Low Power Clocking Strategies in Deep Submicron Technologies

M. Samy Hosny, SiliconPro, Ottawa, ON K2K 2E2 Canada. shosny@silicon-pro.com,
Yuejian Wu, Nortel Networks, Ottawa ON K2H 8E9 Canada. yuejian@nortel.com

Abstract—Recent silicon process technology advancements have given chip designers integration capabilities never were possible before, and have led to a new wave of complex ASICs (Applied Specific Integrated Circuits). These advanced processes come with new challenges. This paper presents some of the challenges in deep submicron technologies, which require new design practices. We demonstrate some issues related to clocking strategies and timing closure that were encountered during the design of a FEC40 ASIC, and a methodology is proposed to mitigate some of these issues. The FEC40 ASIC is a Forward Error Correction Chip designed for Nortel’s 40Gb/s coherent optical transmission system. The chip has about 11 million gates at a core frequency of 350MHz and has 70 clock domains. The chip was fabricated in a 90nm technology.

Index Terms—Clock Gating, Static Timing Analysis, deep submicron technologies, On Chip Variation (OCV)

I. INTRODUCTION

Historically, designers have analyzed digital chips at the various voltage, temperature and process corners without modeling process variation in the same chip. In deep submicron technologies, designers can no longer ignore the variation between transistors and interconnect characteristics on the same die. In this paper we will study how these variations could affect the design process and discuss some of the precautions that should be taken by designers to avoid increase in power dissipation and potential problems at the layout phase and final timing closure.

II. UNDERSTANDING OCV

The manufacturing process of Integrated Circuits is inherently imperfect. Slight variations in the durations, temperature and chemical concentration at each step result in variations in transistors and interconnect characteristics. These variations can be classified into two categories [1]. D2D variations, resulting from lot-to-lot, wafer-to-wafer and within wafer process variations, affect all transistors and interconnect on a die in a similar way. Speed or Fmax binning techniques have been used effectively for high performance microprocessors to partially mitigate D2D variation impacts. Recently, post-silicon, chip-level compensation techniques such as adaptive supply voltage (ASV) and adaptive body bias (ABB) [2-3] have been proposed to further alleviate impacts of D2D variations.

On the other hand, on-chip variation (OCV) is the result of process variations between cells and interconnects within the die. In earlier technologies, OCV was negligible and it was necessary to guard-band the design only for potential process variations between wafers, where only 2 corner cases need be considered for PVT (Process, Voltage and Temperature): minimum and maximum. In sub 100nm technologies, the structure on the die is so small even a minimal process variation can result in a large effect with regard to timing, power or signal integrity. For example, a variation of 1nm at the 100nm technology node equates to only 1% of the nominal value. By comparison, the same 1nm variation at the 65nm node equates to more than 15% of the nominal value. The impact of such variation on timing cannot be ignored. The following sections will study the impact of such variation on design clocking structures in low power designs.

A. OCV effect on Timing

One of the main factors that complicate timing analysis in deep submicron technologies is OCV. The many variables involved in the manufacturing process that affects cell and wire delays are a Gaussian distribution with mean and standard deviation determined by the cell design and these process variations. Fig. 1 shows a cell delay distribution under PVT effect in case of sub 100nm, and an older technology such as 180nm technology. The path delay distribution in the case of the 180nm, for example, is very narrow (shown in dark color in Fig. 1) and thus it is sufficient to perform timing analysis at only 2 extreme corner cases, Min PVT and Max

Fig. 1. Effect of OCV on path delays
PVT conditions. In sub 100nm technologies, however, it was found in the case of the FEC40 ASIC that the path delay distribution widens up such that the worst and best path delay spread at the Min PVT and Max PVT becomes more than a factor of 3.

B. OCV effect on Clocking

As technology advances, a chip may have millions of gates with very complex structure. OCV effect is most evident on the clock network in which it changes clock latency to logically dependant flip-flops. Such changes can lead to setup and hold violations. To account for these variations, a built-in pessimism is included in the Static Timing Analysis (STA) by derating the paths by a certain value, typically, between 10% to 30% for many 90nm and 65nm technologies. Such derating can be too pessimistic if launch and capture flops share a common branch of a clock tree, as illustrated in Fig. 2. This is because any OCV variation on the common portion of the clock tree would affect both the launch and the capture flops in exactly the same way and by the same amount. Most STA tools account for this effect by ways of Clock Reconvergence Pessimism Removal (CRPR), i.e., removing the OCV derating from the portions of a clock tree shared by both launch and capture flops. CRPR makes timing analysis under OCV more realistic.

III. CLOCK GATING IN SUB-100NM TECHNOLOGIES

Clock gating is a very effective technique for power reduction, where branches of a clock tree are turned off when the circuitry on the branches is not needed for certain duration of time. With the clock gated-off, the circuitry does not consume any dynamic power for no switching activity exists inside the circuitry. When the function of the circuitry is required, a controller turns on the clock and the circuit is reactivated. In the past, clock gating was easy to implement since most layout tools were able to balance gated clock tree leaves with their un-gated counterparts, achieving the so-called “zero” clock skew.

With the introduction of deep submicron technologies, however, clock gating must be carefully planned for successful timing closure since the layout tools can no longer guarantee the “zero” clock skew under the influence of OCV. To illustrate the impact of OCV on timing closure, consider a multi-level Gated Clock structure such as that shown in Fig. 3, where GTC1, GTC2 and GTC3 are Gated Clock cells at different levels of the clock tree. An example clock gating cell is also shown in Fig. 4. The following sub-sections analyze the OCV impact on data paths and clock enable paths respectively and discuss potential solutions.

A. OCV Impact on Data Paths

Let us first examine the data paths between flip-flops FF_A and FF_B in Fig. 3. Without considering OCV, the layout tools can often achieve “zero” skews between the clock tree leaves driving FF_A and FF_B. As a result, signals propagating through the combinational circuitry have roughly a whole clock period to propagate from FF_A to FF_B. Assuming the clock period to be $T$, the setup margin in this case is given by

$$SU_{AB} = T - t_{AB} - t_{SU-B} \quad \ldots \ldots \ldots \ldots \ldots (1)$$

Where $t_{AB}$ is the combined CLK to Q delay of the launching flop plus the propagation delay of the combinational circuitry between the flops; and $t_{SU-B}$ is the library setup requirement of FF_B.

Now let us take the OCV effects into consideration. As discussed in the previous sections, OCV only applies to the portions of a clock tree that are not common to FF_A and FF_B. To a first order of approximation, the setup margin under the influence of OCV can be given by:

$$SU_{AB-OCV} = T - (CID_{LEAF} - CID_3)*C_{OCV} - t_{AB} - t_{SU-B} \ldots \ldots (2)$$

Where $CID_{LEAF}$ is the total clock insertion delay to FF_A and FF_B; $C_{OCV}$ is an OCV factor, which is often between 10-30% depending on the underline technology; $CID_3$ is the common clock tree portion shared by both FF_A and FF_B. Thus, $(CID_{LEAF} - CID_3)$ represents the portion of clock tree branches that are not common to FF_A and FF_B, and affected by OCV.

B. OCV Impact on Clock Enable path

OCV impact on clock enable paths is more critical than data path. This is mainly because a clock enable signal is often generated from a flop at a clock tree leaf and the clock gating cell is often further up on a clock tree, which narrows down the setup window to start with. For high speed clocks and large designs, the clock tree insertion delay from a gating cell to its leaves that generate the clock enable signal can be at the same order of magnitude as the clock period. This is illustrated
in Fig. 4, where CLK_LEAF is the clock on a leaf cell; CLK_GTC3 is the clock driving the clock gating cell GTC3; and CEN_GTC3 is the generated clock enable signal. Without OCV, it can be challenging enough to meet timing since CEN_GTC3 is generated late in comparison to CTG3 clock, CLK_GTC3. To make matters worse, most clock gating cells expect to capture their clock enable at a falling clock edge. The setup margin in this case is given by

\[ SU_{CEN} = \frac{\tau}{2} - (CILD_{LEAF} - CILD) - t_{GTC3} - t_{SU-GTC} \quad \ldots (3) \]

Where \( t_{GTC3} \) is the combined CLK to Q of the launching flop plus the propagation delay through the path of the clock enable signal and \( t_{SU-GTC} \) is the setup requirement of the clock gating cell.

The introduction of OCV further reduces the setup margin and makes the timing closure even more difficult. To a first order approximation, the setup margin considering OCV is as follows:

\[ SU_{CEN-OCV} = \frac{\tau}{2} - (CILD_{LEAF} - CILD)^*(1+COCV) - t_{GTC3} - t_{SU} \quad \ldots (4) \]

As shown in equation (4), if the clock is fast enough or \( \tau/2 \) is small enough, we may have \( (CILD_{LEAF} - CILD)^*(1+COCV) \geq \tau/2 \), thus making it very difficult to close timing.

There are, generally, 3 ways to deal with this problem. The first is to minimize the overall clock tree insertion delays. Ideally, if the clock insertion delay is negligible compared to the clock period, there will be significantly less issues related to OCV. However, achieving this goal is very difficult with today’s large designs.

The second solution is to pipeline the clock enable signal and use a slightly earlier clock for each pipeline stage, which will requires careful manual intervention in layout. In this case, detailed instructions must be provided to layout engineers.

The third approach is to move clock tree gating cells close to their leaves, i.e., maintaining CILD_{LEAF} but increasing CILD in the earlier examples, which reduces the overall \( (CILD_{LEAF} - CILD) \). An extreme case is to have one clock gating cell drive only one or a very few flip-flops, which will minimize the OCV impacts, but may significantly compromise power savings on clock tree buffers. On the other hand, for maximal power savings, one would place a clock gating cell as far away to its leaves as possible, which will, not only saves power in the flip-flops and associated circuitry but also saves power on all the clock tree buffers affected.

IV. DESIGN METHODOLOGY FOR GATED CLOCKS

From the previous discussions, it was shown that in sub 100nm technologies, designers now must balance the amount of power savings against potential timing closure difficulties due to excessive OCV.

Fig. 5 shows a proposed flow for a gated clocking strategy for sub 100nm technologies. Initially, a trial synthesis run is performed to get an estimate of the number of flops on a target clock domain. If the clock domain is heavily loaded, then a multi-level gated clocking structure, as the one shown in Fig 6, can be created, where GTC1 is placed at the root of the clock tree. This cell, in turn, will be driving 2 other clock gated cells GTC2, each will, in turn, be driving 4 GTC3 gated clock cells. This way the load of the clock tree will be split as follow:

- GTC1 gated clock cell will be loaded with \( n \) flops, where \( n \) is the total number of flops in this clock domain. GTC1 will be physically located at root of the clock tree.
- GTC2 gated clock cells will be loaded with \( n/2 \) flops each, and will be located in the middle of the clock tree.
- GTC3 gated clock cells will be loaded with \( n/8 \) flops each, and will be the closest to the leaf of the clock tree.

The design is then synthesized and the loading on the clock gating cell is checked. The netlist is then pushed through a full place and route cycle. After extraction, the Clock insertion delays at the gated clock cells are verified and compared against the one at the leaf of the clock tree.

The ideal level of clock gating cells is selected based on the clock insertion delay at the clock gating cells. The

Fig. 5. Methodology for Clock Gating
level closest to the clock trunk that satisfies equation (4) is then selected.

V. CLOCK GATING FOR LOWER LEAKAGE POWER

The process of inserting gated clocks into a design can be performed in two different ways. The easiest way is to have the synthesis tools do the job by specifying the maximum and minimum number of flops associated with each clock gating cell, and let the synthesis tool do the rest. The disadvantage of this method is that if the enable signal is a repetitive pattern of the clock, such as a divide by an integer value of the clock, the data path of the set of flops associated with this gated clocks could be over constraint. This might lead to unnecessarily using high leakage/high speed cells or oversized logic to meet timing in this particular data path. A better method for clock gating is to add the gated clock cell in the RTL. This will give the possibility of changing the timing constraints accordingly and relaxing the timing constraints on this path. A second advantage of this methodology is that it could substantially simplify the design of derived clock domains (i.e. divide by 2, by 3, ...), since they no longer need to be designed as a separate clock domain compared to the sources clock, which simplifies clock tree synthesis and data hand-off. Fig. 7 illustrates how this flow works. If we consider a certain section of the circuit operates on a half of the clock frequency of the main clock, the instantiating the gated clock cell and adding proper timing constraints as shown in the bottom part of the figure will allow the synthesis tools to select slower and less leaky cells from the library.

VI. CONCLUDING REMARKS

Present EDA tools suffer from some shortcomings for some of the new design challenges that emerged with sub 100nm technologies. In this paper, we have demonstrated how a clock gating cell can be optimally placed within the clock tree for better power savings while easing timing closure considering OCV. Although this methodology requires the instantiation of the clock gating in the RTL, this could be automated by simple feature enhancement in the RTL synthesis tools. We have also demonstrated the value of redefining clock frequencies after inserting the gated clock cells in critical paths. When gated a clock is a divided down version of the original clock, leakage power may be reduced on critical paths if clock frequencies are redefined at the output of the gated clock cells.

Fig 8 shows the floorplan of the FEC40 ASIC of which we implemented some of these methodologies successfully.

REFERENCES