More Than Two Orders of Magnitude Leakage Current Reduction in Look-Up Table for FPGA's

Canh Q. Tran, Hiroshi Kawaguchi and Takayasu Sakurai
Institute of Industrial Science, University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505 Japan
{canh, kawapy, tsakurai}@iis.u-tokyo.ac.jp

Abstract: A leakage current reduction scheme based on ZSCCMOS (zigzag super cutoff CMOS) is proposed for a LUT (look-up table) using input forcing and an overdriven power gates. A fabricated chip demonstrates that the leakage current of the LUT can be reduced by more than 2 orders of magnitude. The wake-up time of the proposed LUT is 10 times as short as that of the LUT using SCCMOS (super cutoff CMOS). The area and delay overheads are 15% and 8%, respectively.

I. Introduction

FPGA's (field programmable gate arrays) are very attractive because of their low NRE (non-recurring engineering) cost and short time-to-market [1], although the FPGA's use much more transistors per function than ASIC in order to achieve programmability. This results in higher leakage even though low leakage characteristics are getting important on a current market. The leakage current is increasing due to low VTH used in scaled technologies. In a 65-nm technology with VDD=0.7V, VTH=0.1V, a leakage current of a three-input LUT (look-up table) could be above 10µA. Since a common FPGA's chip can have 10^5 LUTs today, a chip fabricated in the 65-nm CMOS technology node would draw an unacceptable leakage current of 1A.

Since FPGA's usually have 20X-40X area overhead and 3X-10X delay overhead compared to a custom logic circuit, most of previous FPGA's studies have mainly focused on the area and performance, and thus, there have been little works on low-power FPGA's [2-4]. In [2-4], dynamic power is considered assuming that the leakage power is a small, which does not hold any more. Recently, there have been a few works on leakage power in FPGA's [5-7]. In [5], some low-leakage design techniques for FPGA's are evaluated, and the gate biasing, use of redundant SRAM cells, and integration of multi-threshold voltage technology are shown to reduce leakage current from 1/2 to 1/4. In [6], it is shown that a leakage current of FPGA's depends on its inputs for fractional reduction of leakage current. The other work proposed a method to reduce a leakage current by 25% on average [7].

On the other hand, there have been extensive studies on low-leakage design techniques for custom logic design, such as SCCMOS (super cutoff CMOS) [8] and MTCMOS (multi-threshold CMOS) [9]. These techniques can cut off effectively leakage current but suffer from a long wake-up time, which means the time to recover from a standby mode to an active mode. The ZSCCMOS (zigzag super cutoff CMOS) can suppress leakage current by overdriving power gates while having a short wake-up time that is less than a fraction of a clock cycle [10]. Thus ZSCCMOS can be a substitute technique for clock gating, being used even in active mode. The measured wake-up time of a 16-bit Brent-Kung adder with ZSCCMOS is 16% of the cycle time [11], which is ten times as short as the fastest wake-up time ever reported [12].

In this paper, the ZSCCMOS technique is adopted for a LUT to cut off leakage current and obtain shorter wake-up time. The straight-forward application of the ZSCCMOS is not possible because of the sneak paths in the multiplexer network in the LUT's. Although the technique is applied to the LUT in FPGA's in this paper, the concept is effective for general multiplexers excluding sneak path problem.

II. Novel Leakage-Reduced LUT Design

Figure 1 shows a three-input LUT. The SRAM cells (SC0~SC7) just keep table information, and do not need to be fast. Therefore, SRAM cells can use high VTH and the leakage current for the SRAM cells can be ignored. When some outputs of the SRAM cells are high and some are low level, a leakage current problem occurs in the LUT since the leakage current flows not only through the inverters in the circuit but also through transmission gates network in the LUT.

In the proposed ZSCCMOS scheme, the INV's near the SRAM cells (SC0~SC7) just keep table information, and do not need to be fast. Therefore, SRAM cells can use high VTH and the leakage current for the SRAM cells can be ignored. When some outputs of the SRAM cells are high and some are low level, a leakage current problem occurs in the LUT since the leakage current flows not only through the inverters in the circuit but also through transmission gates network in the LUT.

In the proposed ZSCCMOS scheme, the INV's near the SRAM cells are replaced to NAND's. The outputs of the NAND's are forced to the same level as shown in Figure 2 by using Srby-bar signal. This is called phase forcing technique.
Since all outputs of the NAND's are H, the leakage current shown in Figure 1 does not flow any more. The leakage current of the NAND's and INV's are cut off by using the ZSCCMOS scheme.

Figure 1. Conventional LUT. SC stands for SRAM cell.

Figure 2. Proposed LUT. SC stands for SRAM cell.

Figure 3 shows the circuits of NAND's, INV's that use the ZSCCMOS scheme. The drain voltages of the cutoff transistors MN and MP are $V_{DDV}$ and $V_{SSV}$ respectively. They are neither at $V_{DD}$ nor at $V_{SS}$ but stand between $V_{DD}$ and $V_{SS}$. Thus, the voltage of $V_{SS}-V_{OD}$ and $V_{DD}+V_{OD}$ could be assigned to gates of $NCUT$ and $PCUT$ respectively with no over-stress for power switches, where $V_{OD}$ is an overdrive voltage. Figure 4 shows the simulation result using the BPTM (Berkeley predictive technology model) for 65-nm CMOS, where $V_{DDV}=0.7V$ and $V_{TH}=0.15V$ [13]. When $V_{OD}$ is set to 0.15V, the leakage current is reduced by 96X. This leakage current reduction is up to the cutoff of the leakage current that flows through transmission gates and leakage current of INV's.

If the frequencies of the LUT input signals ($IN_0$-$IN_2$) are set at 100MHz and the LUT is configured as a 2-bit adder, the active power of the conventional and the proposed LUT are 2.73μW and 2.94μW, respectively. Thus, the dynamic power of the proposed LUT is increased by 8% over that of the conventional one. In the leakage dominant era, assuming the leakage power makes up only 50% of the total power, with this increased dynamic power the proposed scheme is still effective enough because it can reduce more than 98% of the leakage power and therefore, 45% of the total power.

Figure 3. NAND and INV that use ZSCCMOS. The cutoff transistors MN and MP are shared for all NANDs and INVs. The waveforms of the $NCUT$, $PCUT$ and virtual power lines $V_{DDV}$, $V_{SSV}$ are also shown.

Figure 4. Leakage current vs. overdriven voltage, $V_{OD}$ in LUT.

In ZSCCMOS, the power switches, MN and MP are inserted in off paths thanks to phase forcing, which means that MN does not need to recover “L” nodes nor does MP need to recover “H” nodes in a wake-up process. The virtual power lines $V_{DDV}$ and $V_{SSV}$ stand near $V_{DD}$ and $V_{SS}$ respectively. Thus, the wake-up time from the standby mode to the active mode is much shorter than SCCMOS. Figure 5 shows the
simulated waveforms of $V_{DDV}$ and $V_{SSV}$ in the proposed ZSCCMOS LUT and conventional SCCMOS LUT. The wake-up time can be shortened to 105ps while the wake-up time of the conventional SCCMOS LUT is 997ps.

Figure 5. The waveforms of $V_{DDV}$ and $V_{SSV}$ in ZSCCMOS and SCCMOS LUTs. 65-nm BPTM is used.

Figure 6 shows the wake-up time of the proposed ZSCCMOS LUT and SCCMOS LUT as the gate widths of the power switches are varied. The wake-up time of the proposed ZSCCMOS LUT is ten times as short as that of the SCCMOS LUT.

Figure 6. Simulated wake-up time of the proposed LUT and the one that uses SCCMOS scheme. $W_{CUT}$ and $W_{TOTAL}$ are the total gate width of the cutoff transistors and the whole circuits respectively. The BPTM 65nm CMOS model is used.

III. Measurement results

The proposed and conventional 3-input LUTs are manufactured to verify the effect of the leakage current reduction scheme. Since low $V_{TH}$ process was not available, the leakage emulator [14] was used to emulate the low-$V_{TH}$ process. The prototype chip is fabricated by 0.35µm triple-metal CMOS technology with high $V_{TH}$ transistor and nominal supply voltage of 3.3 volts. Figure 7 shows the relation of the leakage current and the reduced threshold voltage, $\Delta V_{TH}$ using the leakage emulator. As shown in the figure, the leakage current is increased by one order when $V_{TH}$ is reduced by 0.1V (the reduced threshold voltage $\Delta V_{TH}$ increases by 0.1V) which corresponds to the theoretical value.

Figure 7. Leakage current vs. the reduced threshold voltage using leakage emulator. 0.35µm CMOS process is used.

Figure 8 shows the relation of the leakage current and the overdrive voltage when the leakage emulator is used. $\Delta V_{TH}$ is set to 0.6V. The leakage current is reduced by 200X when $V_{OD}=0.2V$. The microphotograph of the fabricated chip is shown in Figure 9. The area overhead of the proposed LUT is 15% over the conventional one when the leakage emulator is not included. The delay overhead is simulated to be 8% as shown in Figure 10. Although the area overhead for a single LUT is 15% but for a whole FPGA, the area overhead becomes less than 5%, since the LUTs occupy only 30% of the area. The interconnection blocks, I/O and other circuits occupy 70% of the total chip. It is to be noted that the proposed approach can be applied for
multiplexers in general.

Figure 9. Microphotographs of the fabricated chip. (a) proposed LUT with the leakage emulator and (b) conventional LUT with the leakage emulator.

Figure 10. Area overhead vs. simulated delay overhead of the proposed LUT. The 0.35um CMOS process is used.

SUMMARY

In this paper a method to reduce the leakage current of a LUT in FPGA's is proposed. The proposed method can reduce the leakage current of LUT more than 2 orders of magnitude, and the area overhead for a single LUT is 15% at the performance penalty of 8%. The wake-up time of the proposed LUT is ten times as short as the conventional power-gated LUT. Thus ZSCCMOS can be used as a substitute technique for clock gating in the leakage-dominant era, being used even in an active mode.

ACKNOWLEDGEMENT

The authors thank Semiconductor Technology Academic Research Center (STARC) for valuable support. The chip fabrication is supported by VLSI Design and Education Center (VDEC), the University of Tokyo with the collaboration by Rohm Corp. and Dai Nippon Printing Corp.

REFERENCES