Look-Up Table Leakage Reduction for FPGAs

Navid Azizi and Farid N. Najm

Department of Electrical & Computer Engineering

University of Toronto, Toronto, Ontario, Canada

{nazizi,najm}@eecg.utoronto.ca

Abstract— We propose new programmable FPGA Look-up Tables (LUTs) that can operate in two different modes: high-performance or low-power. Selection between the two modes is realized by an extra SRAM cell that can be shared by a number of LUTs. In high-performance mode, the LUTs provide similar power and performance to a conventional LUT. In low-power mode, one LUT reduces leakage by 53%, while another reduces leakage by 53% and 80% when outputting a logic-0 and logic-1 respectively, which can lead to an average leakage reduction of up to 76%. In low-power mode, delay is increased by 5% to 20% compared to a conventional LUT. The technique scales well and reduces further leakage for new FPGA architectures that use larger size LUTs.

I. INTRODUCTION

Since the MOS transistor sub-threshold leakage current increases exponentially with a reduced threshold voltage, and the MOS transistor gate tunneling leakage current increases exponentially with a reduced oxide thickness [1], leakage power dissipation has grown to be a significant fraction of overall chip power dissipation in modern processes and it is expected to grow significantly in future processes [2].

Due to the increasing complexity of modern digital designs, Field-Programmable Gate Arrays (FPGAs) and other reconfigurable architectures have become an attractive implementation option. While FPGAs have continued to improve in performance and cost they are less power-efficient compared to an Application Specific Integrated Circuit (ASIC) implementation [3].

Most of the early work on low-power FPGAs [4][5][6] has focused on dynamic power consumption, however, leakage power can now compose over 50% of total FPGA power [7].

FPGAs consist of an array of logic blocks that are connected through a network of routing switches, all programmed by SRAM cells. The programmable logic blocks are composed of Look-up Tables (LUTs) and flip-flops. Recent work has concentrated on reducing the leakage within the routing switch which accounts for 60%-70% of total FPGA leakage [8][9][10]. Since the SRAM bits are not performance critical, they can be made low-leakage and the leakage within flip-flops can also be reduced [11]. The leakage of the LUTs which currently comprise 20%-30% of total chip power, however, has not been targeted. Given that recent low-leakage techniques [8][9] have reduced the leakage in the routing switches by 36% to 75%, leakage in the logic can compose anywhere from a third to a half of the total FPGA leakage. Furthermore, given the trend in some new commercial FPGAs to use a larger LUT [12] which increases the percentage of
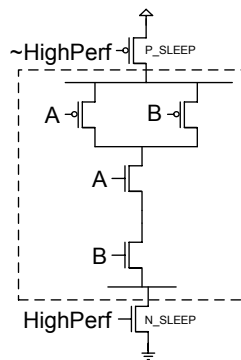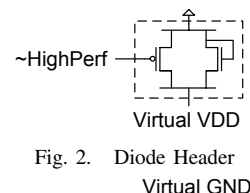

Fig. 1.   Supply Gating
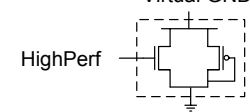

Fig. 2.   Diode Header

Fig. 3.   Diode Footer

total leakage attributed to the LUT [7] it becomes important to reduce the leakage within LUT structures.

We first apply a general leakage reduction technique to a conventional LUT; the new LUT can be programmed into two modes, high-performance or low-power. In high-performance mode there is a 1% increase in delay compared to a conventional LUT and little leakage reduction; in low-power mode leakage is reduced by up to 53%. The selection between low-power and high-performance modes can be realized through an additional SRAM cell in each LUTs or, to reduce the area overhead, a cluster of LUTs.

We then present a novel FPGA LUT that further reduces the leakage in low-power mode by another 27% leading to a total leakage reduction of 80% when the LUT is outputting a logic-1. When used with a technique [13] that skews the static probability of signals to be logic-1, the total leakage can be reduced by up to 76%. This new LUT has, however, a 3% increase in delay in high-performance mode, and thus we present an alternative design with slightly increased area and a leakage reduction of 77% when outputting a logic-1 that has no performance penalty.

The rest of this paper is organized as follows: Section II presents related work and necessary background information. Section III presents the proposed LUT design. Results for the low-power LUT are provided in Section IV, and finally Section V concludes.

II. BACKGROUND

A well-known technique to lower the leakage of logic circuits is the inclusion of sleep transistors [14] within the N-network (footer) and/or P-network (header) in CMOS gates as shown in Fig. 1, to create a virtual V_{DD} node, V_{VDD},
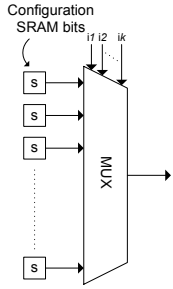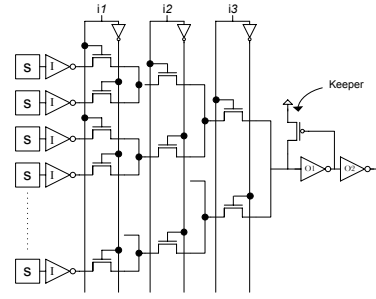
Fig. 4.   Abstract View of LUT

Fig. 5.   Transistor-Level View of LUT

Fig. 6.   Apply Diode Header to LUT

Fig. 7.   New Footer

and a virtual V_{SS} node, V_{VSS}. When the sleep transistors are ON, the gate functions as intended, and when the HighPerf signal goes low, the sleep transistors are turned OFF, inducing a stack effect and limiting the leakage through the gate. The gate, however, no longer functions as a proper logic gate.

A variation of this idea is to include diodes in parallel with the sleep transistors in the header and footer [15] as shown in Figs. 2 and 3. In high performance mode, the sleep transistors are ON, allowing for the full supply voltage to be available to the gate. In low power mode, the sleep transistors are OFF and $V_{VDD} = V_{DD} - V_{tn}$ and $V_{VSS} = |V_{tp}|$, thus reducing the voltage across the transistors within the logic block which leads to reduced gate and subthreshold leakage. While the gate does not reduce the leakage as much as if the diodes where not included, the gate is still able to function, albeit at a reduced performance. This form of supply gating has been used successfully in the header of the output buffer of an FPGA routing switch [8].

An FPGA k-LUT is composed of k input signals selecting a single output from 2^k bits as shown in Fig. 4. The SRAM configuration bits in the LUT allow for any logic function of k-inputs to be provided by the LUT. A transistor-level view of the switch can be seen in Fig. 5. The buffer at the output of the LUT is level-restoring as the weak keeper restores the voltage at the input of the buffer to V_{DD} when a weak logic-1 is passed through the pass-gate multiplexor.

III. LOW-POWER LUT DESIGN

In this section we describe how logic headers and footers, such as in Figs. 2 and 3, can be used within an FPGA LUT and the limitations of their use, and then describe a novel logic footer that provides a reduced leakage over the standard logic footer.

Among the locations within a LUT where supply gating can be used are the input buffers (inverters I in Fig. 5) and the output buffer (inverters O1 and O2). A header or footer for use in the output buffer would be of little value, since there would be increased leakage in any subsequent stages as the weak logic-1 or logic-0 which is propagated would be unable to completely turn off downstream transistors[1].

We can instead add a header for the input inverters as in Fig. 6. When using a diode header, the LUT would function

[1]This is not a problem in a routing switch, since most routing switches feed other NMOSs pass-gate structures [8].
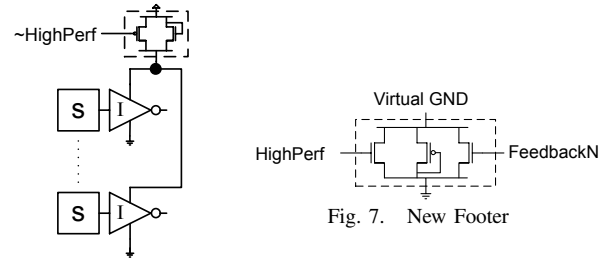
as expected in high-performance mode. In low-Power mode, there would be a reduced supply voltage across the input inverters, and thus the leakage through the input inverters and the pass-gate structure would be reduced. Furthermore, since a V_{tn} drop is already incurred through the pass-gate multiplexor, the V_{tn} drop due to the diode header would only cause a slight increase in the rise-time.

A diode footer for the input inverters, however, would cause increased leakage current in the LUT due to the weak logic-0 which would be propagated to the output buffer. There would also be a large increase in delay in low-power mode since inverter O1 would not have its PMOS conducting completely. Thus we propose a new footer design to be used in LUTs as shown in Fig 7. The new footer is a diode footer augmented by an extra transistor which is controlled by a feedback signal from the output of the multiplexor as shown in Fig. 8.

The new LUT works as so: in high-performance mode transistors NHP and PHP are ON, bypassing the effects of transistors ND, PD, and NFB, thus providing a full V_{DD} across all the input inverters and allowing the LUT to work as normal.

In low-power mode transistors NHP and PHP are turned OFF; as described above the virtual supply node would be at $V_{DD} - V_{tn}$ allowing for a reduced voltage across the input inverters thus reducing the leakage in the LUT with no impact on the functionality of the LUT as the output buffer is level-restoring.

Now let us assume that, in low-power mode, the output of the multiplexor is a logic-0, which causes transistor NFB to be ON due to inverter FN. In this case, $V_{VSS} = V_{SS}$ allowing a strong logic-0 to be fed to the output of the multiplexor. In this configuration, those inverters outputting a logic-0, there is reduced subthreshold leakage through their pull-up transistor, and for those inverters outputting a logic-1, there is a reduced subthreshold leakage through their pull-down transistor, and reduced subthreshold logic into the multiplexor, and reduced gate leakage in both of their transistors.

Now consider that, due to the LUT inputs changing, a logic-1 is fed to the output of the multiplexor. As the output of the multiplexor starts to rise, inverter FN would switch, turning transistor NFB off, completing a feedback loop. The virtual ground rail is now connected to V_{SS} through the diode connected transistor PD and is $V_{VSS} = |V_{tp}|$. Instead of full V_{DD}, $V_{DD} - |V_{tp}| - V_{tn}$ appears across the input inverters, thus reducing their leakage even further.

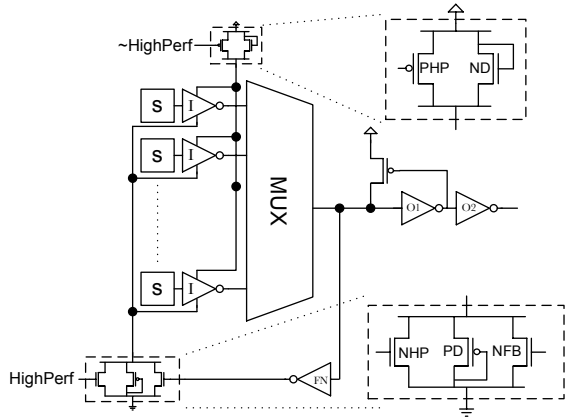For inverters outputting a logic-0, there is a reduced

Fig. 8.   Novel LUT

Fig. 9.   Delay of new LUTs relative to a conventional LUT

TABLE I

INCREASE IN LUT AREA

	% increase
DH	2%
DHNF	10%
ADH	4%
ADHNF	18%

subthreshold leakage through its pull-up transistor, reduced subthreshold leakage from the multiplexor and reduced gate leakage through both of its transistors. For inverters outputting a logic-1, there is a reduced subthreshold leakage through their pull-down transistor and reduced gate leakage through the PMOS.

To increase the performance of the LUT in low-power mode inverter FN is heavily skewed. Only inverter FN's rise time is important to the performance of the LUT, since it turns ON transistor NFB when a logic-0 is being propagated to the multiplexor output, and thus its NMOS is made minimum size, and its PMOS is made larger than normal.

Thus the addition of transistor NFB to the footer and the feedback signal allows a the circuit to enter a low power state when a logic-0 is not needed, but at the same time allowing a strong logic-0 to appear when needed, and avoiding the problems of the standard diode footer.

IV. RESULTS

All simulation results reported in this paper are based on HSPICE, using Berkeley Predictive Technology Models (BPTM) [16] for a $70nm$ technology. The transistor models were extended to include gate tunneling leakage which was modeled using a combination of four voltage-controlled current-sources, as in [17]. All simulations presented were performed at $110^{o}C$.

To study the proposed LUTs, first a traditional 4-LUT was developed, and sized for equal worst-case rise and fall times. Then two additional LUTs were created: (a) a conventional LUT with only a diode header for its input inverters (DH), as in Fig. 6 and, (b) a conventional LUT with a diode header and a new footer for its input inverters (DHNF), as in Fig. 8.

A. Performance

By adding the header and/or the footer to the conventional LUT, the rise and fall times of the new LUTs are no longer equal in high-performance mode, but they were made equal (thus minimizing the worst-case delay) by skewing inverter O1 and keeping the area of the LUT the same. The delay in high-performance and low-power modes of the new LUTs in comparison to a conventional LUT are shown in Fig. 9. In
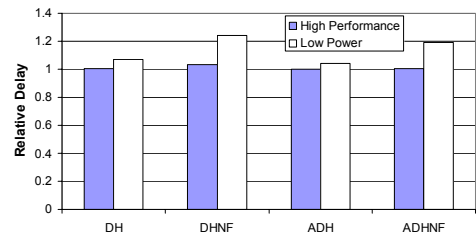
high-performance mode, the designs have a slight increase in the delay; the DH LUT has a 1% increase in delay, while DHNF LUT has a 3% increase in delay. In low-power mode the DH LUT has a 7% increase in delay and the DHNF design has a 24% increase in delay.

To minimize the performance penalty in high-performance mode, we have resized some of the transistors in the input inverters and output buffers in the DH and DHNF LUTs to equalize performance to that of a conventional LUT. The performance of these *alternate designs*, designated as ADH and ADHNF, is also shown in Fig. 9. In high-performance mode there is no performance penalty and in low-power mode there is a 4% and 19% for the ADH and ADHNF designs respectively.

B. Area

The inclusion of the headers, footers and the feedback inverter increase the area of the LUT. Using the alternate designs further increase the area of the LUT. The area overhead of the designs is shown in Table I.

C. Power

To measure the leakage of the different LUT designs, random values were placed in the SRAM bits and the leakage measured. The leakage in the SRAM cells is not included since the SRAM cells are not performance critical and can be constructed to have very low leakage. Fig. 10 shows the average leakage when the multiplexor output is logic-0 and logic-1 in both high-performance and low-power modes for the various LUTs. In high-performance mode there is some leakage reduction for most of the circuits, ranging from 12% to 15%, due to the stacking effect. The alternate DHNF design, however, increases leakage by 5% due to the increase in the transistor sizing.

In low-power mode the DH design reduces leakage by approximately 53% when the multiplexor output is either logic-0 or logic-1. Using the new footer in combination with the diode header, further decreases the leakage. When the multiplexor is outputting a logic-1 there is a 53% reduction in leakage and more importantly when outputting a logic-0 there is a 80% leakage reduction.
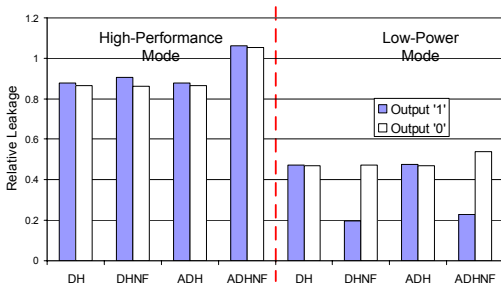
Fig. 10. Leakage of LUTs

Fig. 11. Leakage of Different Sized LUTs

This large difference in the leakage reduction when outputting different logic values can be used to reduce the average leakage of the LUTs. Using the no-cost transformation from [13] we can increase the static probability of signals in the FPGA to be in the logic-1 state to be much higher than 50%. In [13], industry designs could be skewed so that the static probabilities of all signals to be logic-1 was 62% to 84%. With these static probabilities, the new LUT can save on average 70% to 76% of the LUT leakage.

For the alternate DH and DHNF designs, due to the increased area of the designs, the leakage has increased slightly compared to the original designs; the ADH design reduces leakage by 52% regardless of output state and the ADHNF design reduces leakage by 47% and 77% when outputting a logic-0 and logic-1 respectively. Combining the ADHNF design with the technique in [13] reduces total leakage by 65% to 72%.

The use of the new footer incurs, however, a slight increase in the dynamic energy due to the inclusion of the extra inverters and the charging and discharging of the transistor in the footer. The DHNF design incurs a 3% increase the dynamic power consumption while a diode header by itself incurs a 0.5% increase. The alternate DHNF design has a further 7% increase in the dynamic power consumption. Since leakage in FPGAs is now surpassing 50% of total power consumption [9], the leakage reduction outweighs the increase in dynamic power consumption.

D. Different LUT-sizes

Due to the trend in some new commercial FPGAs which use larger LUTs [12] we have also tested the new low-leakage LUT circuits on a 6-LUT. Fig. 11 shows the leakage reduction of the various LUTs, in low power mode. Since the input inverters comprise a larger percentage of the total leakage, there is more room for leakage reduction in the footers/headers. The DH and ADH designs reduce leakage by 62% and 61% respectively; the DHNF reduces leakage by 64% and 94% when outputting a logic-0 and logic-1 signal respectively leading to an total LUT leakage reduction of 83% to 89%. The ADHNF reduces leakage by 52% and 92% when outputting a logic-0 and logic-1 signal respectively leading to an total LUT leakage reduction of 77% to 86%.

V. CONCLUSION

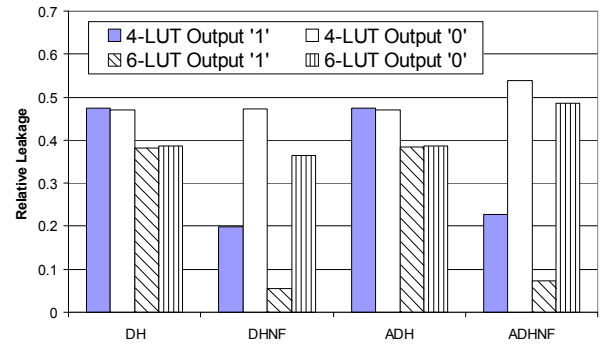As technology continues to scale, reducing leakage in FPGAs is becoming increasingly important. Most recent work on reducing leakage has concentrated on reducing the leakage in the interconnect fabric; given this leakage reduction and the increasing LUT size in commercial FPGAs, reducing the leakage in the LUT will become more important. In this paper we present new designs for low-power LUTs; one LUT, which uses a conventional low-leakage technique, reduces leakage by 53% with an area increase of 2%; a second low-leakage LUT reduces leakage by 72% with an area increase of 18% and no increase in delay in high-performance mode.

REFERENCES

[1] W.K. Henson et al. Analysis of leakage currents and impact on off-state power consumption for CMOS technology in the 100-nm regime. *IEEE Transactions on Electron Devices*, 47(2):440–447, February 2000.
[2] *2002 International Technology Roadmap for Semiconductors*.
[3] V. George and J. Rabaey. *Low Energy FPGAs: Architecture and Design*. Kluwer Academic Publishers, Boston, MA, 2001.
[4] I. Brynjolfson and Z. Zilic. Dynamic clock management for low power applications in FPGAs. *Custom Integrated Circuits Conference*, 2000.
[5] Amit Singh and Malgorzata Merk-Sadowska. Efficient circuit clustering for area nd power reduction in FPGAs. *ACM/SIGDA Symposium on Field-Programmable Gate Arrays*, 2002.
[6] V. George, H. Zhang, and J. Rabaey. The design of a low energy FPGA. *International Symposum on Low Power Electronics and Design*, 1999.
[7] Fei Li, Deming Chen, Lei He, and Jason Cong. Architecture evaluation for power-efficient FPGAs. *ACM/SIGDA Symposium on Field-Programmable Gate Arrays*, 2003.
[8] Jason Anderson and Farid. N. Najm. A novel low-power FPGA routing switch. *IEEE Custom Integrated Circuits Conference*, 2004.
[9] Arifur Rahman and Vijay Polavarapuv. Evaluation of low-leakage design techniques for field programmable gate arrays. *ACM/SIGDA Symposium on Field-Programmable Gate Arrays*, 2004.
[10] Tim Tuan and Bocheng Lai. Leakage power analysis of a 90nm FPGA. *IEEE Custom Integrated Circuits Conference*, 2003.
[11] P.R. van der Meer, A.van Staveren, and A.H.M van Roermund. Ultra-low standby-currents for deep sub-micron VLSI CMOS circuits: Smart series switch. *International Symposium on Circuits and Systems*, 2000.
[12] David Lewis et al. The Stratix II logic and routing architecture. *ACM/SIGDA Symposium on Field-Programmable Gate Arrays*, 2005.
[13] Jason Anderson, Farid. N. Najm, and Tim Tuan. Active leakage power optimization for FPGAs. *ACM/SIGDA Symposium on Field-Programmable Gate Arrays*, 2004.
[14] S. Mutoh, T. Douseki, Y. Matsuya, T. Aoki, S. Shigematsu, and J. Yamada. 1-V power supply high-speed digital circuit technology with multithreshold-voltage CMOS. *IEEE Journal of Solid-State Circuits*, 30(8):847–854, August 1995.
[15] K. Kumagai, H. Iwaki, H. Yoshida, H. Suzuki, T. Yamada, and S. Kurosawa. A novel powering-down scheme for low Vt CMOS circuits. *Symposium on VLSI Circuits*, 1998.
[16] http://www-device.eecs.berkeley.edu/~ptm/.
[17] D. Lee, W. Kwong, D. Blaauw, and D. Sylvester. Simultaneous subthreshold and gate-oxide tunneling leakage current analysis in nanometer CMOS design. *ISQED*, pages 287–292, 2003.