# Power Gating to reduce Leakage Current in Low Power CMOS Circuits

## Sharath B.P, Suhas K.V

[1]*(Telecommunication Engineering, APSCE/VTU, INDIA)*
[2]*(Telecommunication Engineering, APSCE/VTU, INDIA)*

***Abstract :*** *Power-gating has proved to be one of the most effective solutions for reducing stand-by leakage power in nanometer scale CMOS circuits, and different strategies and algorithms for its application have been proposed recently. Unfortunately, power- gating comes with its own set of costs: Performance degradation, area increase, dynamic power increase and routing congestion. A decision to power-gate a design has to be taken, pros and cons of power-gating have to be properly weighted to achieve optimal results. In this paper, "Figures of Merit" (FoMs) for power-gating can be used by designers to better understand the benefits and costs of power-gating, thereby allowing them to achieve optimal results.*
***Keywords:*** *Figures of Merit, Leakage current, Power-Gating, Sleep transistor size, Turn off/on time.*

## I.        Introduction

Leakage power is a major concern in sub-90nm CMOS technologies and numerous design techniques have been introduced to reduce stand-by leakage in digital circuits [1], [2]. Out of the available solutions, power-gating has proved to be very efficient in reducing the stand-by leakage currents in the idle mode. In its most common exploitation, shown in Figure 1, power-gating relies on the insertion of a high- threshold NMOS transistor between the source terminal of a cell/block and the ground rail. The insertion point of the sleep transistor is called virtual ground line (VGND).
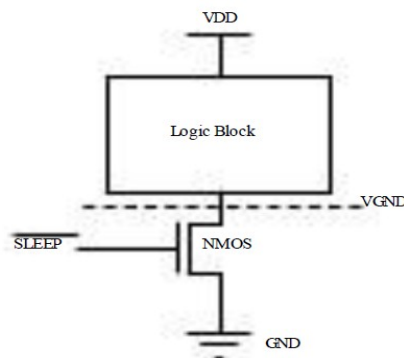


Fig 1: A logic block with a nMos Sleep transistor

A sleep transistor can be added on a cell-by-cell basis or can be shared among a number of cells belonging to the same block/cluster.  Clustered sleep transistor insertion (STI) implies a smaller sleep transistor area compared to adding asleep transistor to every single cell, due to discharge current balancing of cells connected to it. In addition, it ensures a lower congestion overhead for sleep signal routing and lower buffering overhead required for sleep signal driving. These are the main reasons why clustered STI has emerged as the preferred choice for industrial application of power-gating. In this paper, we will thus consider clustered STI as the reference power-gating methodology.

Clustered power-gating offers a multitude of challenges and optimization opportunities. The major one concerns the place- ment of the sleep transistors and the routing of the virtual ground lines. STI is done as a post-layout step; therefore, care should be taken to guarantee little perturbation of the original layout to achieve final design closure. An equally important issue is related to sleep transistor sizing to ensure compliancy with timing constraints. More specifically, since STI intrinsically results in performance degradation due to slowing down of the cells connected to the NMOS footers, the sleep transistor should be up-sized to minimize the virtual ground voltage ($V_{VGND}$), thus the timing overhead. On the other hand, large sleep transistors would lead to an increase in the overall sleep transistor area and the associated buffers, long power-mode transition times (turn on/off times) and a non-negligible energy spent during the power-mode transition. Depending on the timing

criticality of the design and the ratio between stand-by time and active time, one can choose how large the sleep transistors should be to achieve optimal power savings.

To make sure that the total power/area/performance overhead due to power-gating does not overwhelm the benefits of power-gating, a careful investigation of the Figures of Merit (FoMs) for power gating has to be considered to achieve optimal results. This is exactly the main goal of this paper. In particular, we first define a set of FoMs which are suitable for a thorough assessment of the effectiveness of power- gating; next, we discuss how such FoMs can provide useful guidelines to designers for the best exploitation of clustered power-gating. We will refer to realistic data collected on circuit layouts implemented with a state-of-the-art, CMOS process (i.e., the 65nm library by STMicroelectronics), and to an industry-strength design flow for clustered power-gating.

## II. Figures Of Merit For Power-Gating

Recently, a few works have specifically addressed the issue of quantifying the benefits and costs of power-gating. In [3], the authors discuss system level aspects of power- gating and provide insightful discussions on various costs and benefits of this technique. They conclude that, for sub-90nm technologies, the benefits achieved by power-gating are quite substantial compared to the overhead due to it, thus making it a very attractive option to achieve power savings. However, the authors did not address the issues at design/block level and on how one can achieve optimum savings by trading different parameters, such as performance degradation and area. In [4], the authors define the efficiency of power-gating and show how the efficiency varies with different design metrics, such as sleep transistor area, length and also throw light on methodologies, which can be used for sleep transistor power-mode transition. However, the authors do not provide substantial results on real circuits.

As anticipated, this paper takes a step further in this direction by defining various FoMs for power-gating, by benchmarking them against industry-strength design implementations and, finally, by providing useful design tips for applying power- gating in practice. The remainder of this section is devoted to the definition of the FoMs.

1.1 Performance Degradation ($\alpha$): This is a design specification or an input specification, and defines the maximum allowable delay increase in the design/logic block. This is, in fact, the increase in the original critical path delay permitted when the design/logic block is power-gated. We denote this parameter by $\alpha$, expressed as the percentage increase in the original critical path delay. This metric quantifies the degradation, in speed or performance, can be tolerated when power-gating a design.

1.2 Sleep Transistor Size ($W_{sleep}$): Sleep transistor size depends mainly on two parameters: (i) The virtual ground voltage, $V_{VGN D}$ ; (ii) The peak discharge current of the design/logic block, $I_{peak}$ . For a given peak current value Ipeak, one can have a lower $V_{VGND}$ value to obtain a lower speed degradation of the cells in the power-gated design, thus resulting in a larger sleep transistor. On the other hand, a higher value of $V_{V GN D}$ would lead to a smaller sleep transistor, but to a higher delay degradation of the cells in the power-gated design.The relationship between $W_{sleep}$ and $V_{VGND}$ can be written approximately as:

$$Wsleep = \frac{L}{\mu n * Cox * Ron * (Vdd - Vth)}$$

Where L and Vth denote the channel length and the threshold voltage of the sleep transistor, respectively, $\mu_n$ represents the mobility of the electrons and Cox represents the cell capacitance per unit area of the sleep transistor. Finally, Ron denotes the channel resistance of the sleep transistor operating in the linearity region.

While a smaller sleep transistor results in higher performance degradation of a design/block, a larger sleep transistor leads to an increase in chip area, in the number of associated buffers to drive the input sleep signal of the sleep transistor and in the energy dissipated during power-mode transition. Apart from these, as indicated by [3], as the percentage of gated area on a chip increases, more de-cap insertion in the active circuits is needed to make sure that the ground bounce is under the maximum allowable limits, which again increases leakage and total chip area. Thus, sleep transistor size is a very important design metric in power-gating.

1.3 Leakage Power Savings: The main benefit of power-gating is how much leakage power can be saved. As discussed earlier in the paper, performance degradation and sleep transistor area are opposing design constraints and hence one has to trade these metrics very carefully to achieve optimum power savings. For given $\alpha$ and Wsleep constraint, one can partially power-gate a design, thus leaving out timing critical cells from being power-gated, thereby satisfying both the constraints while still achieving considerable leakage savings.

1.4 Power Mode Transition Time ($P_{Mt}$): Given a power-gated design/block, we define the turn-off time as the time required by the design/block to go from the active to the stand-by mode, and the turn-on time as the time required by the design/block to make the opposite mode transition. In some cases, a design/block is active most of the time, and it only provides small windows of stand-by time; the availability of short turn-off and turn-on times may enable leakage reductions corresponding to such short stand-by periods. Turn-off and turn-on times mainly depend on the size of the sleep transistor. In particular, the turn-off time depends on the total input capacitance of the sleep transistor and on the time required to pull-up the virtual ground to a stable value. On the other hand, the turn-on time depends on the total capacitance of the virtual ground line and on the net charge flow from the power-gated design/block to the ground. The virtual ground line capacitance has two components, namely: (i) The virtual ground line parasitic capacitance; (ii) The output capacitance of the sleep transistor. We observe that we do not consider the turn-off time of the sleep transistor as an important parameter since, although it is not negligible, it does not introduce any performance penalty in the context of power-gating because it overlaps with the stand-by period. For this reason, in this paper we provide results only for the turn-on time.

1.5 Power-mode transition energy ($P_{Me}$): This quantity de- notes a non-negligible energy dissipation during turn-on/off of the sleep transistor. The energy loss occurs in the charging and discharging of the virtual ground line capacitance and in the buffers which drive the sleep transistor. This parameter is strongly coupled with the size of the sleep transistor. A large sleep transistor would mean higher energy dissipation during turn-on/off and vice versa. As indicated in [3], the total energy overhead due to STI should be very low compared to the leakage savings to achieve benefits from power-gating.

## III.    Power- Gating Flow

To quantify the FoMs of power-gating, we adopted a flow based on the row-based methodology of [5]. The methodology realistically assumes that one or more rows are dedicated to the insertion of the sleep transistors and are appended at the end of the existing layout. The methodology guarantees marginal layout changes, that is, it is very much "layout friendly"; as such, it is easily usable within industrial design flow.

1.1 Clustering and Sizing: As discussed in [5], the clustering algorithm is based on using rows of the layout as atomic clustering objects. It takes two input parameters, namely, the performance ($\alpha$) and the area ($W_{sleep}$) constraints. $W_{sleep}$ can be mapped onto an optimal $V_{VGND}$, which sets the individual cell delay degradation. Then, depending on $\alpha$, the algorithm finds an optimal sub-set of rows of the design that can be power-gated, thus satisfying both the size and the performance constraints. Sleep transistor sizing requires an accurate estimation of the peak discharge current ($I_{peak}$) of the design/block to which power-gating is applied. We use the technique of [6] for this task, which builds temporal current waveforms for each cell based on the possible time intervals during which the cell can switch. Using this waveform for each cell, the peak current waveform for the entire design/cluster can be constructed and the actual peak current value can then be obtained.

1.2 Turn-On/Off Methodology: To measure the $P_{Mt}$ and $P_{Me}$ , we use the following methodology. We implement a n-stage power-mode transition scheme, as shown in Fig 2. Given a sleep transistor size, we design an optimal inverter chain for driving the sleep transistor input with the goal of achieving minimum turn-off time. For turn-on, we implement a parallel sleep transistor architecture where each sleep transistor is optimally sized to reduce the turn-on time while making sure that the ground bounce is lower than a specified maximum value to avoid sudden current surge on the supply rails. We set a maximum current equal to $I_{peak}$ of the design as the allowable peak current during turn-on. Since different circuits will have different $I_{peak}$, the P Mt of the circuits also varies accordingly. Note that in the figure, the $V_{GND}$ pull-up transistor is used to pull the virtual ground and the internal nodes of the power-gated design/block to a logical value since some internal nodes become floating when a design/block goes into the power-down mode.
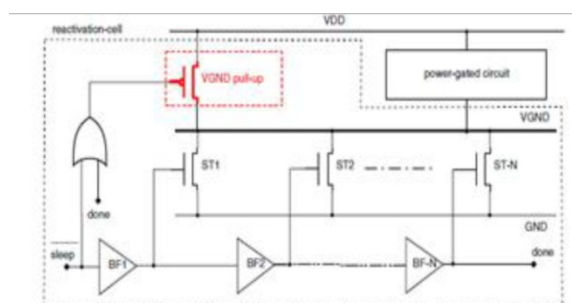

Fig 2: Sleep Transistor Cell Architecture

## IV.    Experimental Results

1.1 Experimental Setup: We applied the power gating flow to the total of four designs. Emulated the working of a processor chip by designing a 16 bit full adder, 8 bit Multiplier, Flip Flops and a 4 bit Asynchronous Counter. Each design was placed and routed on to a 45nm technology using a Synopsys Physical Compiler.  The experiments we performed targeted the exploration of the achievable leakage savings (i.e., *Savings*), of the power mode transition time (i.e., *PMt*) as functions of $\alpha$ and *Wsleep*. For sweeping $\alpha$ parameter, we first compute the critical path delay for each design and then we set $\alpha$ as the percentage increase in critical path delay For this setting, the sleep transistor size is kept constant at 40u. For sweeping the *Wsleep* parameter, we first set $\alpha$ at a constant Value of 5%, then we compute the total sleep transistor size required to power-gate the entire design with this $\alpha$ value.
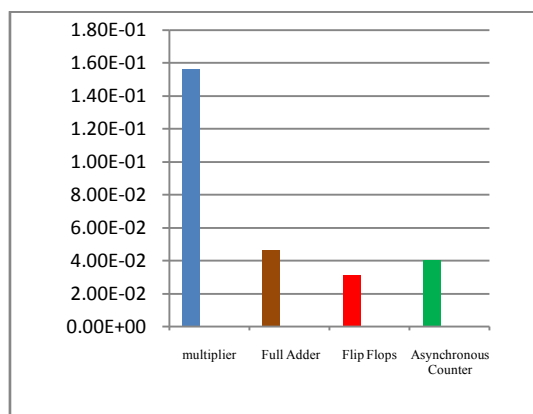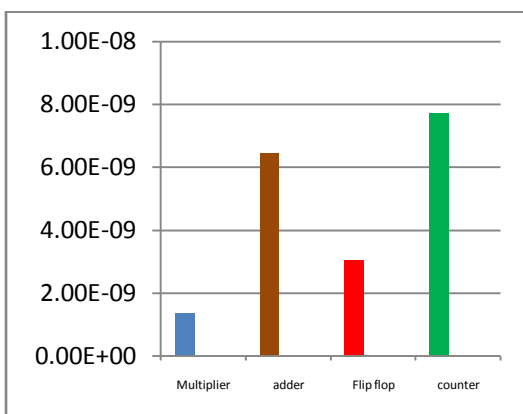
| Fig 3: Leakage current without Power gating | Fig 4: Leakage current with Power gating |
|---|---|

1.2 Results: From the chart of Figure 4 we can clearly evidence that, when the allowed performance degradation increases, the amount of savings increases as well since, for the same area constraint (i.e., for the same sleep transistor), more cells can be powergated. The variation of leakage savings with *Wsleep* follows the same trend. If we can accommodate more area for the placement of the sleep transistors, we can achieve savings for a given performance degradation. The achievable savings clearly depend on the type of circuit. However in some cases, increasing *Wsleep* might not generate enough slack to power-gate more and hence the savings will not increase. Regarding the *PMt* parameter, we recall that this mainly depends on the amount of cells that are power-gated and on the size of the sleep transistor. Transition time increases as we increase the allowed performance degradation. This is due to the fact that, when we increase $\alpha$, we increase the amount of cells that can be powergated, thereby increasing the charge flow from the power-gated circuit to the ground.

## V.    Conclusions

Power-gating has proved to be a very efficient way to reduce leakage power in stand-by mode, but its full exploitation entails a good understanding and weighing of the costs and benefits that are attached to this technique. In this paper, we have carried out a detailed investigation of the pros and cons of power-gating by defining Figures of Merit (FoMs) for power- gating. We have applied a power-gating methodology to a set of designs synthesized in the 65nm CMOS technology, we have studied how the FoMs vary for different design constraints, and we have derived some guidelines for orienting the designers in the use of the power-gating technology.

## References

[1].    F. Fallah, M. Pedram, "Standby and Active Leakage Current Control and Minimization in CMOS VLSI Circuits*," IEICE Trans. on Electronics, Special Section on Low-Power LSI and Low-Power IP, Vol. E88-C, No. 4, pp. 509-519, Apr. 2005.*

[2]    K. Roy, et al., "Leakage Current Mechanisms and Leakage Reduction Techniques in Deep-Submicrometer CMOS Circuits," *Proceedings of the IEEE, Vol. 91, No. 2, pp. 305-327, Feb. 2003.*

[3]    H. Jiang, M. Marek-Sadowska, S. R. Nassif, "Benefits and Costs of Power-Gating Technique," *ICCD-05, pp. 559-566, Oct. 2005.*

[4]    K. Shi, D. Howard, "Challenges in Sleep Transistor Design and Implementation in Low-Power Designs," *DAC-06, pp. 113-116, Jun. 2007.*

[5]    A. Sathanur, A. Pullini, L. Benini, A. Macii, E. Macii, M. Poncino, "Timing Driven Row-Based Power Gating," *ISLPED-07, pp. 104-109, Aug. 2007.*

[6]    A. Sathanur, A. Calimera, L. Benini, A. Macii, E. Macii, M. Poncino, "Efficient Computation of Discharge Current Upper Bounds for Clustered Sleep Transistor Sizing," *DATE-07, pp. 1-6, Apr. 2007.*