock enabling circuit. The clock buffer BUFGCE input clock should be connected to a flip flop to achieve glitch-free clock gating [7].

B. Clock enabling circuit:

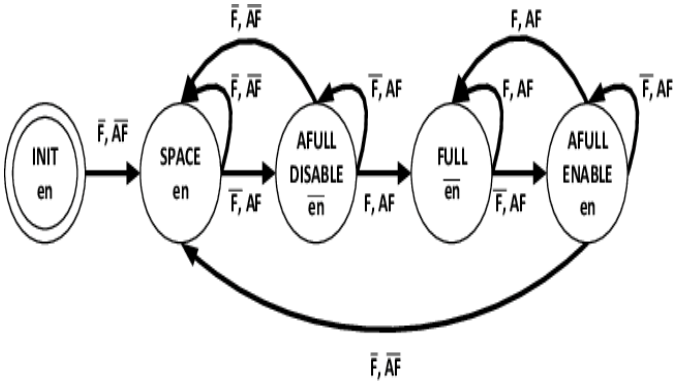


Figure.2 Clock enabling circuit implemented as Finite State Machine

The clock enabling circuit is shown in above fig.2. The shown fig. is implemented as a finite state machine (FSM).

FSM is having a clock, a reset, input Full(F), input Almost Full (AF) and an output enable (EN). The Almost Full AF input becomes active high only when there is one space left to full in its FIFO queue [5]. The above shown FSM has 5 states.

S= {INIT , SPACE , AFULL_DISABLE , FULL , AFULL ENABLE }

The clock-enabling controller circuit starts with INIT state and maintains the ENABLE EN output at active HIGH until F and AF become active LOW. The active high ENABLE EN is maintained during SPACE state. As soon as a queue becomes full, the state will change to AFULL_DISABLE. Here the queue gets disable. In this case, EN output becomes active LOW.

An approach is used in this state as BUFGCE a clock buffer disables the output clock from high-to-low level edge. The enable clock entering BUFGCE must be synchronised to input clock. When a token is taken from the queue, it goes to the AFULL_ENABLE and activates the clock. Then based on the output of buffer Full F (or) Almost Full AF , the state

ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, VOLUME-4, ISSUE-12, 2017
DOI:10.21276/ijcesr.2017.4.12.30

changes to Full (or) to SPACE state.

The user can choose the actor to be clock gated by using mapping configuration. For this, an attribute is given to each actor. The outputs of FIFO queues F and AF are connected to the clock enable controller circuit , if an actor is selected for clock- gating. Output of the queues can be connected directly to a queue (or) through fanout. In the first case, actor output is directly connected to queue without fanout. In the second case, the controller results are connected to other port. In this case one of the fanouts in the queue is full and the fanout should command the actor , but not to produce a token.

III. EXPERIMENTAL RESULTS

This section gives the power reduction gain of the previously mentioned methodology, which is evaluated by applying it to a video decoder design. There are many RVC- CAL applications for dataflow programs [11]. INTRA MPEG4 simple profile decoder is one of the applications. Due to the limitation of number of clock buffers in Xilinx FPGA the design selected was redesigned to result in 32 actors.

MPEG4 video coding is characterized by its scalability and high flexibility. It is a method of compression of audio and the visual digital data. Here the Intra MPEG 4 Simple Profile (SP) description consists of 32 actors. It is a 4:2:0 decoder consists of 8 blocks. Out of these blocks, 4 blocks are luminance (Y) and the other are chrominance U and V, for each two blocks. The parser block contains the bitstream and variable length decoding process, in which U, V and Y are used for texture implements (Tex Y, U,V). The texture decoding consists of variable length coding (VLC), inverse scan test of DCT, inverse quantization, inverse DCT. Whereas, MOT Y, U, V are used to realize the motion compensation stage. Due to the nature of experiments, the MOT stage consists of only one residual error actor. For each queue in the decoder, the minimum queue size is determined [4] by using TURNUS profiler.

For hardware experiment and evaluation, VC707 FPGA kit was used. The code was generated by xronos and was synthesized by Xilinx XST synthesizer. Inorder to produce a netlist, synthesis, routing and placing were applied. The final netlist generated was simulated inorder to extract the Switching Activity Information File(SAIF) of the design. After that, the Xilinx power analyzer was used to get power consumed by using design constraints, design netlist and SAIF inputs also.

The following Table I shows the synthesis results of Intra MPEG4 SP decoder with and without clock-gating.

Logic utilization	Non clock gated	Clock gated	Available
Slices	9214	12776	60720
DSPs	18	18	2850
BRAMs	21499	725	1030
LUTs	109	126	30360
Max freq.		109	-

Table I : Synthesis results with and without clock-gating

A safe option when finalizing your figures is to strip out the fonts before you save the files, creating “outline” type. This converts fonts to artwork what will appear uniformly on any screen.The above table I shows that clock gated decoder use more slices than non-clock gated. Here the clock gating needs only 15% more than LUTs. A 50 MHz clock is given as synthesis constraint. The Table II shown below gives the power consumption of Clock Gating strategy. In this, two tests are considered. Clock gating enabled and disabled [5], when decoding at maximum output.

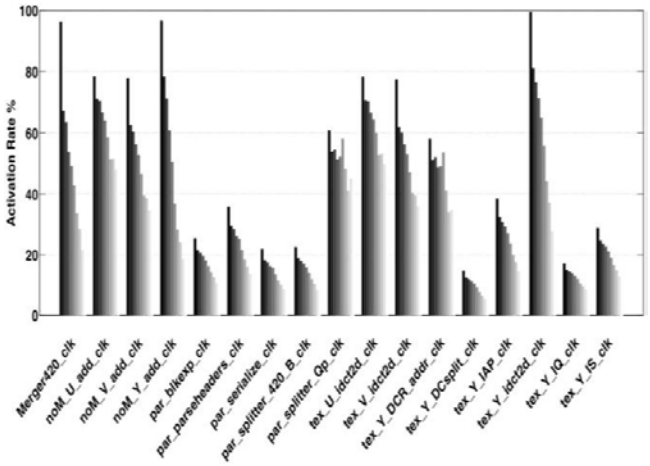
ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, VOLUME-4, ISSUE-12, 2017
DOI:10.21276/ijcesr.2017.4.12.30

Table II : power consumption when clock gating enabled and disabled

Clock gating	Disabled(mW)	Enabled(mW)
Actors clock	58	43
Logic	25	24
Signals	42	41
C		
l		
t		

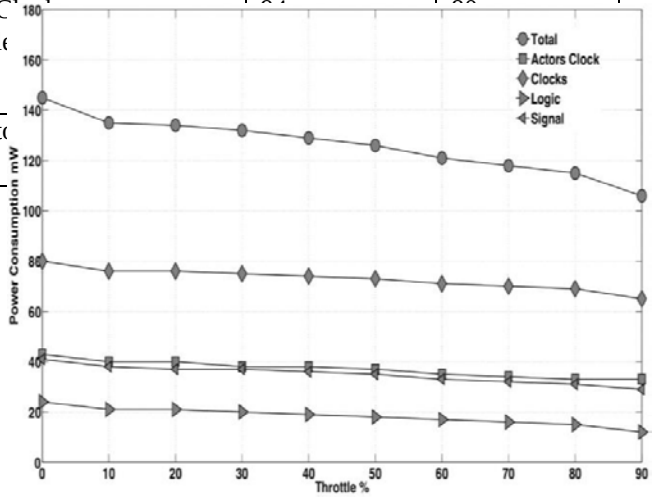
From Table II, the actor clocks show only the power consumption of the actor. Where as, the clock will consist of 50 MHz clock net enabling of clock nets. As a result, the actor clocks consume 26% less power due to clock gating and due to decoder running at high speed, the activation rate of signals and logic result in decrease of 4% in total power.

As from Table I, the maximum decoder output rate is 350 frames/second, for QCIF image of 176x144 pixels. Here the decoder is throttled, such that to decode only 30 images/second for two resolutions. They are CIF(384 x 288 pixels) and QCIF.



The following fig.4 shows (a) the actor power consumption and (b) shows activation rate of actor for each actor clock. The activation rate shows that some of them has an activation rate less than 10%. From this, the power consumption on clocks has reduced by 53.7% for QCIF and 47.6% for CIF resolution. The decoder consumes 54mW less for CIF resolution and 59mW less for QCIF resolution when compared to overall power consumption. From 31 actors, 15 will be always ON, this implies that the 15 actors will never fill up their output. Further, the actors which are not needed for this methodology should find out and eliminate the installation of unwanted additional logic.

In the case of power saving efficiency and bandwidth, the decoder will be throttled from 0 to 90% [5] by simulating channel with different



consumption rates. It is a case where clock gating is applied for a general application. It is shown in fig.3 below.

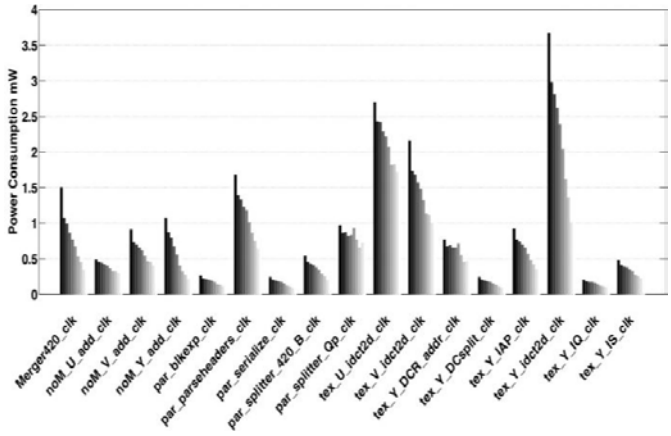


Figure.3 power consumption of clocks, signals, logic and total dynamic power consumption of INTRA MPEG4 SP decoder

From the figure shown above, it is depicted that the total dynamic power has reduced from 145mW to 106mW, a total reduction of 27% power.

The dynamic power of clock gating has been reduced by 34% when compared to non-clock gating. From fig.4 it is shown that, the 15 actors data has been removed due to their activation rate being higher than 99%.

Figure. 4(a) Actor clock power consumption

ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, VOLUME-4, ISSUE-12, 2017
DOI:10.21276/ijcesr.2017.4.12.30

Figure.4(b) Actor clock activation rate

The clock activation rates of actors will be decreased with increase of throttle(except for two cases par_splitter_QP_clk & tex_Y_DCR_addr_clk where power increased slightly). Here, the decoder used is [5] YUV 420. When the power reaches 60%, the chrominance decoding remain active where as, the decoder throttles luminance. It is also occurred during a behavioral simulation in ModelSim.

IV. CONCLUSION

This clock-gating (CG) methodology can be applied to any application inorder to reduce power consumption and reduces the effort during design process at dataflow level. The clock gating(CG) logic is introduced at synthesis stage of HLS design flow. The important component in clock gating is clock enable controller, which is used to give the additional clock. The results show that, when the main block is inactive due to the disabled clock it reduces the switching power. This methodology is very useful where the power dissipation is major challenge.

This method is a simple and effective method inorder to recover power from an idle state. Further developments are necessary inorder to develop the tools used for complex applications onto the limited number of clock domains for more implementations.

V. REFERENCES

[1]. Massoud Pedram, “Power minimization in ic design: Principles and applications,” ACM Trans. Des. Autom. Electron. Syst., vol. 1, no. 1, pp. 3–56, Jan. 1996.

[2]. QingWu, M. Pedram, and XunweiWu, “Clock-gating and its application to low power design of sequential circuits,” Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on, vol. 47, no. 3, p. 415–420, Mar 2000.

[3]. G.E. Tellez, A. Farrahi, and M. Sarrafzadeh, “Activity- driven clock design for low power circuits,” in Computer-Aided Design, 1995. ICCAD-95. Digest of Technical Papers., 1995 IEEE/ACM International Conference on, Nov 1995, pp. 62–65.

[4]. M. Canale, S. Casale-Brunet, E. Bezati, M. Mattavelli, and J. Janneck, “Dataflow programs analysis and optimization using model predictive control techniques,” Journal of Signal Processing Systems, pp. 1–11, 2015.

[5]. Endri Bezati, Simone Casale-Brunet, Member, IEEE Marco Mattavelli, and Jorn W. Janneck, Member, IEEE “clock-gating of streaming aplications for energy efficient implementations on FPGAs” DEC-2015.

[6]. H Mahmoodi, V Tirumala shetty, M Cooke and K Roy,” Ultra Low-Power Clocking Scheme Using Energy Recovery and Clock Gating,” in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 17, no. 1, pp. 33-44, Jan. 2009.

[7]. Xilinx, Analysis of Power Savings from Intelligent Clock Gating, August 2012, XAPP790.

[8]. S. C. Brunet, M. Mattavelli and J. W. Janneck, "Buffer optimization based on critical path analysis of a dataflow program design," 2013 IEEE International Symposium on Circuits and Systems (ISCAS2013), Beijing, pp. 1384-1387,2013.

[9]. B. Ghavami and H. Pedram, “High performance asynchronous design flow using a novel state performance analysis method, Computational Electrical Engineering, vol.35, no.6, pp. 929-941, November 2009.

[10]. S Suhaib, D Mathaikutty, and S Shukla, “Dataow architectures for GALS,” Electron. Notes Theor. Comput. Sci., vol. 200, no. 1, pp. 33–50, 2008.

[11]. “Open RVC-CAL Applications,” 2014, http://github.com/orcc/orc-apps, accessed 25-February- 2014.