

Simultaneous Control of Sub threshold and Gate Leakage Current in Nanometre-Scale CMOS Circuits

Y.Kishore, U.Vankata Sivaiah

Abstract— Leakage power has become a serious concern in Nanometre CMOS technologies, and power-gating has shown to offer a viable solution to the problem with a small penalty in performance. This paper focuses on leakage power reduction through automatic insertion of sleep transistors for power-gating. In particular, we propose a novel, layout-aware methodology that facilitates sleep transistor insertion and virtual-ground routing on row-based layouts. We also introduce a clustering algorithm that is able to handle simultaneously timing and area constraints, and we extend it to the case of multi-sleep transistors to increase leakage savings. The results we have obtained on a set of benchmark circuits show that the leakage savings we can achieve are, by far, superior to those obtained using existing power-gating solutions and with much tighter timing and area constraints. Leakage power is a major concern in sub-90-nm CMOS technology. The exponential increase in the leakage component of the total chip power can be attributed to threshold voltage scaling, which is essential to maintain high performance in active mode, since supply voltages are scaled. Numerous design techniques have been proposed to reduce standby leakage in digital circuits. Out of this rich set of solutions, power gating has proven to be a very effective approach to minimize standby leakage while keeping high speed in the active mode. It is based on the principle of adding devices, called sleep transistors in series to the pull-up and/or the pull-down of the logic gates, and turning them off when the circuit is idle, thereby decreasing the leakage component due to IDS sub-threshold currents. When an NMOS sleep transistor is used on the pull down path, a SLEEP signal controls its active/standby mode.

Index Terms— Low power, sleep transistor, power gating

I. INTRODUCTION

Several power-gating styles have been proposed, differing in the *granularity* of the blocks that the sleep transistors have to control. Granularity may range from individual cells (the *fine-grained* sleep transistor insertion approach) to large chip sub-units, in which very large sleep transistors are placed on the root of the power distribution networks of large chip areas. While the fine-grained approach suffers from high area overhead and excessive buffering of the sleep signals, power-gating applied to large chip sub-units has the disadvantage of having long transition delays between sleep and active state, caused by the large resistance-capacitance

(*RC*) time constant of the sub-unit's power distribution network and large IR drop on virtual ground rails. Multi-threshold CMOS is a valuable leakage reduction method in circuit standby mode. Reducing leakage current through fine grain sleep transistor insertion (*FGSTI*) makes it easier to guarantee circuit functionality and improves circuit noise margin. In this paper, we first indicate the negligible dependence of ST size on the amount of leakage saving which makes the two-phase *FGSTI* reasonable based on our leakage current and delay models. Then we introduce a novel two-phase *FGSTI* technique: a) ST placement and b) ST sizing, which are formally modelled as two linear programming (LP) models respectively. Our experimental results show that the two-phase *FGSTI* technique can achieve 78.91%, 92.55%, 97.97% leakage saving when the circuit slowdown is 0%, 3%, 5% respectively. Comparing to the simultaneous ST placement and sizing method using mix integer linear programming (MLP) [1], our technique leads to on average 2% more leakage current reduction while at least 10X runtime saving since fewer variables and constraints with less approximation are used in the LP models.

When the Sleep transistors in industrial power-gating designs are custom designed with an optimal size. Consequently, sleep transistor P/G network optimization becomes a problem of finding the optimal number of sleep transistors and their placement as well as optimal P/G network grids, wire widths and layers. This paper presents a fake via based sleep transistor P/G network synthesis method, which addresses the requirements from industrial power-gating designs. The method produces optimal sleep transistor P/G networks by simultaneously optimizing sleep transistor insertion and placement as well as the power network grids and wires for minimum area, maximum rout ability with a given IR-drop target. The most popular MTCMOS technique is gating the power of sizable blocks using large sleep transistors which is concluded as *block based ST insertion (BBSTI)* technique. In *BBSTI* techniques, all the gates in the block are assumed to have a fixed slowdown, so it is also called fixed slowdown method. The existing literatures on *BBSTI* techniques present some details in clustering gates into blocks in order to optimize the leakage current and ST size. All these literatures focus on how to reduce the ST area penalty along with a remarkable leakage saving: first gives out a mutual exclusion method; present several fast heuristic techniques for efficient gate clustering; propose a distributed sleep transistor network (*DSTN*) approach which assumes that all the sleep devices are connected to further reduce the area penalty. Although *BBSTI* techniques greatly reduce the area penalty, they induce large ground bounce in the P/G network which has adverse effects on circuit speed and noise immunity. What is more, ST size is determined by the worst case current of the clustering block which is quite difficult to determine without comprehensive simulation. Thus it is harder to guarantee circuit functionality for large blocks with only one ST. In recent years gate level

Manuscript received July 17, 2016

Y.Kishore, Final M.Tech student, ECE, IITS, A.P.

U.Vankata Sivaiah, Asst.Professor, E.C.E, IITS, A.P, INDIA.

ST insertion, which can be also called *fine-grain ST insertion (FGSTI)* technique. It is easier to guarantee circuit functionality in an *FGSTI* technique as ST sizes are not determined by the worst case current of large circuit blocks. And *FGSTI* technique leads to a smaller simultaneous switching current when the circuit changes between standby mode and active mode, thus improves circuit noise margins. Furthermore, better circuit slack utilization can be achieved as the slowdown of each gate is not fixed, and then leads to a further reduction of leakage and area. As shown in, *FGSTI* technique corresponds to an area penalty of roughly 5% using standard cell placement. Cell-based sleep transistor implementations were proposed, where each cell had a built-in sleep transistor. In this case the sizing became a small scale local problem and was based on the cell's worst case current and timing criticality and temporal currents. One of the major issues of the cell-based implementations is the large area penalty introduced by adding a sleep transistor in every cell. Another issue in cell-based sleep transistor implementations is the increased design sensitivity to PVT variations due to the power supply variations in individual cells introduced by sleep transistors. Currently, most, if not all, industrial power-gating designs adopt distributed sleep transistor network implementation, where sleep transistors are connected between the permanent power supply and the virtual power supply networks as shown in Fig. 1.

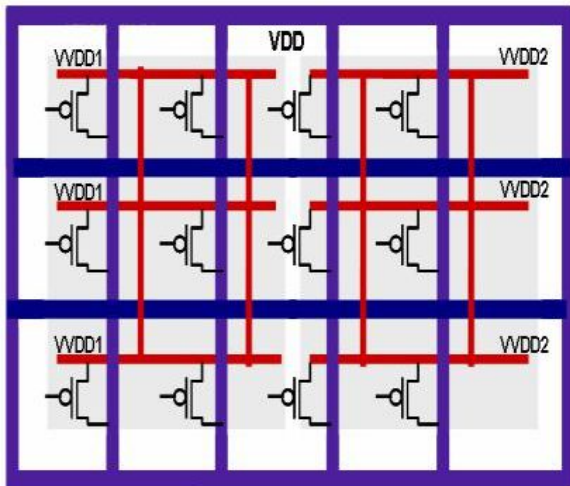


Fig.1: Distributed sleep transistor implementations

The main advantage of the distributed sleep transistor implementation is the ability to share current charge or discharge among the sleep transistors. Consequently, it is less sensitive to PVT variation and introduces smaller IR drop variations than the cell-based and cluster-based implementations. The sleep transistor sharing also reduces the area overhead significantly. Power/Ground (P/G) network synthesis becomes a challenge in the distributed sleep transistor implementations because the P/G network consists of three components: a permanent power network, a virtual power network and an array of the sleep transistors that connect the permanent and virtual power networks. All of these components contribute to the quality of the sleep transistor P/G network in terms of IR drops, routing. A number of P/G network synthesis methods for the distributed sleep transistor P/G network have been reported. In these methods, the permanent and virtual power networks are generated by conventional power network synthesis

methods. Then, the sleep transistors are inserted and sized based on current drawn and IR-drop requirement. The sleep transistor insertion is defined by the user, based on either circuit clusters or design heuristic. The sleep transistor is sized based on the current through the sleep transistor branches to satisfy the IR-drop target. The IR-drops and the current of the sleep transistor branches are calculated by conventional P/G resistive network methods. The work reported in moved the sleep transistor P/G network synthesis a step forward by simultaneously sizing sleep transistors and P/G network wires using the sequential linear programming method. The sleep transistor insertion points are determined based on cut sets of P/G branches that disconnect all cells from the power supply. Such a cut-set can be found by getting all P/G branches connected to a cell. The size of the sleep transistor of each cut-set is determined based on the current of the cells in the cut-set and the IR-drop target using a constant sheet channel resistance to model the sleep transistor's drive. The optimization variables include the resistances of not only the P/G wire branches but also the sleep transistor branches. However, the method still relies on the pre-defined or pre-synthesized power network grids. The number of the sleep transistors and their insertion positions are defined by the user. Recently, a delay degradation effect based power gating network synthesis method was proposed. In this method, a simple close-form analytic equation was proposed to model the delay degradation effect on a design due to the IR-drop on the sleep transistors. Differing from other methods, the method sizes the sleep transistors based on delay degradation constraints, namely trying to reduce sleep transistor size and hence leakage while meeting the design speed target considering the sleep transistor introduced delay degradation.

II. SLEEP TRANSISTOR P/G NETWORK MODEL:

The sleep transistor power network consists of a permanent power (VDD) network, a virtual power (VVDD) network and distributed sleep transistors that connect the two networks. The VDD and VVDD networks can be represented by two resistive networks as show in Fig. 2.

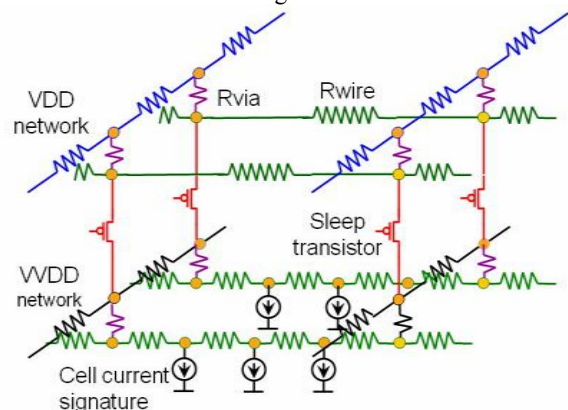


Fig.2: Sleep transistor power network

The branch resistance R_{wire} of the network is defined based on the wire length, width and thickness of the branch segment as follows:

$$R_{wire} = \rho * l_s / w_s$$

Where l_s and w_s are the length and the width of a branch segment of the network respectively. ρ is the sheet resistance

per square R_{via} is via array resistance. The current source at a VVDD network node is the worst-case current of the cells connected to the node. Although the current signature of a cell is dynamic in nature, namely, it varies with time as signal transiting, the worst-case average current signatures are commonly used in industrial P/G network synthesis, because accurate dynamic current is not available at the power planning stage. For the same reason, the sleep transistor power network is synthesized based on the worst-case average cell currents to meet user defined IR drop and EM targets. The dynamic IR drops are controlled by on-chip decoupling capacitor insertion and placement. In a power-gating design, active cells receive power supply from the virtual power network through the sleep sleep transistors are conducting in operating mode, the virtual and permanent power networks effectively become a single network. To achieve an optimal sleep transistor power network, the sleep transistor distribution and the permanent and virtual power networks should be optimized simultaneously and as a whole in the sleep transistor power network synthesis.

Sleep transistors can be inserted following many different schemes, depending on the target technology and the available library of sleep transistors. In practice, however, the usable insertion architectures are limited to two kinds: ring and grid style. *Ring Style Sleep Transistor Insertion:* In this approach, the sleep transistors are placed as a ring around the block to be power-gated. The sleep transistors connect the real supply and ground lines present on the outer ring to the virtual ground mesh inside the ring. This solution features a simple power plane and has a small impact on placement and routing in the standard cell area. On the negative side, ring style insertion is applicable only if the entire block is power-gated, since the cells inside the block do not have access to the power-supply during the power-down mode. *Grid Style Sleep Transistor Insertion:* In this scheme, sleep transistors are placed either in columns, or in dedicated rows inside the standard cell design. The grid style approach has advantages over the ring style in the sense that it requires less sleep transistor area, it provides access to the power-supply and ground lines inside the standard cell design and it offers a lower IR drop, since the sleep transistors are closer to the standard cells to which they are connected. On the negative side, grid style insertion (especially the column-based version) disrupts the placement and routing of the original placed circuit, thus complicating design convergence. For our experiments, we have chosen the row-based, grid style insertion scheme; in order to make it applicable in practice, we have addressed some of the open issues posed in regarding this type of sleep transistor insertion.

In particular, as follows. We use a power-grid for power supply distribution, with a modified grid in which we interlace also the virtual ground line between the power-supply and the ground lines, thereby providing accessibility of all the supply lines to all the rows of the layout. • Since our clustering approach flags each ground line in lower metal layer (horizontal supply lines) either as ground or virtual ground, our approach naturally allows a mix of power-down and always-on cells in the circuit.

The existing literature on MTCMOS circuits present clustering based approaches for sleep transistor insertion and sizing in circuits. Clustering sleep transistors reduces the area penalty and leakage power but has adverse effects on circuit performance due to the virtual ground bounce problem

associated with clustering. Since many logic gates share a common sleep transistor, the virtual ground bounce can have a severe effect on gate speed and noise immunity. The RC time constant of the associated virtual ground line can be used to address this problem through accurate sleep transistor sizing which is not possible if the sleep transistor is shared between several logic gates. also discusses the presence of reverse conduction paths in clustered MTCMOS logic blocks which degrade the noise margins and could eventually cause the circuit to fail logically. tries to limit these adverse effects in their proposed approach at the cost of increased leakage and dynamic power by accepting a certain level of speed and noise degradation. Other works do not address this problem in their proposed approaches. In this work, we propose a fine-grained sleep transistor placement methodology which allows us to control the size of the sleep transistor placed at each gate and eliminate the noise immunity drawbacks associated with clustering. We evaluated the virtual ground bounce problem using the Cadence custom IC design tool with 0.18 micron technology library. We built a chain of custom made inverters as a test benchmark. Three test cases, one with clustered sleep transistors, one with fine-grained sleep transistors and the third with no sleep transistors were evaluated for performance. We observed the signal at the output of one of the intermediate inverters in the benchmarks. The signal after the high to low transition. We can see that the signal from the fine-grained methodology maintain its noise immunity similar to the case with no sleep transistor whereas the signal from the clustered scheme is noisy and reduces the noise immunity. Thus our fine-grained sleep transistor placement and sizing scheme preserves the noise immunity of the logic gates while reducing the leakage power. Potentially a fine-grained sleep transistor placement approach could suffer from an area explosion. In this work, we have shown that there is no area explosion in our proposed approach as compared with the existing clustering based approaches. In , a DSTN approach is presented which does clustering based sleep transistor insertion. They evaluate the area penalty of their scheme using a Custom Layout Design. However, they do not extend the DSTN approach to a standard cell based design methodology which could result in a significant routing overhead. a standard cell layout is considered to assess the area penalty cause by the sleep transistor insertion. They propose to create a row of sleep transistors below each row of standard cells.

III. SIMULATIONS AND CONCLUSION

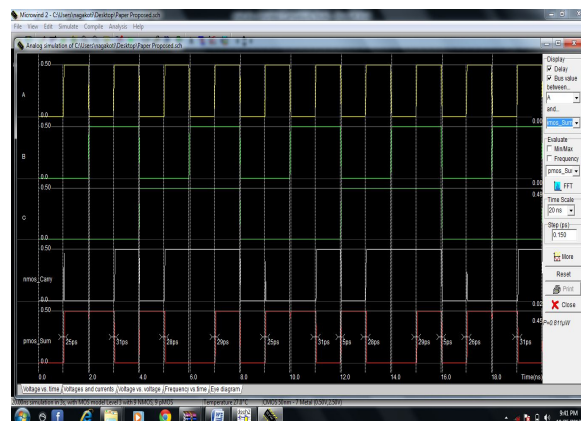


Fig. 3: Analysis for voltage vs time of Design

The Fig.3 shows the Analysis for voltage vs time of Design2. In this case active power is $0.81\mu\text{W}$. The power is reduced by 83.64% compare to Base case.

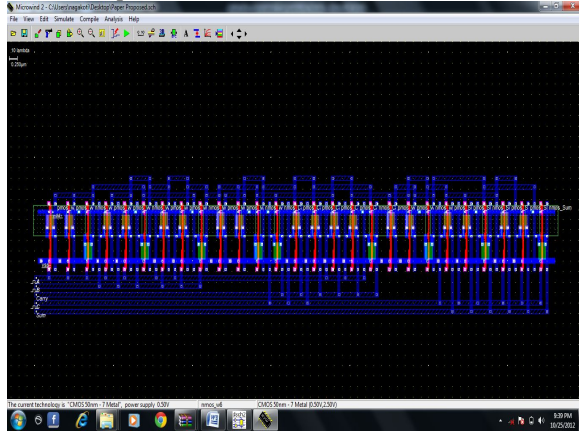


Fig.4: Layout for design2

The fig.4 shows the Layout for Design 2, Micro wind is used to draw the MOS layout and simulate its behaviour. The layout window features a grid, scaled in lambda (λ) units. The lambda unit is fixed to half of the minimum available lithography of the technology. The technology is a CMOS 7-metal layers, 50nm.

A new, layout-aware power gating methodology for leakage power reduction in nano meter CMOS circuits. Our methodology allows row-based, clustered power-gating and it features minimal perturbation of the original layout compared to existing sleep transistor insertion techniques. This favours fast design closure and makes the methodology suitable for the implementation as a CAD tool. We have introduced a clustering flow which considers both timing and area constraints, and proposed algorithms for achieving maximum leakage savings under such constraints. Finally, we have developed a multi- sleep transistor synthesis technique which enables to further reduce the total leakage w.r.t. the case of high- sleep transistors. All the algorithms have been integrated into a commercial tool flow, which has allowed us to run experiments on realistic test cases using an industrial, 65-nm CMOS technology. The leakage savings we have achieved are excellent; this proves the effectiveness

REFERENCES

- [1]Y. Wang, H. Lin, H.Z. Yang, R. Luo, H. Wang,"Simultaneous Finegrain Sleep Transistor Placement and Sizing for Leakage Optimization," in Proc. of ISQED'06, 2006, pp. 723-728.
- [2]G. Moore, "No exponential is forever: But forever can be delayed," in IEEE ISSCC Dig. Tech. Papers, 2003,pp 20-23.
- [3]D. Duarte, N. Vijaykrishnan, M. J. Irwin, and M.Kandemir,"Formulation and validation of an energy dissipation model for the clock generation circuitry and distribution networks," in Proc. Of VLSI Design, 2001,pp: 248 - 253.
- [4]J. Kao, S. Narendra, A. Chandrakasan, "Sub threshold Leakage modeling and reduction techniques", in Proc. of ICCAD, 2002, pp 141 – 149.
- [5]K. Roy, S. Mukhopadhyay, H. Mahmoodi-Meimand, "Leakage Current Mechanisms and Leakage Reduction Techniques in Deep-Submicrometer

CMOS Circuits", in Proc. of the IEEE, Vol. 91, No.2, February 2003 pp 305-327.

S.Narendra et.al, "Forward body bias for microprocessors in 130-nm technology generation and beyond", in *IEEE JSSC*, Vol. 38 , No. 5 ,May 2003 pp. 696 - 701.

C.H. Kim, K. Roy, "Dynamic VTH scaling scheme for active leakage power reduction", in *Proc. of DATE*

2002 pp.163 - 167.

S. Mukhopadhyay et. al., "Gate Leakage Reduction for Scaled Devices Using Transistor Stacking", in *IEEE TVLSI*, Vol. 11, No. 4.