Chapter 5
The HL Approach: A Low-Leakage ASIC Design Methodology

5.1 Overview

One of the most popular ways of reducing leakage is through the use high-$V_T$ power gating transistors (as in the MTCMOS technique [8,13] mentioned in Chap. 2). The HL approach is a variant of this technique that uses these power gating transistors selectively. In the HL approach we first create two low-leakage variants of each cell in a standard-cell library. If the inputs of a cell during the standby mode of operation are such that the output has a high value, we minimize the leakage in the pull-down network. Similarly we minimize leakage in the pull-up network if the output has a low value. In this manner, two low-leakage variants of each standard cell are obtained. While technology mapping a circuit, we determine the particular variant to utilize in each instance, so as to minimize leakage of the final mapped design.

This chapter is organized as follows. The philosophy of the HL approach is explained in Sect. 5.2. Related previous work is discussed in Sect. 5.3. In Sect. 5.4, details of the HL approach are presented. In Sect. 5.5, we present experimental results that compare placed-and-routed area, leakage and delay of this new methodology against MTCMOS and a regular standard-cell-based design style. The results show that the HL approach has better speed and area characteristics than MTCMOS implementations. The leakage current for HL designs can be dramatically lower than the worst-case leakage of MTCMOS-based designs and two orders of magnitude lower than the leakage of traditional standard cells. An ASIC design implemented in MTCMOS would require the use of separate power and ground supplies for latches and combinational logic, while our methodology does away with such a requirement. Another advantage of our methodology is that the leakage is precisely estimable, in contrast with MTCMOS. The primary contribution of the work presented in this chapter is a new low-leakage design style for static CMOS designs.

In Sect. 5.6, we present some experiments that explore the feasibility of using gate length biasing (minor changes to the channel length of a transistor) instead of changing the $V_T$. In Sect. 5.7, we discuss techniques to reduce leakage in dynamic (domino logic) designs and a summary is presented in Sect. 5.8.
5.2 Philosophy of the HL Approach

The leakage current for a PMOS or NMOS device corresponds to the $I_{ds}$ of the device when the device is in the cut-off or sub-threshold region of operation. The expression for this current [1] is:

$$I_{sub} = \frac{W}{L} I_{D0} e^{\frac{V_{gs}-V_{T}-V_{off}}{n \cdot v_t}} \left[ 1 - e^{-\frac{v_t}{V_{ds}} \cdot v_t} \right]$$  \hspace{1cm} (5.1)

Here $I_{D0}$ and $V_{off}$ (typically $V_{off} = -0.08V$) are constants, while $v_t$ is the thermal voltage (26 mV at 300°K) and $n$ is the sub-threshold swing parameter.

We note that $I_{ds}$ increases exponentially with a decrease in $V_T$. This is why a reduction in supply voltage (which is accompanied by a reduction in threshold voltage) results in exponential increase in leakage.

Another observation that can be made from (5.1) is that $I_{ds}$ is significantly larger when $V_{ds} \gg n v_t$. For typical devices, this is satisfied when $V_{ds} \approx VDD$. The reason for this is not only that the last term of (5.1) is close to unity, but also that with a large value of $V_{ds}$, $V_T$ would be lowered due to drain-induced barrier lowering – DIBL ($V_T$ decreases approximately linearly with increasing $V_{ds}$) [1, 15]. Therefore, leakage reduction techniques should ensure that the supply voltage is not applied across a single device, as far as possible.

Our approach to leakage reduction attempts to ensure that the supply voltage is applied across more than one turned-off device and one of those devices is a high-$V_T$ device. This is achieved by selectively introducing a high-$V_T$ PMOS or NMOS supply gating device in either the pull-up network of a gate (if the output is low in standby) or the pull-down network of a gate (if the output is high in standby). By this design choice, we obtain standard cells with both low and predictable standby leakage currents, unlike MTCMOS-based approaches.

5.3 Related Previous Work

Previous design approaches have suggested the use of dual-threshold devices [8] in an MTCMOS configuration which MTCMOS utilizes NMOS and PMOS power supply gating devices. The authors propose a MTCMOS standby device sizing algorithm, which is based on mutually exclusive discharging of gates. This technique is hard to utilize for random logic circuits as opposed to the extremely regular circuits, which are used as illustrative examples in [8]. In [13], the authors describe an MTCMOS implementation of a PLL using a 0.5-μm process. In both these works, the problem of estimating the leakage of an MTCMOS design is not addressed. In practice, the leakage of such a design can vary widely and is hard to control or predict. The threshold voltage is modified by bulk bias (via body effect) and DIBL, which are determined in part by the voltages of the bulk/source and source/drain nodes. Since cell inputs and outputs as well as bulk nodes float in an MTCMOS design.