Power and Delay Comparison of Binary and Quaternary Arithmetic Circuits

Cristiano Lazzari
INESC-ID
Lisbon - Portugal
Email: lazzari@inesc-id.pt

Paulo Flores
INESC-ID / IST, TU Lisbon
Lisbon - Portugal
Email: pff@inesc-id.pt

José Monteiro
INESC-ID / IST, TU Lisbon
Lisbon - Portugal
Email: jcm@inesc-id.pt

***Abstract*—Interconnections play a crucial role in deep sub-micron designs since they dominate the delay, power and area. This is especially critical for modern million-gates FPGAs, where as much as 90% of chip area is devoted to interconnections. Multiple-valued logic allows for the reduction of the required number of signals in the circuit, hence can serve as a means to effectively curtail the impact of interconnections. We present in this paper a comparison of binary and quaternary implementations of arithmetic modules based on lookup table structures using voltage-mode circuits. Our assessment demonstrates that significant power reduction is possible through the use of quaternary structures, with very low delay penalties.**

***Index Terms*—Quaternary Logic, Arithmetic Circuits, FPGA Synthesis, Delay and Power Consumption.**

I. Introduction

The large number of components in modern systems on chip (SoCs) presents new challenges to designers. The high integration of different systems increases the number and length of interconnections, hence the overall complexity involving the connections of these systems.

Moreover, interconnections are becoming the dominant aspect of the circuit delay for state-of-the-art circuits due to the advent of deep sub-micron technologies (DSM). This fact is becoming even more significant with each new technology generation [8]. In DSM technologies, the gate speed, density and power scaling follows the Moore's law. On the other hand, the interconnection resistance-capacitance product increases exponentially with the technology node leading to an increase of network delay. Even after modifications in interconnections, from aluminum to copper and low-k inter metal dielectric materials, the problem remains and it is getting more significant [1].

Interconnections play an even more crucial role in Field Programmable Gate Arrays (FPGA), because they not only dominate the delay, but they also severely impact on power consumption [11] and area [15]. This is explained by the fact that in modern million-gates FPGAs, as much as 90% of chip area is devoted to interconnections [4].

Multiple-valued logic (MVL) has received increased attention in the last decades because of the possibility to represent the information with more than two discrete levels [6]. Representing data in a MVL system is more effective than the binary-based representation because the number of interconnections can be significantly reduced, with major impact in all design parameters: less area dedicated to interconnections; shorter interconnections, leading to increased performance; lower interconnect switched capacitance, hence lower power dissipation.

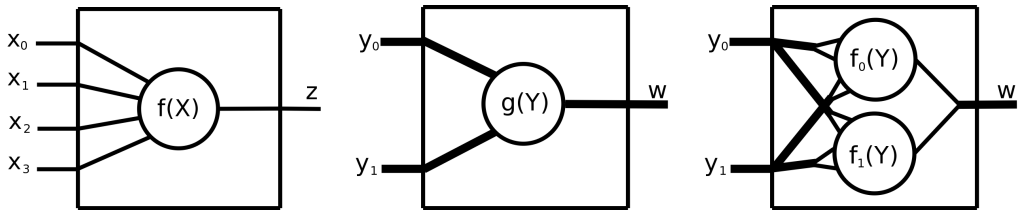
The possibility to represent the information using MVL is not recent. MVL has been successfully accomplished in flash memories [10], for example, where a single memory cell can hold different logic values. Some combinational circuits such as adders [7] and multipliers [9], as well as FPGAs [17], [14] were also proposed.

The main drawback of these systems is that they are based on current-mode devices. They present successful improvements in reducing area, but their excessive power consumption and implementation complexities has prevented these systems from being until now a viable alternative to standard CMOS designs.

A voltage-mode MVL technique is presented in [5] in order to deal with the static power dissipation problem using a standard CMOS process, and still maintain the logic compaction allowed by MVL. This methodology intends to reduce the number of interconnections present in current binary-based systems without incurring on power consumption penalties. The benefits of this MVL methodology can be directly applied to the FPGA domain. A new FPGA structure was proposed in [4] where information is represented by quaternary values. A new quaternary logic cell was presented and results demonstrate interesting area and power reductions in comparison to equivalent binary structures.

In this paper, we present a comparative study of delay and power consumption of binary and quaternary circuit implementations. Circuit implementations are based on lookup table (LUTs) as proposed by [4]. As a case study, we perform an analysis of ripple carry adders. Our experiments were developed using a 45nm process technology [16]. We also developed some experiments with reduced V_{DD} in order to present a more adequate comparison of binary and quaternary methodologies.

This paper is organized as follows. Section II introduces essential background about binary and quaternary lookup tables. Differences in threshold voltages among transistors in the binary and quaternary lookup tables are discussed in Section III. Section IV discusses the behavior of binary and quaternary lookup tables. Experimental results with ripple



(a) 4-input BLUT. (b) 2-input QLUT. (c) Equivalent QLUT using binary functions.

Fig. 1. Binary and quaternary functions.

carry adders are presented in Section V. Finally, Section VI concludes the paper and outlines future work.

II. Binary and Quaternary Lookup Tables

General lookup tables are basically memories, which implement a given logic function. Function results are initially stored in the lookup table structure, and once inputs are applied, the logic value in the addressed position is assigned to the output. The capacity of a lookup table (LUT) is given by

$$|C| = n \times b^k \tag{1}$$

where n is the number of outputs, k is the number of inputs and b is the number of logic values. For example, a 4-input binary lookup table is able to store $1 \times 2^4 = 16$ Boolean values.

For the purpose of this work, only 1-output LUTs ($n = 1$) are discussed in this paper. More details about the differences between binary and quaternary lookup tables are given in the following section.

A. Preliminaries

A binary function implemented by a lookup table (BLUT) is defined as f: $\mathbf{B}^k \rightarrow \mathbf{B}$, over a set of variables $X = (x_0, \cdots, x_i, \cdots, x_{k-1})$, where each variable x_i represents Boolean values. The total number of different functions $|F|$ that can be implemented in a BLUT with k input variables is given by

$$|F| = b^{|C|} \tag{2}$$

where $b = |B|$ (i.e. $b = 2$ in the binary case) and $|C|$ is given by (1). Figure 1(a) illustrates a binary function where $k = 4$. Thus, a lookup table with 4 inputs can implement one of $|F| = 65,536$ different functions.

Quaternary functions are basically generalizations of binary functions. A quaternary function implemented by a quaternary lookup table (QLUT) is defined as g: $\mathbf{Q}^k \rightarrow \mathbf{Q}$, over a set of quaternary variables $Y = (y_0, \cdots, y_i, \cdots, y_{k-1})$, where the values of a variable y_i, as the values of the function $g(Y)$, are in $\mathbf{Q}$= $\{0, 1, 2, 3\}$. As in the binary case, the number of possible function in QLUTs is given by (2), where $b = 4$. In this case, the number of functions that can be represented is around 4×10^9 for a QLUT with only two inputs, which is much larger than the BLUT. Figure 1(b) illustrates a 2-input quaternary function implemented in a QLUT.

Note that the function $g(Y)$ performs exactly the same function as two binary BLUTs, $f_0(Y)$ and $f_1(Y)$, as depicted in Figure 1(c), where f_0 represents the least significant Boolean values and f_1 represents the most significant ones.

Since a quaternary variable y is capable of representing twice as much information as a binary variable x, we consider the cardinality of $|Q| = 2 \times |B|$ in our experiments. In other words, we assume that two binary variables can be grouped in order to represent a quaternary variable. Such characteristic aims at reducing the total number of connections and the number of gates as well.

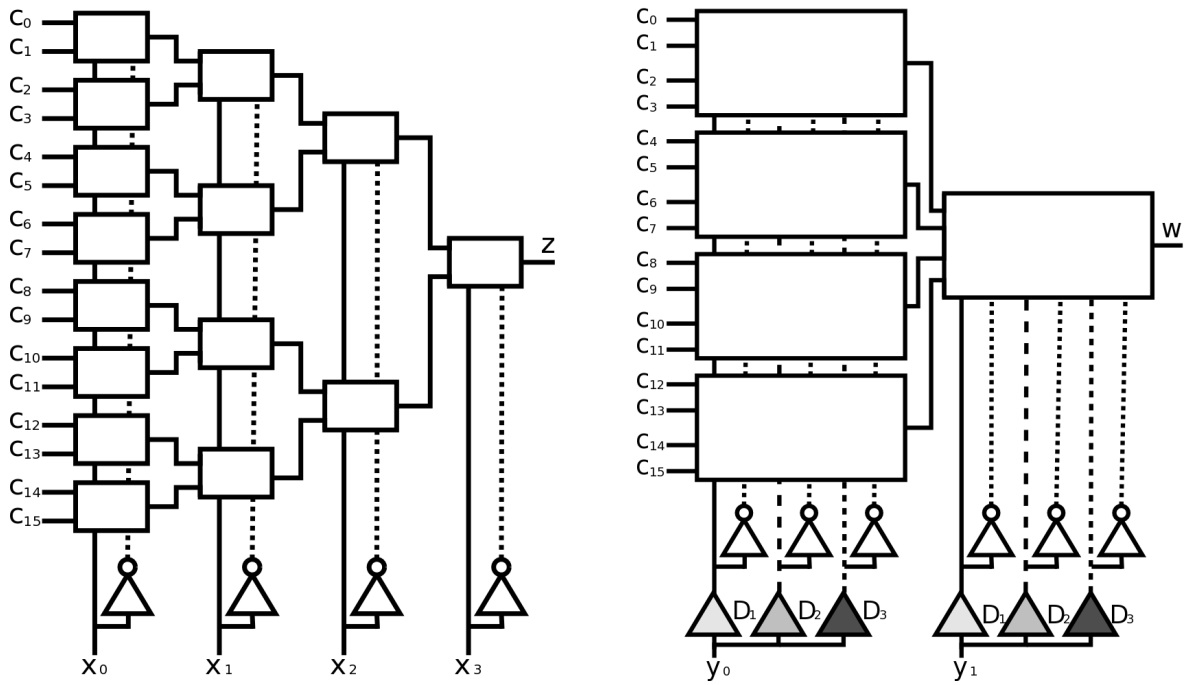
B. Lookup Tables Implementation

In our experiments, the implementation of binary and quaternary lookup tables were implemented by a set of multiplexers as presented in [4] and illustrated in Figure 2.

Figure 2(a) shows a binary 4-BLUT implementation ($b = 2, |X| = k = 4, |C| = 16$) where $x_i \in X$ are the inputs, $c_i \in C$ form the lookup table configuration and z is the output. The BLUT is composed of four stages as a consequence of the number of inputs. Multiplexers are responsible for propagating configuration values to the BLUT output. The multiplexers are composed of pass gates, which receive selection signals from the four BLUT inputs and associated inverters.

A quaternary lookup table (QLUT) follows the same structure as the BLUTs. However, Down Literal Circuits (DLCs) [5] structures are inserted to determine which configuration value must be propagated to the output. Figure 2(b) illustrates the implementation of a 2-input QLUT ($b = 4, |Y| = k = 2, |C| = 16$). As in the binary case, $c_i \in C$ are the lookup table configuration, $y_i \in Y$ are the inputs and w is the output. Due to the quaternary representation, each multiplexer has four configuration inputs, therefore only two multiplexer stages are required.

The Down literal circuits (DLCs) (D_1, D_2 and D_3 in Figure 2(b)) have structures similar to inverters (with 1 PMOS and 1 NMOS transistor). Transistors in each DLC circuit have modified V_{th} values in order to allow the switching at different input voltages. This way, the 3 DLCs circuits work as a thermometer system. Output values are only '0' (GND) or '3' (V_{DD}), according to the logic value applied to their inputs. Table I shows the DLC output logic values as function of the inputs.



(a) 4-input BLUT. (b) 2-input QLUT.

Fig. 2. Binary and quaternary lookup tables implementation.

Although transistors used in the implementation of the DLCs have different threshold voltages (V_{th}), to allow the desired behavior, it is important to highlight that standard CMOS technology is used in the whole QLUT. Only the DLC structures are composed of 6 transistors with different threshold voltages (3 PMOS and 3 NMOS). The quaternary multiplexers are composed of transistors with the same V_{th} as the ones used in the binary multiplexers.

More details about the threshold voltages are discussed in the following Section.

TABLE I
DOWN LITERAL CIRCUITS (DLCS) BEHAVIOR ACCORDING TO THE LOGIC VALUE AT THE INPUT.

Input	D1	D2	D3
0	3	3	3
1	0	3	3
2	0	0	3
3	0	0	0

III. V_{th} ADJUSTMENT FOR QUATERNARY REPRESENTATION

For our experiments, a 45nm process technology was used [16]. The V_{DD} voltage of this technology is 1V. In general, binary logic values are represented by 0V, meaning the logic value '0', and 1V, meaning the logic value '1'.

In a quaternary representation, four discrete voltages are necessary in order to describe the four logic levels. As proposed by Cunha *et al* [4] for a 0.18μm technology ($V_{DD} = 1.8V$), we decide to maintain the V_{DD} voltage as the same voltage used in the binary representation. By maintaining the V_{DD} voltage we avoid the power increase due to the V_{DD}^2 effect.

Thus, the quaternary logic levels '0', '1', '2' and '3' are represented by the discrete voltages 0V, 0.33V, 0.66V and 1V, respectively. Figure 3 shows the behavior of the 3 pairs of transistors used in the DLC structure regarding its drain to source (I_{DS}) current as function of the voltage applied to the gate (V_G). Different from the binary BLUT, where only two different transistors are required, 6 different transistors with particular threshold voltages are required for the QLUT.

With respect to the DLC structures, transistors PMOS1 and NMOS1 are used in $D1$. When $V_G = 0V$ ('0'), the current cross PMOS1 (NMOS1 is open) and 1V ('3') is applied to the output. On the other hand, if $V_G = 1V$ ('3'), PMOS1 is open and NMOS1 is discharging the output value in the GND.

$D2$ is composed of transistors PMOS2 and NMOS2. These transistors are basically the same as the binary transistors. PMOS1 is conducting when $V_G = 0V$ ('0') and $V_G = 0.33V$ ('1'). Differently, NMOS1 is closed and PMOS1 is open if $V_G = 0.66V$ ('2') and $V_G = 1.0V$ ('3'). PMOS3 and NMOS3 are the transistors in the $D3$. PMOS3 conducts when $V_G < 1.0V$ ('0','1','2') and NMOS3 is conducting, otherwise.

Table II summarizes in column 2 the threshold voltages for each transistor used in the DLCs. The V_{th} adjustment was performed such that we obtain the thermometer behavior and forcing the current through a transistor that should be in cut-off region to be zero or negligible.

It is clear that a comparison between binary and quaternary

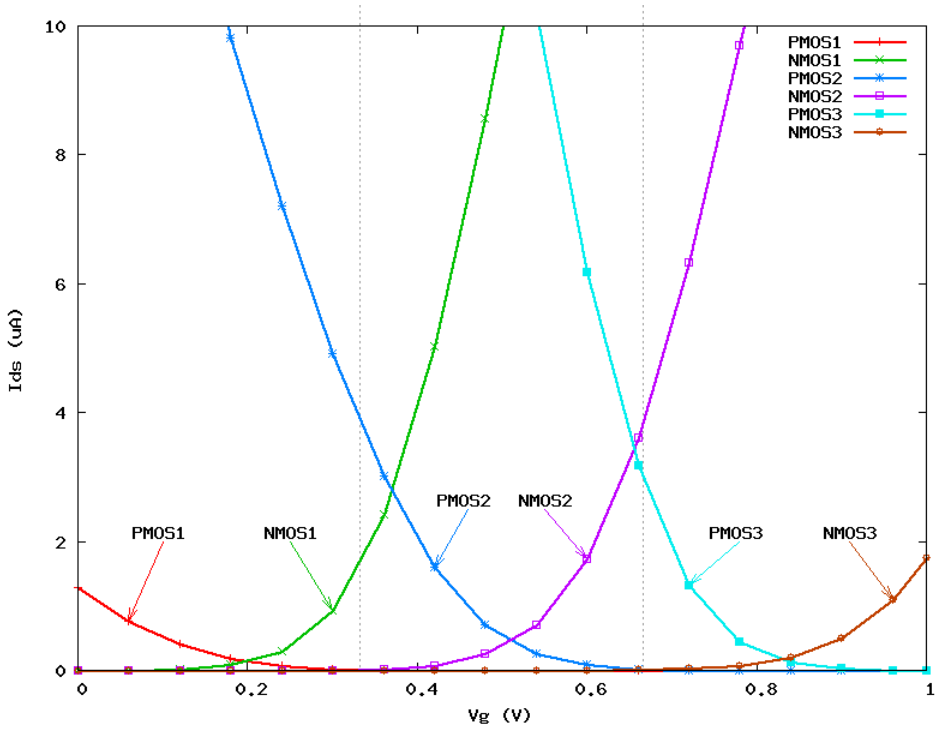


Fig. 3. I_{DS} versus V_G for the transistors used in the quaternay lookup table (QLUT).

TABLE II
THRESHOLD VOLTAGES V_{th} USED IN OUR EXPERIMENTS.

Transistor	Quaternary	Binary	
		$V_{DD} = 1.0V$	$V_{DD} = 0.33V$
PMOS 1	-0.829	-0.418	-0.244
NMOS 1	0.396	0.469	0.236
PMOS 2	-0.556		
NMOS 2	0.642		
PMOS 3	-0.370		
NMOS 3	0.945		

structures is not so fair when different noise margins are considered (i.e. similar V_{DD} voltages for both cases). Figure 4(a) shows I_{DS} as function of V_G when noise margins are larger than in the quaternary representation. Curves show the current behavior according to the gate voltage V_G, as required to the good functioning with $V_{DD} = 1V$.

For this reason, we add in our experiments a new set of transistors for the binary BLUT, where noise margins are reduced to the same level as the quaternary. Figure 4(b) shows the drain-to-source current. These new transistors have different V_{th}s, for which the BLUT functioning is guaranteed with a $V_{DD} = 0.33V$. The voltage thresholds for each transistor used with different $V_{DD}s$ are presented in Table II (last two columns).

IV. POWER CONSUMPTION AND DELAY BEHAVIOR

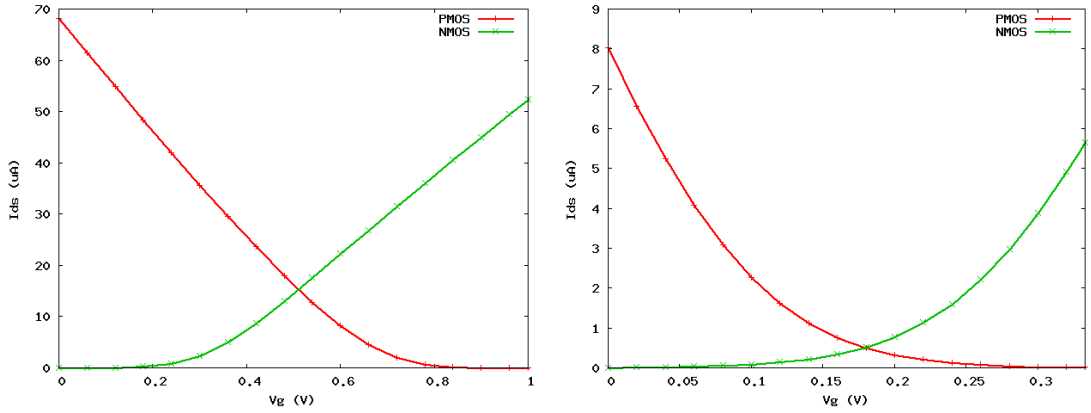
After the adjustment of V_{th}, a fine tuning was performed. Binary and Quaternary lookup tables were properly sized to obtain the best possible delay and power consumption trade-off. As a starting point, the relation between PMOS and NMOS transistors width, in the inverters and DLCs, were tuned in order to obtain an adequate voltage transfer characteristic. Specifically, we define the W_n/W_p relation for the inverters and DLCs circuits, such that their output switches from $0V$ to V_{DD} at a given input voltage V_M

$$V_M = \frac{V_{DD} - |V_{tp}| + V_{tn}\sqrt{\frac{\mu_n W_n}{\mu_p W_p}}}{1 + \sqrt{\frac{\mu_n W_n}{\mu_p W_p}}}, \quad (3)$$

where V_{tp} is the threshold voltage of the PMOS transistor, V_{tn} is the threshold voltage of the NMOS transistor, μ_n is NMOS channel mobility and μ_p is PMOS channel mobility [13]. It is important to highlight that channel length is kept the minimum to avoid a performance reduction and, for this reason, it is not present in (3).

With respect to the gate dimensions, different versions were implemented. For the binary BLUT with $V_{DD} = 1V$, we implemented versions with different transistors widths and we choose the one with the best delay-power trade-off. Transistors for quaternary and binary circuits with $V_{DD} = 0.33V$ were sized in order to present similar timing characteristics to the binary version with $V_{DD} = 1V$.

Figure 5 shows the delay versus power consumption behavior as a function of the load capacitances ($0.1fF$, $1fF$, $10fF$ and $100fF$) for the binary and quaternary lookup tables used in our experiments. These results were obtained using Cadence Spectre simulations [3] and using transistors models created in the Predictive Technology Model (PTM) website [16],



(a) VDD=1V (b) VDD=0.33V

Fig. 4. I_{DS} versus V_G for the transistors used in the binary lookup table (BLUT) with VDD=1V and VDD=0.33V.

according to the specifications presented in Section III.

For load capacitances from $0.1fF$ to $100fF$, binary BLUTs present small variation concerning power consumption, but the delay varies more than 10x. The quaternary structure is more sensitive than the binary BLUTs in respect to power consumption, which increases 2x. On the other hand, the delay increases around 4x when the load capacitance goes from $0.1fF$ to $100fF$.

Based on the development of these structures and according to the obtained simulation results, the following considerations can be reported.

1) The power consumption can be considerably reduced by decreasing the V_{DD} voltage as expected, but with a delay penalty. The binary structure with reduced V_{DD} never reaches the same speed as the traditional one, even with sizing and modified V_{th}.
2) Quaternary structures with similar delay characteristics to binary ones always present better power consumption. Besides, QLUTs are less sensitive to load capacitance variations concerning timing.

V. Experimental Results

In order to determine the timing and power consumption of the binary and quaternary LUTs in circuits, we implemented ripple carry adders using these structures. Binary adders were synthesized with Berkeley ABC [12]. BLUT mapping were also performed with the ABC tool to 4-input structures.

To the best of our knowledge, there is no automated way to map quaternary circuits into QLUTs. Some tools such as the Berkeley MVSIS [2] are able to perform logic synthesis of multiple-valued circuits, but the technology mapping generates only binary circuits. For this reason, we manually generated the whole architecture of the quaternary circuits mapped to 2-input QLUTS.

Quaternary circuits were generated taking the binary descriptions as basis. Binary nets were paired and the BLUT structures are grouped aiming at generating circuits with the smallest number of QLUTs as possible.

Table III shows some experimental results, obtained from the comparison of binary and quaternary ripple carry adders. These results were obtained with the fast spice simulator Cadence Ultrasim [3]. The delay of these circuits was obtained by applying input vectors in which carry signals are propagated through the whole circuit. Power consumption was obtained by the simulation of 1024 random transitions for each circuit.

The binary circuits with $V_{DD} = 1V$ present better performance than the quaternary ones. The main reason is that the DLC circuits are slower than the binary inverter. The DLC $D1$ is composed by a very slow PMOS transistor (PMOS1 In Figure 3) as a consequence of its V_{th}. Similarly, the transistor NMOS3 in $D3$ is very slow to propagate a '0'.

On the other hand, significant power reduction is obtained with quaternary circuits. The results show that quaternary circuits dissipate in average 65% of the power consumption of the binary circuits. Binary circuits with $V_{DD} = 0.33V$ present very low power consumption but timing limitations make them prohibitive for many applications.

VI. Conclusion

This work presents for the first time a comparison between binary and quaternary implementations of ripple carry adders circuits in lookup table structures. Results show reduced power consumption of the quaternary circuits with low impact on the timing.

We are working on methods for automating the technology mapping in order to make possible the synthesis of any type of logic circuits into QLUTs. Significant improvement concerning timing and power consumption is expected in random logic due to the reduction of number of connections and gates.

References

[1] K. Banerjee, S.J. Souri, P. Kapur, and K.C. Saraswat. 3-D ICs: a novel chip design for improving deep-submicrometer interconnect performance and systems-on-chip integration. *Proceedings of the IEEE*, 89(5):602–633, May 2001.

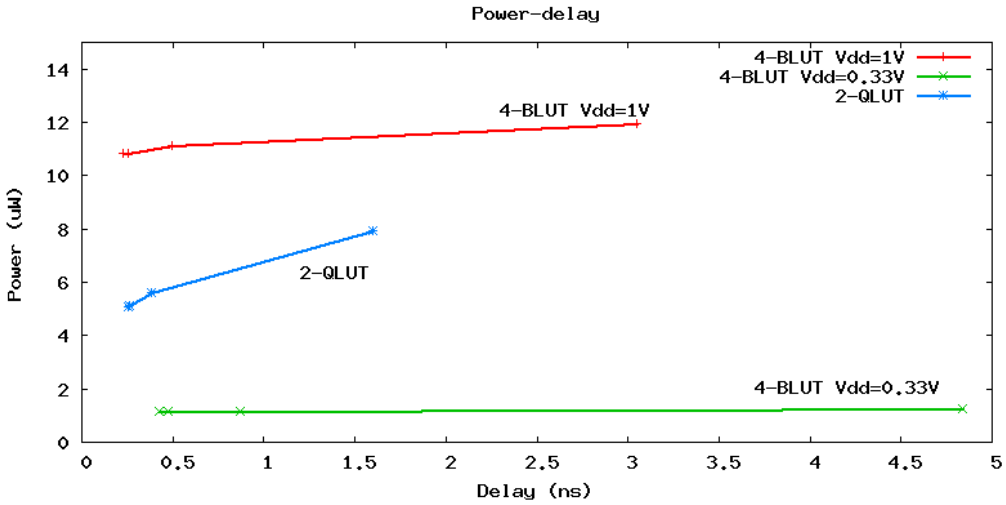


Fig. 5. Delay versus power consumption for binary and quaternary structures with load capacitances from $0.1fF$ to $100fF$.

TABLE III
DELAY AND POWER CONSUMPTION OF BINARY AND QUATERNARY CIRCUITS MAPPED TO LOOKUP TABLES (UNITS ARE ns FOR DELAY AND $fJ/operation$ FOR POWER CONSUMPTION).

Circuit	**Binary**					**Quaternary**		
	#	$V_{DD} = 1V$		$V_{DD} = 0.33V$				
	BLUTs	**Delay**	**Energy**	**Delay**	**Energy**	**#QLUTs**	**Delay**	**Energy**
FA 16	32	16.8	367	25.4	91	16	17.9	232
FA 32	64	33.1	755	49.7	176	32	35.3	553
FA 64	128	66.7	1159	101.8	258	64	68.0	712
Average Gain							-4.9%	65.8%

[2] Robert K. Brayton and Sunil P. Khatri. Multi-valued logic synthesis. In *VLSID '99: Proceedings of the 12th International Conference on VLSI Design - 'VLSI for the Information Appliance'*, page 196, Washington, DC, USA, 1999. IEEE Computer Society.

[3] Cadence. Available at http://www.cadence.com/. Visited on May, 2009, 2009.

[4] R. Cunha, H. Boudinov, and L. Carro. Quaternary look-up tables using voltage-mode CMOS logic design. *ISMVL 2007. 37th International Symposium on Multiple-Valued Logic*, pages 56–56, May 2007.

[5] R.C.G. da Silva, H. Boudinov, and L. Carro. A novel voltage-mode CMOS quaternary logic design. *IEEE Transactions on Electron Devices*, 53(6):1480–1483, June 2006.

[6] E. Dubrova. Multiple-valued logic in VLSI: challenges and opportunities. In *Proceedings of NORCHIP'99*, pages 340–350, 1999.

[7] A.F. Gonzalez and P. Mazumder. Multiple-valued signed digit adder using negative differential resistance devices. *IEEE Transactions on Computers*, 47(9):947–959, Sep 1998.

[8] Amit K. Gupta and William J. Dally. Topology optimization of interconnection networks. *IEEE Comput. Archit. Lett.*, 5(1):3, 2006.

[9] T. Hanyu and M. Kameyama. A 200 MHz pipelined multiplier using 1.5 v-supply multiple-valued mos current-mode circuits with dual-rail source-coupled logic. *IEEE Journal of Solid-State Circuits*, 30(11):1239–1245, Nov 1995.

[10] Tae-Sung Jung, Young-Joon Choi, Kang-Deog Suh, Byung-Hoon Suh, Jin-Ki Kim, Young-Ho Lim, Yong-Nam Koh, Jong-Wook Park, Ki-Jong Lee, Jung-Hoon Park, Kee-Tae Park, Jhang-Rae Kim, Jeong-Hyong Yi, and Hyung-Kyu Lim. A 117-mm2 3.3-v only 128-mb multilevel NAND flash memory for mass storage applications. *IEEE Journal of Solid-State Circuits*, 31(11):1575–1583, Nov 1996.

[11] Fei Li, Y. Lin, Lei He, Deming Chen, and J. Cong. Power modeling and characteristics of field programmable gate arrays. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 24(11):1712–1724, Nov. 2005.

[12] Alan Mishchenko, Satrajit Chatterjee, and Robert Brayton. Dag-aware aig rewriting a fresh look at combinational logic synthesis. In *DAC '06: Proceedings of the 43rd annual conference on Design automation*, pages 532–535, New York, NY, USA, 2006. ACM.

[13] A. S. Sedra and K. C. Smith. *Microelectronic Circuits*. Oxford University Press, 4 edition, 1998.

[14] A. Sheikholeslami, R. Yoshimura, and P. Gulak. Look-up tables (LUTs) for multiple-valued, combinational logic. In *ISMVL '98: Proceedings of the The 28th International Symposium on Multiple-Valued Logic*, page 264, Washington, DC, USA, 1998. IEEE Computer Society.

[15] Amit Singh and Malgorzata Marek-Sadowska. Efficient circuit clustering for area and power reduction in FPGAs. In *FPGA '02: Proceedings of the 2002 ACM/SIGDA tenth international symposium on Field-programmable gate arrays*, pages 59–66, New York, NY, USA, 2002. ACM.

[16] Wei Zhao and Yu Cao. New generation of predictive technology model for sub-45nm design exploration. *International Symposium on Quality Electronic Design*, pages 585–590, 2006.

[17] Z. Zilic and Z.G. Vranesic. Multiple-valued logic in FPGAs. pages 1553–1556 vol.2, Aug 1993.