TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 5516 - 5522

Power and Performance analysis using Multibit Technology

Manjunatha Visweswaraiah Dr. Somashekar K.
R&D, Design Compiler, Synopsys India. Professor, SJBIT, Bengaluru, India.

Article Info
Volume 82
Page Number: 5516 - 5522
Publication Issue:
January-February 2020

Article History
Article Received*: 18 May 2019*
Revised*: 14 July 2019*
Accepted*: 22 December 2019*
Publication*: 27 January 2020*

Abstract:
The implementation of deep submicron technology has resulted in increased devices on chip making power management a challenging task during design cycle. There are many different techniques developed to reduce the power consumption. Clock-gating, Voltage and frequency scaling, power gating etc. are some of the techniques used to reduce power consumption. Multibit technology is also used widely to help reduce the power. The reduction of power comes with compromise in performance of the integrated circuit. The paper analyses power and performance of face detection chips when multibit is used.

Keywords: *Submicron technology, chip making, power management, multibit*

Introduction

The advances in manufacturing technology have enabled tremendous increase in device count on the chip. The increased gates have multiplied the number of nets used for interconnections. The complexity of the chip increases the power consumption. Reduction of power is an important factor or many applications such as hand held devices like mobile phone, notepads etc [1].

The power consumption in a chip can be divided into three major categories as Dynamic Power, Short Circuit Dissipation, and Leakage Power Dissipation. Each of these categories and their components are discussed below in detail.

Leakage Power Dissipation: This component of power dissipation is getting the most attention these days. Not all the components of leakage consumption existed or dominated for quarter micron and above nodes and thus, it contributed a negligible portion of the overall power consumption. However, with the shrinking of MOS due to technology advancements, the quantum mechanical effects started coming into picture and resulted into many of these leakage current components. This is the component of energy dissipation which affects operation of chip largely in the standby operation as other components seize to play during that period.

It has been observed that during Scan and TDF tests, clock signals are supplied simultaneously to numerous flip-flops in Dynamic Energy Consumption: Dynamic Energy consumption is the consumption due to toggling of the cells and nets because of toggle in the input. This is also known as Switching Energy. When a cell changes its state from logical high to logical low or vice versa, various internal capacitors (junction, interconnect and diffusion capacitances) charge or discharge accordingly. Energy is drawn from the supply to charge these capacitors, known as dynamic power.

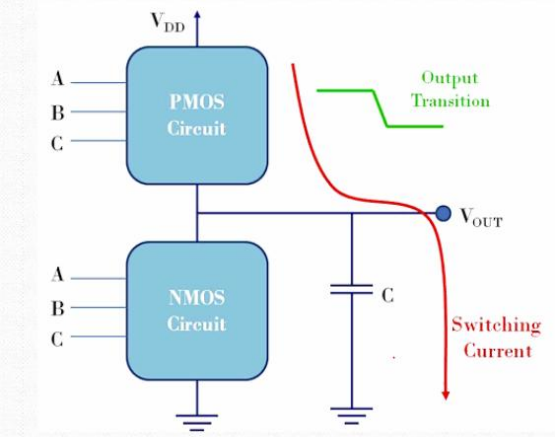


Fig: 1

January - February 2020
ISSN: 0193 - 4120 Page No. 5516 - 5522

If all the parasitic capacitances in a CMOS cell are lumped into load capacitance C, then, if the output level changes from VDD to Ground, there is a total energy consumption of CV_{DD}^2. Half of energy is stored in the load capacitor C and rest half of the energy is dissipated. Similarly, when output changes back to ground, similar energy dissipation of energy takes place. Therefore, this switching energy consumption is directly related to VDD and switching frequency. As a result, reduction of supply voltage is one way of reducing dynamic consumption. However, reduction of VDD causes cells to become slower, therefore, effectively reducing the maximum frequency of the operation. Besides, reduction in frequency causes the same operation to take more time. Average switching energy consumption is:

Pav = f·C·V2

where, f is the frequency of operation. This power consumption is independent of the rise and fall time of the input and output signals

The other component of switching energy consumption is loss due to dynamic hazards and glitches. Glitches may arise in a circuit due to unbalanced delays in the paths of various inputs coming in or in the path internal to the circuit. Consider the circuit as shown below

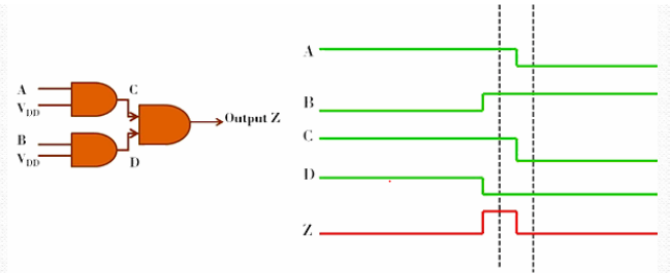


Figure 2: Glitch generation circuit and timing diagram

Consider the case where two of the inputs are at logical one, shown by VDD, and signals A and B transition with some delay, as shown in adjacent timing diagram. Due to unbalanced delays between arrival of A and B, output signal Z is asserted to 1 for a short duration of time. Such transitions are known as glitches/hazards. On the other hand, had A dropped earlier than assertion of B, there would not have been any glitch at the output as one of the input of output AND gate would have toggled to zero before assertion of other input. Therefore, timing is met in such a way that such glitches are either removed or minimized. However, in some cases, this behavior may be intentional to stop race conditions in a circuit. For this purposes, not all the inputs are toggled at the same time. Conditions where such glitches cannot be removed altogether, logic may be placed at the output to absorb such glitches to arrest their propagation to following logic, e.g. adding some buffers in the path to absorb such glitches and balance the timing of the path.

There are many different techniques to reduce power consumption.

Clock-gating: This technique is a very popular Dynamic Power reduction technique. Dynamic power is the sum of transient power consumption (transient) and capacitive load power (cap) consumption. transient represents the amount of power consumed when the device changes logic states, i.e. "0" bit to "1" bit or vice versa. Capacitive load power consumption as its name suggests, represents the power used to charge the load capacitance. Total dynamic power is represented as

$$P_{dynamic} = P_{cap} + P_{transient} = (CL + C)\, V_{dd}^2\, f\, N^3$$

where CL is the load capacitance, C is the internal capacitance of the chip, f is the frequency of operation, and N is the number of bits that are switching. As dynamic power consumption is directly linked to toggling of the MOS cells, gating the clock when not required helps reduce the dynamic current. This technique help preserves the state of the design while only limiting the transient currents. Designers frequently use AND/NOR gates to gate a clock, however, latch based clock gating is the most favored technique as it also saves designs from hazards which canotherwise

January - February 2020
ISSN: 0193 - 4120 Page No. 5516 - 5522

introduce additional power consumption, inherent in dynamic power consumption.

Variable Frequency Islands [1]: In a big chip, not all the blocks should be clocked at highest possible frequency in order to achieve the desired level of performance. There can be few blocks which inherently work slow (e.g., slow communication blocks like I2C, UART, etc.) and, therefore, can be clocked at slower clock than blocks like core/processor which require high frequency clock for maximum throughput. Therefore, by providing different frequency clocks to different blocks, one can reduce localized dynamic consumption.

Power Gating: There can be applications where certain blocks of the chip might not be required to function in some of the low power modes like sleep, deep-sleep, standby mode, etc. and only a part of the device is required to function. In such cases, it makes sense to power off non-functional blocks so that device does not have to power unused blocks. This not only helps reduce the dynamic consumption but leakage power is also saved for such a power gated block. However, while dealing with such a technique, design has to make sure that signals coming in from power-gated blocks do not affect the functioning blocks while operating in low power. For this purpose, isolation blocks are placed in the path so that functionality corruption does not take place, as can be seen in figure 2. Please note that the isolation signals are not required for signals going out of always-ON domain to other power domains as they are never supposed to go non-deterministic

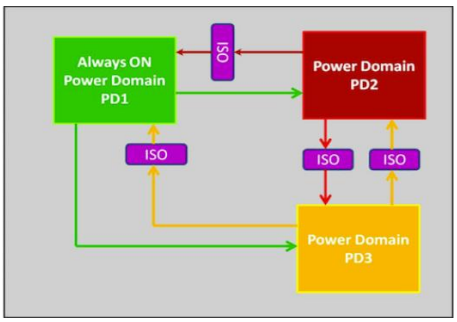


Figure 3: Power Gating

Apart from the above there are other techniques which are process based

Mutli VDD Technique: As we can see from the equation above, there is a quadratic relationship between device voltage VDD and dynamic power consumption. Therefore, one can reduce the dynamic voltage substantially by reducing the supply voltage.

However, voltage reduction has its downside as well. Propagation delay of a cell is as below :

$$T_D = C \,.V_{DD}\, /k.(V_{DD} - V_T)^2$$

As one can see from equation above, reduction in the VDD increases the delay of the cell. thus, the operating frequency of the cell reduces when one reduces the supply voltage. Therefore, one must maintain a balance between voltage supply and associated performance. A solution to this challenge can be to create voltage islands in the design where low performance slow peripherals can be powered using lower supply voltage and performance critical blocks can be powered using higher voltage. However, we have to make sure that appropriate voltage level shifters are placed on those signals which talk across the voltage domains

Dynamic Voltage and Frequency Scaling: Voltage Island technique, also known as Static Voltage Scaling presents few constraints while operating the device. This technique is not adaptive to the application needs and voltage supply to a block cannot be changed once designed. However, Dynamic Voltage Scaling technique liberates designer and customer of such limitations. This technique makes use of a regulator which can be programmed to deliver voltage levels as required. Therefore, various blocks can get configurable voltage and the customer/user can change the voltage settings as per the application settings. This can help save the power dynamically. Various solutions have also been used where the design freed the software to make changes to voltage

January - February 2020
ISSN: 0193 - 4120 Page No. 5516 - 5522

scaling. The design itself senses the current-load requirement in the device and makes the voltage adjustments accordingly. This technique helps reduce power consumption in a more adaptive manner

Multibit technology Multibit cells have more than once cell embedded in a single library cells. The main advantage of using multibit cells areas below

- Reduction in area due to shared transistors and optimized transistor-level layout
- Reduction in the total length of the clock tree net
- Reduction in clock tree buffers and clock tree power

An example of two-bit multibit cell is as below

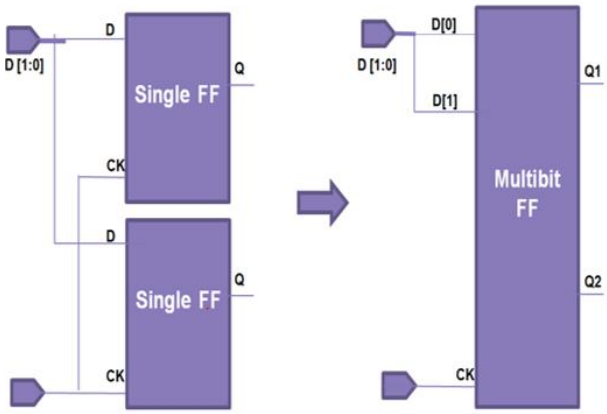


Figure 4: 2 bit multibit cell

The area of the 2-bit cell is less than that of two 1-bit cells due to transistor-level optimization of the cell layout, which might include shared logic, shared power supply connections, and a shared substrate well.

Details of Work Done

The objective of the work is to analyze the power and performance of face detection chips. The face detection chips process real time data and there will be high degree of computation occurring which needs more power. In this work the power and performance analysis of the two designs has been done with and without using multibit cells.

The library is 28nm TSMC having rich flavors of combinational and sequential cells.

The multibit library cells are sequential cells having 2,4, 8 and 16 bit bus width.

Design1:900 K instances, 64 macros, Rectangular Floor Plan.

Design2: 835K instances, 15 macros, Rectangular Floor Plan

Tools Used: HDL compiler for RTL elaboration, DFT compiler for Scan insertion and Design Compiler Graphical for physical synthesis

Design Flow:

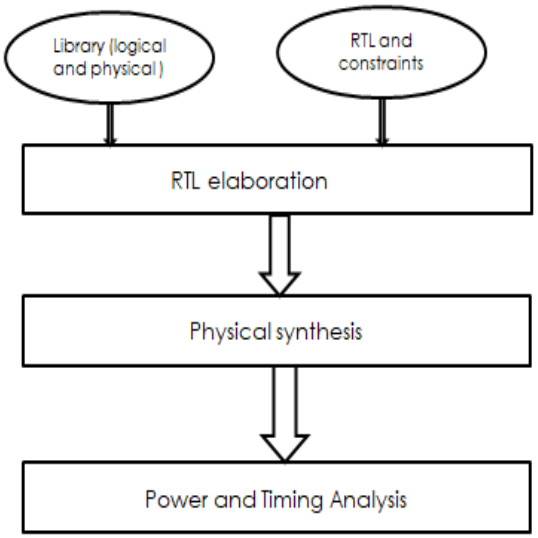


Figure 5: Design flow

The input is the RTL, logical and physical library, Timing constraints and floorplan file which has macro placement information.

The Design flow involved performing first pass of synthesis, then insert the design for testability logic and then perform second pass of synthesis. And then perform analysis of power and timing.

Three different runs conducted on two designs. The details of the runs are as below

Run1: This is called the base flow wherein each design is run through the flow without any multibit mapping of sequential cells

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 5516 - 5522

Run2: In this flow the designs are through the same flow as base flow but multibit mapping is enabled. Multibit mapping is through two steps.
In the initial synthesis before scan insertion the bussed registers were mapped to multibit library cells.
In the second pass of multibit mapping which is called physically aware multibit bankingis performed before scan insertion, in which the sequential cells which are physically nearer to each other are combined and mapped to multibit library cell as shown in figure 6

Figure 6: physically aware multibit banking

Run3: In this flow not all the cells are allowed to be mapped to multibit as it may degrade the timing results. The sequential cells on the timing critical path are not allowed to be mapped to multibit. This flow is expected to yield optimal results in terms of timing, area, and power.

Latest version (2018.06)of Design compiler tool from Synopsys is used for physical synthesis.
The above mentioned flows and runs were performed.
Once the runs were performed, timing and area reports were generated along with power numbers.
The congestion map of the design was also generated to compare the effect of multibit mapping on design congestion.

The following tables provide details about the designs selected.

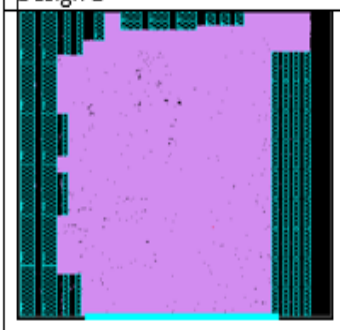
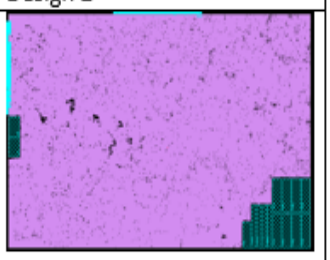
Design 1	Design 2
	
900 K instance 64 Path groups 12 path groups Frequency: 600 MHz	800 K instance 12 Macros 8 path groups Frequency: 400 MHz

Table 1: Design details

The Quality of Results metrics like total negative slack (TNS), Worst negative slack (WNS), Area, Power is compared between the runs for both the designs.

The multibit banking ratio [2]provide details about the number of sequential cells mapped to multibit cells.

If more number of sequential cells are mapped to multibit, the more power and area is reduced.

In Run2, where full multibit mapping is enabled, the packing ratio results are as below

For Design1: The banking ratio is 90.43% as below

Total number of sequential cells:	71216
Number of single-bit flip-flops:	21058
Number of single-bit latches:	5
Number of multi-bit flip-flops:	50153
Number of multi-bit latches:	0
Total number of single-bit equivalent sequential cells:	221675
(A) Single-bit flip-flops:	21050
(B) Single-bit latches:	5
(C) Multi-bit flip-flops:	200700
(D) Multi-bit latches:	0
Sequential cells banking ratio ((C + D) / (A + B + C + D)):	91.0%
Flip-Flop cells banking ratio ((C) / (A + C)):	91.0%

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 5516 - 5522

For Design2: The banking ratio is 94.49% as below

```
Total number of sequential cells:                    27117
   Number of single-bit flip-flops:                  7490
   Number of single-bit latches:                     0
   Number of multi-bit flip-flops:                   19627
   Number of multi-bit latches:                      0

Total number of single-bit equivalent sequential cells:    135894
   (A) Single-bit flip-flops:                        7490
   (B) Single-bit latches:                           0
   (C) Multi-bit flip-flops:                         128404
   (D) Multi-bit latches:                            0

Sequential cells banking ratio ((C + D) / (A + B + C + D)):   94.49%
Flip-Flop cells banking ratio  ((C) / (A + C)):               94.49%
```

In Run3, where timing critical paths are not allowed for multibit mapping, the banking ratio results are as below

For Design1: The banking ratio is 90.43% as below

```
Total number of sequential cells:                    71324
   Number of single-bit flip-flops:                  21208
   Number of single-bit latches:                     5
   Number of multi-bit flip-flops:                   50111
   Number of multi-bit latches:                      0

Total number of single-bit equivalent sequential cells:    221657
   (A) Single-bit flip-flops:                        21208
   (B) Single-bit latches:                           5
   (C) Multi-bit flip-flops:                         200444
   (D) Multi-bit latches:                            0

Sequential cells banking ratio ((C + D) / (A + B + C + D)):   90.43%
Flip-Flop cells banking ratio  ((C) / (A + C)):               90.43%
```

For Design2: The banking ratio is 94.35% as below

```
Total number of sequential cells:                    27464
   Number of single-bit flip-flops:                  7679
   Number of single-bit latches:                     0
   Number of multi-bit flip-flops:                   19785
   Number of multi-bit latches:                      0

Total number of single-bit equivalent sequential cells:    135897
   (A) Single-bit flip-flops:                        7679
   (B) Single-bit latches:                           0
   (C) Multi-bit flip-flops:                         128218
   (D) Multi-bit latches:                            0

Sequential cells banking ratio ((C + D) / (A + B + C + D)):   94.35%
Flip-Flop cells banking ratio  ((C) / (A + C)):               94.35%
```

The Quality of Results (QoR) comparison tablebetween the base run Run1 and Run2 (where full multibit mappingis enabled) for Design1is as below.

	Run1	Run2	% Change
TNS	0.0255	0.0355	10% degradation
WNS	0.13	1.16	23% degradation
AREA	1866796	1819132	2.55% improvement
POWER	1000mW	116 mW	88% improvement

Table2 QoR comparison b/w Run1 and Run 2 for Design1

The Quality of Results (QoR) comparison table between the base run Run1 and Run2 (where full multibit mapping is enabled) for Design2 is as below.

	Run1	Run2	% Change
TNS	0.00164	0.0025	0.04% degradation
WNS	8.01	10.11	25% degradation
AREA	995941	955086	4% improvement
POWER	33 mW	24 mW	27% improvement

Table3: QoR comparison b/w Run1 and Run 2 for Design2

The Quality of Results (QoR) comparison table between the base run Run1 and Run3 (where full multibit mapping is not enabled) for Design1 is as below.

	Run1	Run3	% Change
TNS	0.0255	0.0270	0.02% degradation
WNS	0.13	0.129	No change
AREA	1866796	1829495	4% improvement
POWER	1000mw	600 mW	40% improvement

Table4 QoR comparison b/w Run1 and Run 3 for Design1

The Quality of Results (QoR) comparison table between the base run Run1 and Run3 (where full multibit mapping is not enabled) for Design2is as below

	Run1	Run3	% Change
TNS	0.00164	0.00170	0.04% degradation
WNS	8.01	9.0	12.5% degradation
AREA	995941	950238	4% improvement
POWER	33 mW	25 mW	24% improvement

Table3: QoR comparison b/w Run1 and Run3 for Design2

January - February 2020
ISSN: 0193 - 4120 Page No. 5516 - 5522

The congestion map is also compared between the runs for two designs as below

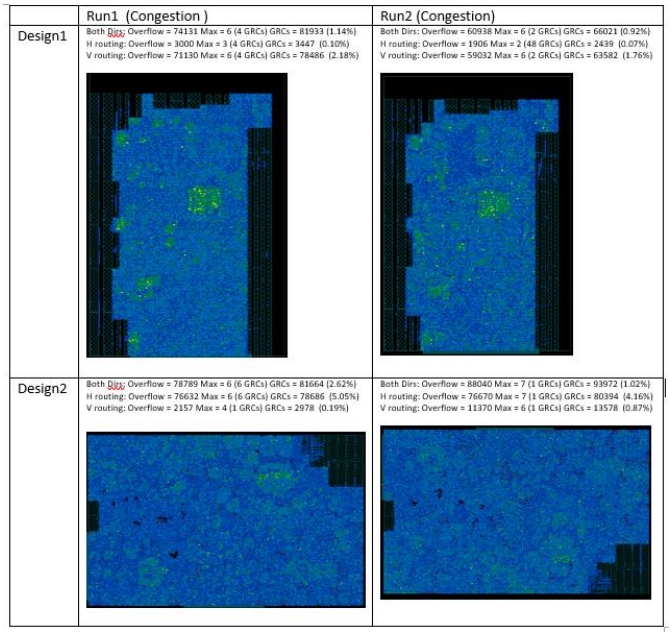
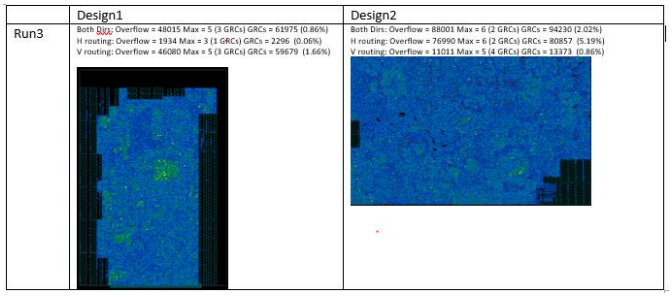


Table4: Congestion comparison

The congestion for run3 is similar to run2. There are no red spots to make the design non routable and congestion is manageable.



Conclusion and Future Work

From the experimental results it is observed that using multibit mapping technology

- The area is improved for both the designs significantly
- There is huge reduction in total power.
- There is no congestion degradation due to mapping to large cells.
- There is slight degradation in worst negative slack (WNS)and it is observed that the degradation is due to multibit cells in the critical path.
- There is sizeable degradation in Total Negative Slack (TNS). The most degradationsare due to multibit cells in timing paths.

It is evident that the multibit technology helps to reduce the area and to reduce dynamic power dissipation for the design but there is a negative effect on the timing performance of the design. and dynamic power consumption.

The technology when used intelligently on the design can help provide better area, dynamic power without compromising on the timing Quality of Result. In run3
approach the sequential cells on timing critical path are not allowed to map to multibit cell to ensure timing degradation does not happen The future work can be comprised of taking the design through full flow in physical implementation like detailed placement and clock tree synthesis to compute the clock tree power and to develop a methodology for more intelligent mapping of multibit cells.

References

1. P. Girard, “Survey of Low-Power Testing of LVSI Circuits”, in proc. IEEE Design and Test of Computers, May-June 2002,
2. Synopsys Design Compiler User Guide
3. Synopsys IC Compiler II User Guide
4. Wai Tung Ng, O Trescases “Power management for Modern VLSI loads using dynamic voltage scaling” in proc Proceedings. 7th International Conference on Solid-State and Integrated Circuits Technology, 2004.
5. Martn D. F. Wong“Low power design with multi-Vdd and voltage islands” 2007 7th International Conference on ASIC