A Combined Gate Replacement and Input Vector Control Approach for Leakage Current Reduction

Lin Yuan and Gang Qu

***Abstract*—Input vector control (IVC) is a popular technique for leakage power reduction. It utilizes the transistor stack effect in CMOS gates by applying a minimum leakage vector (MLV) to the primary inputs of combinational circuits during the standby mode. However, the IVC technique becomes less effective for circuits of large logic depth because the input vector at primary inputs has little impact on leakage of internal gates at high logic levels. In this paper, we propose a technique to overcome this limitation by replacing those internal gates in their worst leakage states by other library gates while maintaining the circuit's correct functionality during the active mode. This modification of the circuit does not require changes of the design flow, but it opens the door for further leakage reduction when the MLV is not effective. We then present a divide-and-conquer approach that integrates gate replacement, an optimal MLV searching algorithm for tree circuits, and a genetic algorithm to connect the tree circuits. Our experimental results on all the MCNC91 benchmark circuits reveal that 1) the gate replacement technique alone can achieve 10% leakage current reduction over the best known IVC methods with no delay penalty and little area increase; 2) the divide-and-conquer approach outperforms the best pure IVC method by 24% and the existing control point insertion method by 12%; and 3) compared with the leakage achieved by optimal MLV in small circuits, the gate replacement heuristic and the divide-and-conquer approach can reduce on average 13% and 17% leakage, respectively.**

***Index Terms*—Gate replacement, leakage reduction, minimum leakage vector (MLV).**

I. Introduction

As THE VLSI technology and supply/threshold voltage continue scaling down, leakage power has become more and more significant in the power dissipation of today's CMOS circuits. For example, it is projected that subthreshold leakage power can contribute as much as 42% of the total power in the 90-nm process generation [11]. Many techniques thus have been proposed recently to reduce the leakage power consumption. Dual threshold voltage process uses devices with higher threshold voltage along noncritical paths to reduce leakage current while maintaining the performance [16]. Multiple-threshold CMOS (MTCMOS) technique places a high $V_{\rm th}$ device in series with low $V_{\rm th}$ circuitry, creating a sleep transistor [13]. In dynamic threshold MOS (DTMOS) [3], the gate and body are tied together and the threshold voltage is altered dynamically to suit the operating state of the circuit. Another technique to dynamically adjust threshold voltages is the variable threshold CMOS (VTCMOS) [14]. All of these approaches require the process technology support.

(a)

INPUT	Leakage(nA)
0	best:100.3
1	worst:**227.2**

(b)

INPUT	Leakage(nA)
00	best: 37.84
01	2nd worst: 100.30
10	95.17
11	worst: **454.50**

(c)

INPUT	Leakage (nA)
000	best: 22.84
001	37.84
010	37.84
011	2nd worst: 100.30
100	37.01
101	95.17
110	94.87
111	worst: **852.40**

Fig. 1. Leakage current of (a) INVERTER, (b) NAND2, and (c) NAND3. Data obtained by simulation in cadence spectre using 0.18-μm process.

The input vector control (IVC) technique is applied to reduce leakage current at circuit level with little or no performance overhead [7]. It is based on the well-known transistor stack effect: a CMOS gate's subthreshold leakage current varies dramatically with the input vector applied to the gate [10]. Recently, Lee *et al.* observed that gate oxide leakage is also dependent on the input vectors to a CMOS gate [12]. Besides, the maximal and minimal leakage vectors are the same for both subthreshold leakage and gate leakage. In our study, we use Cadence Spectre to measure the overall leakage current in a CMOS gate that includes both subthreshold leakage and gate leakage. Fig. 1 lists the overall leakage current in INVERTER, NAND2 and NAND3 gates under all the possible input combinations. We see that the worst case leakage (marked in bold) is much higher than the other cases. The idea of IVC technique is to manipulate the input vector with the help of a sleep signal to reduce the leakage when the circuit is at the standby mode [9]. The associated minimum leakage vector (MLV) problem seeks to find a primary input vector that minimizes the total leakage current in a given circuit. [1], [4] [6], [8]–[10], [15]. The MLV problem is NP-complete and both exact and heuristic approaches have been proposed to search for the MLV. A detailed survey is given in Section II.

In this paper, we consider how to enhance IVC technique with little or no re-design effort. In particular, we study the **MLV+ problem** that seeks to modify a given circuit and determine an input vector such that the circuit's functionality is maintained at the active mode and the circuit leakage is minimized when the circuit is at standby mode. Our solution to this problem is based on the concept of gate replacement that is motivated by the large discrepancy between the worst leakage and the other cases (see Fig. 1). The essence of gate replacement is to replace a logic gate that is at its worst leakage state (WLS) by another library gate. This is illustrated by the following example.

Manuscript received May 28, 2005; revised October 15, 2005.

The authors are with the Electrical and Computer Engineering Department and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742 USA (e-mail: yuanl@eng.umd.edu).

Digital Object Identifier 10.1109/TVLSI.2005.863747

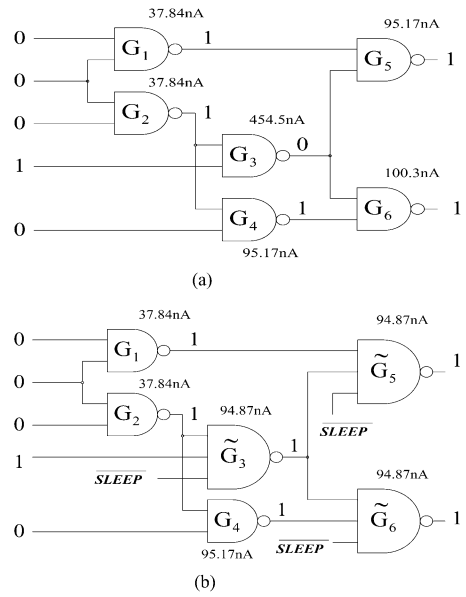

Fig. 2. Motivation example for gate replacement. (a) Original MCNC benchmark circuit C17 with total leakage 831.08 nA under the optimal MLV. (b) New circuit C17 with three gates replaced and total leakage 476.88 nA under the same MLV.

Consider circuit C17 from the MCNC91 benchmark suite [21] [Fig. 2(a)]. An exhaustive search finds the MLV $\{0, 0, 0, 1, 0\}$, with the corresponding minimum total leakage current of 831.08 nA. Note that gate G_3 has its worst leakage current (454.5 nA) with input $\{1, 1\}$, which contributes more than half of the total leakage. In fact, we have observed that a significant portion of the total leakage is often caused by the gates that are in their WLS (see Table II in Section V).

Instead of controlling the primary inputs, we consider replacing these leakage-intensive gates. In particular, we replace the NAND2 gate G_3 by a NAND3 $\tilde{G}_3$, where the third input $\overline{SLEEP}$ is the complement of the *SLEEP* signal [Fig. 2(b)]. At active mode, $\overline{SLEEP} = 1$ and $\tilde{G}_3$ produces the same output as G_3. But at the standby mode, $\overline{SLEEP} = 0$ and $\tilde{G}_3$ has a leakage of 94.87 nA [Fig. 1(b)], which is much smaller than G_3's 454.5 nA.

However, this replacement also changes the output of this gate at the sleep mode and affects the leakage on gates G_5 and G_6. In this case, we replace them in a similar fashion. As a result, the new circuit's total leakage becomes 476.88 nA, a 43% reduction from the original 831.08 nA in Fig. 2(a).

The proposed gate replacement technique is conceptually different from the existing IVC methods. In fact, they are complementary to each other. Specifically, IVC method considers the entire circuit and searches for an appropriate input vector in favor of small leakage. The gate replacement technique targets directly at the logic gates that are in their WLS under a specific input vector and replace them to reduce leakage. This paper has the following contributions.

1) We examine the effectiveness of IVC methods[1] in multilevel circuits. For all the 69 MCNC91 benchmarks, we obtain the optimal MLV for small circuits and the best over 10 000 random input vectors for large circuits. The number of gates in their WLS are on average 15% and 17%, respectively, but they contribute more than 40% of the circuit's total leakage.
2) Motivated by the above observation, we propose the technique to replace gates that are in their WLS by other library gates that will generate less leakage current at those states. Unlike other leakage reduction techniques such as MTCMOS and DTMOS, this modification of the circuit does not require changes of process technology in the design flow. Hence, it will not increase the design complexity or the leakage sensitivity.
3) We implement a fast gate replacement algorithm that gives an average of 10% leakage reduction for a fixed input vector. This algorithm's run time complexity is linear to the number of gates in the circuit in average cases and quadratic in the worst case.
4) We develop a divide-and-conquer approach to combine gate replacement and IVC. It reduces the leakage by 17% and 24% over the optimal/suboptimal MLV mentioned in 1) with little area and delay overhead. The number of gates in their WLS is dropped to 4% and 9%, respectively.

[1]IVC-based approaches such as internal control point insertion [1] will be discussed in Section II.

II. Related Work

In this section, we mainly survey the efforts on IVC-based leakage reduction techniques. A survey on other leakage minimization techniques can be found in [7].

The effect of circuit input logic values on leakage current was observed by Halter and Najm [9]. The underlying reason of this effect was explained by Johnson *et al.* [10] as the transistor stack effect. Authors in [9] proposed a technique to insert a set of latches with MLV stored in to the primary inputs of a circuit, forcing the combinational logic into a low-leakage state when the circuit is idle. Many algorithms have been proposed to find such MLV. Based on the nature of these algorithms, they can be classified into the following groups:

Heuristic Algorithms: These include the random search algorithm developed by Halter and Najm [9] and the genetic algorithm proposed by Chen *et al.* [5].

Johnson *et al.* [10] defined *leakage observability* for each primary input as the degree to which the value of a particular input is observable in the magnitude of leakage current. They iteratively chose the input with the largest leakage observability and assigned it with a value that results in the smallest leakage. The input combination constructed in this greedy fashion was taken as the MLV.

In [15], Rao *et al.* introduced the concept of *node controllability*, which is defined as the minimum number of inputs that have to be assigned to particular values to ensure that a node (or gate) is in a specific state. Based on this, they proposed a fast greedy heuristic to determine the values of the primary inputs that minimize the node's leakage.

Exact Algorithms: The MLV problem can be modeled as a pseudo-Boolean satisfiability (SAT) problem. This formulation allows us to apply the off-the-shelf SAT solvers to find the MLV for leakage reduction [1], [2].