Indo-Iranian Journal of Scientific Research (IIJSR)
Volume 2, Issue 1, Pages 83-93, January-March 2018

High Speed Performance and Energy Efficient 256-bit CMOS Priority Encoder

S.Gnana murgan[1] and K.Tharani Priya[2]

[1]*Assistant Professor, Department of ECE, Vivekanandha College of Engineering for Women, Tiruchengode, Tamilnadu, India.*
[2]*PG Student, Department of ECE, Vivekanandha College of Engineering for Women, Tiruchengode, Tamilnadu, India.*

Article Received: 28 November 2017 *Article Accepted: 31 January 2018* *Article Published: 06 February 2018*

ABSTRACT

A high speed performance and efficient energy 256-bit CMOS priority encoder and realized on transistor level using 32 nm predictive technology. The new circuit is designed with a full custom approach and incorporates 2 novel logic styles: the Multiple-Output Monotonic CMOS (M2CMOS) and the Dynamic Inversion technique (DI). The achieved performance is in the order of O (log2 (N)), It are response to the input size. A simulation-based comparative analysis concludes that, compared to the conventional design, the proposed circuit achieves up to 57% improvement in delay, 8% improvement in energy consumption and 39% improvement in EDP, while maintaining 20% smaller transistor count.

Keywords: Priority encoder, 256-bit, CMOS, VLSI, High-performance and Energy-efficient.

1. INTRODUCTION

The priority encoder (PE) is a fundamental circuit in digital systems and belongs in the field of parallel-prefix computations [1]. Its common functionality involves the settlement among several modules of different priority. In its general form, it has *N* binary inputs/ outputs of incrementing priority, where only one output can be selected (set to logic 1) at a given time while the rest remain at logic 0. The highest priority input that is 1 selects the corresponding output, while preventing all lower priority ones from being selected. PEs have been used in many applications including both combinational circuits and memories. Recent literature is plentiful in full custom PE implementations of varying input size, including 8-bit [2-4], 32-bit [5-8], 64-bit [9, 10] and 256-bit [11] circuits. Certain applications, such as high-capacity content-addressable memories (CAMs) [12], require efficient implementations of 256-bit (and larger) PEs. However, most of the existing design techniques do not scale well for increasing (>64-bit) input size, or do not address the matter at all, therefore further research on this field can be beneficial. This work focuses on the design of a new 256-bit PE architecture that is suitable for scaling in modern technologies and maintains logarithmic performance in the order of O(log2(*N*)), with respect to the input size *N* .

A important amount of literature refers to the PE as the "prioritizer" circuit, whereas the combination of a *N*-*N* prioritizer followed by a standard 2^N-*N* binary encoder is called a PE. By this alternative definition, a PE has 2^N inputs and *N* outputs and displays the bit position of the highest priority input that is 1. References [2-11], as well as the current work, use the first definition that we presented for the PE and observe circuits of parallel functionality. The respite of this paper is organized as follows. Section II provides an overview of the 256-bit multilevel lookahead/ folding architecture [11], which is the only existing full custom design for a 256-bit PE, and will be used as a benchmark for performance. Section III presents a novel 256-bit PE with a full custom design on transistor level. Sections IV and V provide the simulation setup and equivalent results of our comparative analysis. Lastly, conclusion is given in section VI.

2. MULTILEVEL LOOKAHEAD/FOLDING ARCHITECTURE

Fig. 1 illustrates a transistor-level schematic of the 8 -bit PE macro used in [11], which was recreated according to

the original work with some necessary modifications. It is implemented in standard n-type dynamic logic/domino and contains 2 cascaded 4-bit PE cells along with lookahead gates. Transistors drawn with bold lines comprise keeper circuits and week feedback inverters, which were added by us and are not present in the original work. Keepers are used to staticize dynamic nodes and fight leakage in dynamic/domino gates, and their use has been deemed essential for robust operation in modern nanoscale processes [1]. In our comparative analysis, we use this modified version with keepers, which will be referred to as $[11]^k$ in the rest of the text.

Transistor sizing has been performed by us and is adjusted for a 32 nm technology (the original being in 0.6 ɜm CMOS), since the authors did not provide any exact information on this topic. Transistor widths are displayed in multiples of *u* (*u*=60 nm), which we define as the nMOS width of a unit-sized inverter (*Ln*=*Lp*=30 nm, *Wn*=60 nm, *Wp*=120 nm) in 32 nm technology. A more detailed analysis on our transistor sizing methodology is given in section IV.

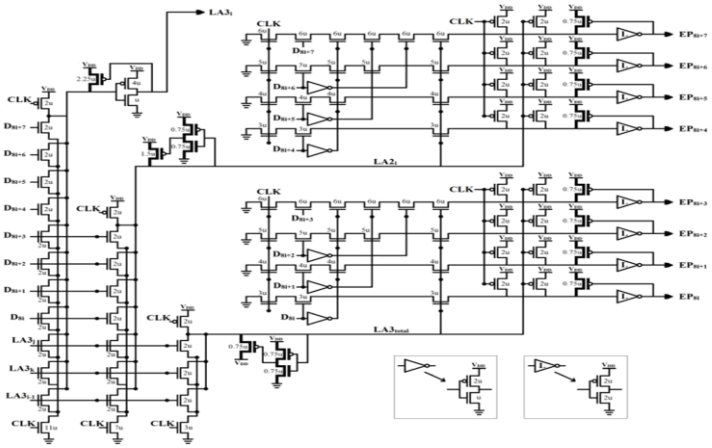

Fig. 1. 8-bit PE macro with keeper circuits $[11]^k$.

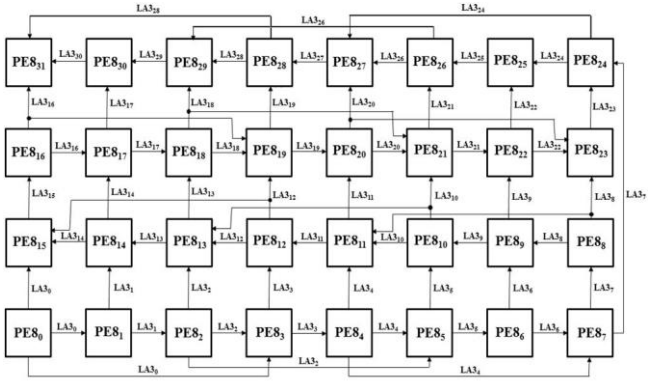

Fig. 2. 256-bit PE with the multilevel lookahead/folding architecture [11]

Fig. 2 illustrates a 256-bit PE with the 3-level lookahead and 3-level folding architecture [11], which comprises 32 8-bit PE macros connected on different levels as shown in the diagram. It has 256 inputs (*Di*) and 256 priority-encoded outputs (*EPi*). Due to its highly symmetric structure, this circuit has several equal critical paths of 8 gate delays (without counting transitional inverters as separate gates), amounting exactly to log2(*N*) performance

(*N*=256, log2(256)=8). The transition we use for calculation of worst-case delay occurs when *D72=D255*=1 and the rest of the inputs are 0. In this case, *EP255* initially rises (0x1) and later resets (1x0) during the evaluation phase.

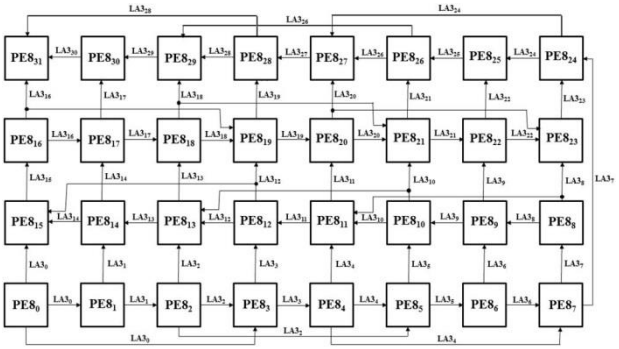

Fig. 3. Lookahead gates used in our design

3. PROPOSED CIRCUIT

This section introduces a recently designed 256-bit PE, as well as two novel CMOS design techniques. The proposed circuit consists of 3 different types of logic modules: lookahead gates, decision blocks and priority encoding cells. They are described in ensuing following subsections A, B and C. Finally, subsection D describes the total 256-bit architecture.

A. Lookahead Gates

The first three stages of the proposed 256-bit PE consist of NOR/NAND lookahead gates implemented in monotonic static CMOS logic (MS-CMOS) [13, 14]. MS-CMOS is a static/dynamic hybrid logic family which seeks to combine the robustness of static CMOS with the high performance of dynamic circuit techniques. Its logic operation is clocked, monotonic and segregated in precharge/evaluation cycles, as in dynamic logic, whereas gates consist of both complementary nMOS and pMOS logic networks.

Static CMOS gates are often sized for equal rise and fall times, whereas MS-CMOS gates are skewed to favor either one of two possible transitions: low-skew (LS) gates precharge high and favor the high-to-low transition whereas high-skew(HS) gates precharge low and favor the low-to-high transition. The skew ratio (SR) is defined as the amount that transistor widths are scaled down in the non-critical switch of the skewed gate [1], in reference to the standard balanced sizing. MS-CMOS gates will always perform their favored transition during evaluation, thus increasing the overall performance. Their weak logic network only serves as a complex keeper to fight leakage and noise during precharge phase.

Fig. 3 illustrates the lookahead gates used in our design: footed 4-bit LS NOR gate (Fig. 3(a)), unfooted 4-bit HS NAND gate (Fig. 3(b)) and unfooted 4-bit LS NOR gate (Fig. 3(c)). All gates are implemented with SR=4. Transistor widths are given in multiples of *u* and different symbols are assigned to each gate (variably colored circles) for distinction.

B. Decision Blocks

Parallel prefix computations, such as priority encoding, involve calculating a set of outputs from a set of inputs, when each output depends on all previous input bits [1]. This often requires computing multiple NAND/NOR functions where one is a sub-function of another. This paper introduces the Multiple Output Monotonic CMOS (M^2CMOS) logic, which is a suitable enhancement to standard MS-CMOS for the aforementioned parallel prefix circuits.

The main idea behind M^2CMOS is to merge gates together by using shared stacked transistors for the non-critical switch networks, whereas the parallel evaluating networks are kept separate. This arrangement yields two significant advantages: firstly, transistor count is reduced, thus saving power and area, and secondly, the input load of gates is reduced, thus improving performance. The three following stages of the proposed 256-bit PE consist of decision blocks implemented in M^2CMOS, which are illustrated in Fig. 4. Again, transistor widths are annotated in multiples of *u* and different blocks are given corresponding symbols (variably colored squares).

The NAND block (Fig. 4(a)) receives 4 inputs *L3<3:0>* and comprises 4 separate gates, which implement the NAND function of inputs *L30*, *L3<1:0>*, *L3<2:0>* and *L3<3:0>* (the first 1-input NAND gate is technically a dynamic inverter). The gates precharge low with nMOS transistors controlled by the clock and calculate with a monotonically rising transition (0X0 or 0X1) executed by pMOS transistors in parallel. The rest of the nMOS transistors serve as a shared keeper network, which branches to all four outputs (hence the "multiple-output" designation) and fights noise and leakage during evaluation.

The NOR block (Fig. 4(b)) operates in a similar but corresponding manner, comprising 4 separate NOR gates that precharge high and evaluate with a monotonically falling transition (1X1 or 1X0). In this case, the precharge/keeper transistors are pMOS whereas the evaluation transistors are nMOS. It can be observed that both fundamental decision blocks are skewed to favor the critical transition, with the non-critical transistors being downsized by a factor of 2 (SR=2) in reference to the corresponding 4-bit gate.

Finally, the AND block (Fig. 4(c)) is merely a NAND block with inverters on the outputs, thus implementing AND logic. The inverters are low-skewed (*Wn* is upsized to 2*u*) to favor the critical transition (see Fig. 1 for the inverter schematic).

C. Priority Encoding Cells

Recent research [4, 10] introduced a novel design technique for the functioning of a very compact and low-power 4-bit PE cell. In this work, an improved 4-bit topology is presented, further reducing transistor count, whereas the aforementioned technique is formally named the Dynamic Inversion (DI). The new 4-bit PE cell is illustrated in Fig. 5. It is implemented in Multiple Output Domino Logic (MODL) enhanced by the DI technique. Compared to the one previously proposed, this cell requires 3 fewer nMOS transistors to implement the priority encoding

function, as a result of merging. The novelty of this cell is displayed by the alternatively placed transistors Z1, Z2 and Z3, whose gates are controlled by dynamic nodes of previous gates instead of input signals or the clock.

It receives 4 data inputs (*Di…Di+3*) and 1 lookahead input *Nj*, and produces 4 encoded outputs (*EPi…EPi+3*), as described by the following boolean functions: signals are derived from corresponding figures 3, 4 and 5. For the sake of clarity, only primary I/O and critical path links are directly displayed, whereas secondary connections between intermediate levels are shown by the arrow flowchart on the left of the tree. The PE tree has a total of 7 computational stages: 3 stages of lookahead logic followed by 3 stages of decision logic and finally the priority encoding stage.

The 1st stage receives primary inputs *D<255:0>* and produces lookahead signals *L<63:0>* which are fed to stages 2 and 6. The 2nd stage receives *L<63:0>* and produces lookahead signals *L2<15:0>* which are fed to stages 3 and 5. The 3rd stage receives *L2<15:0>* and produces lookahead signals *L3<3:0>* which are fed to stage 4. The 4th stage receives L3<3:0> and feeds decision signals *N3<2:0>* to stage 5, which then feeds decision signals *N2<15:0>* to stage 6. Finally, the 7th stage receives decision signals *N<63:0>* produced by stage 6, as well as the primary inputs *D <255:0>*, thus implementing the final priority encoding. The 4th stage produces the *COUT* output, which is the OR function of all inputs *D<255:0>* (if *D<255:0>* = 0…..000 *COUT*=0, else *COUT*=1).

Stages 1-6 operate sequentially and monotonically, whereas stage 7 operates in parallel with the rest. When the evaluation phase begins, the PE cells in stage 7 calculate the priority among their respective inputs while stages 1-6 calculate lookahead and decision logic. The final decision signals *N<63:0>* arrive on the 7th stage later and pull down all outputs except for the one that should remain to logic 1. This arrangement yields an important advantage: stages 1-6 are implemented with high-performance logic (gates that evaluate with parallel transistors), whereas stage 7 is implemented with low-power and compact logic (stacked transistors and branching evaluation paths) without impairing the total performance of the circuit.

The complete 256-bit PE contains a total of 744 logic gates and 3182 transistors. Its worst case delay occurs when *D0=D255*=1 and the rest of the inputs are 0. In this scenario, *EP255* is erroneously set to 1 and has to be pulled down by a lookahead/decision signal that propagates through the critical path of the tree. Fig. 7 illustrates (a) the number of computational stages (without inverters) and (b) total gate count (including inverters) of the proposed PE architecture for various operand lengths between 16 and 4096 bits.

4. SIMULATION SETUP

This section describes the simulation environment and setup used to compare the newly designed 256-bit PE and the multilevel lookahead/folding architecture. Specifically, we use the version with keeper circuitry [11][k] presented in section II.

A.Transistor-level Implementation and Sizing

Both circuits were implemented on transistor level, using a 32 nm predictive technology model for low-power applications (PTM LP), incorporating high-k/metal gate and stress effect [15]. All transistors have the same (minimum) channel length (*Ln*, *Lp*) of 30 nm, whereas transistor widths are sized in multiples of *u* as shown in corresponding Figures 1, 3, 4 and 5.

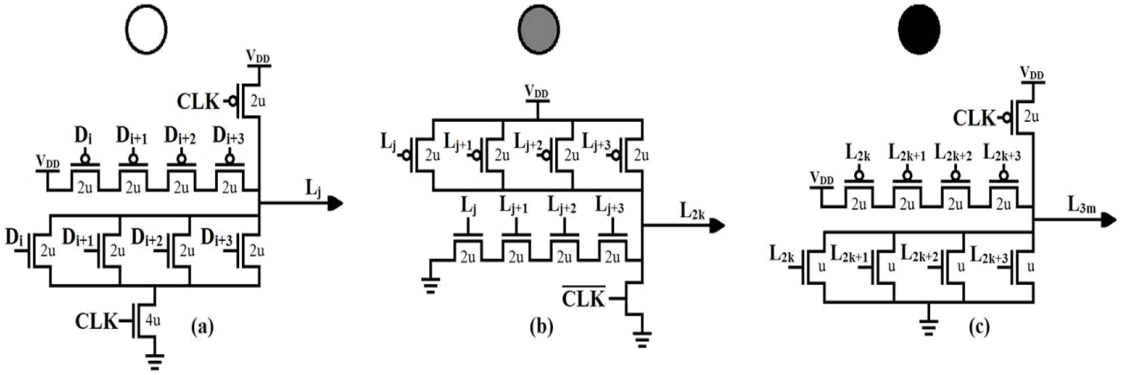

Fig. 4. Lookahead gates implemented in CMOS: (a) footed 4-bit LS NOR, (b) unfooted 4-bit HS NAND and (c) unfooted 4-bit LS NOR

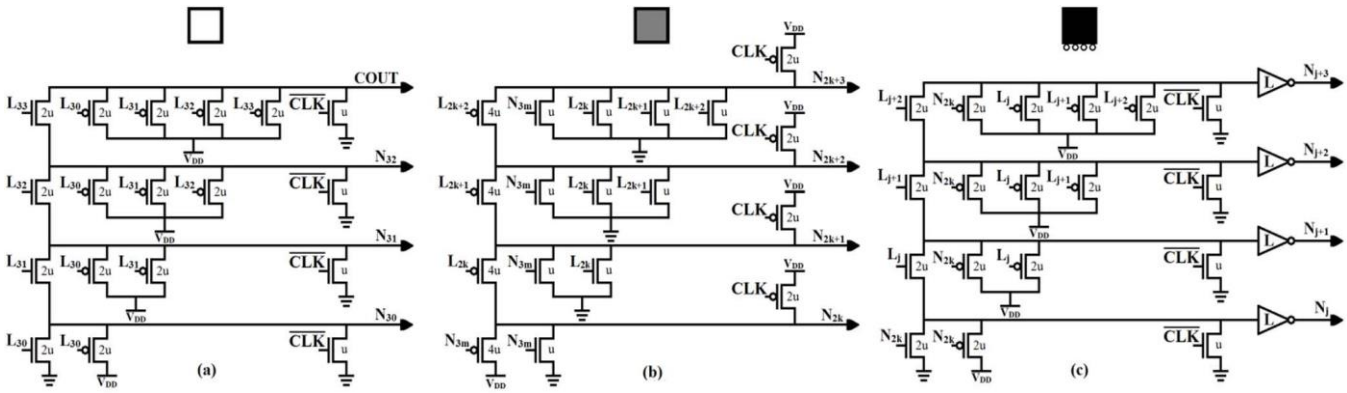

Fig. 5. Decision blocks implemented in M^{2}CMOS: (a) 4-bit NAND block, (b) 4-bit NOR block and (c) 4-bit AND block.

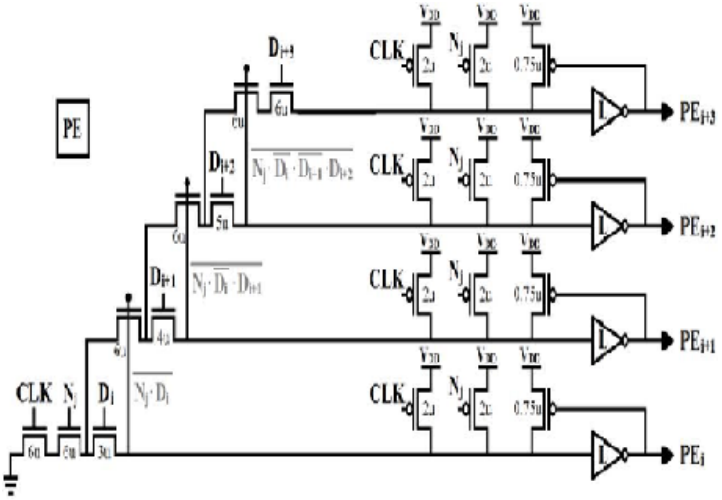

Fig. 6. 4-bit PE cell implemented in MODL enhanced by the DI technique

We used a similar methodology to size both circuits, opting for fair and unbiased comparison. In general, gates are sized to deliver the same current with a unit-sized inverter, however some cases require special reference: a) footer transistors in parallel-evaluating gates are upsized to have at least half the total width of the transistors connected.

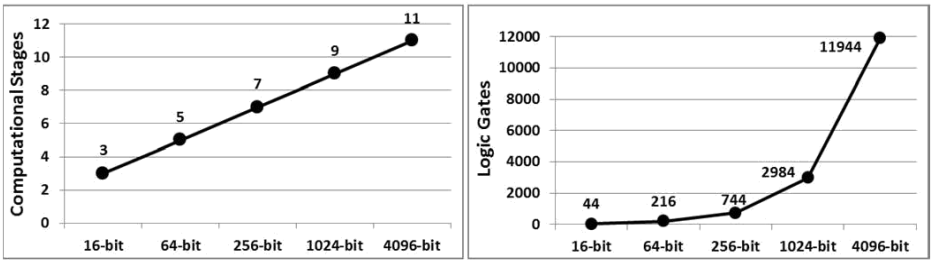

Fig. 7. Scaling of proposed PE tree for various operand lengths, in terms of (a) computational stages and (b) logic gate count

series transistors in monotonic-static gates are downsized (SR=2 or 4, as previously mentioned), since they don't perform critical transitions, c) static inverters following monotonic /dynamic gates are high-skewed (*Wn=60* nm, *Wp*=240 nm) or low-skewed (*Wn*=*Wp*=120 nm) to favor equivalent critical transitions, d) keeper transistors are sized at 10% the width of corresponding pull-down networks with a minimum value of 45 nm (0.75*u*) and e) feedback inverters (added for keeper circuitry in [11][k]) are minimally sized (*Wn*=*Wp*=45 nm).

B. Evaluation and Comparison Methodology

The evaluation procedure was realized with BSIM4-level SPICE simulations. Specifically, transient powerup analysis was used to calculate the delay and power dissipation of the examined circuits under various conditions. Two sets of simulations were performed, one focusing on delay and one focusing on energy and energy-delay-product (EDP).

The first simulation set is used to calculate the critical delay of examined circuits under varied operating temperature (*T*) and supply voltage (*V DD*). For the temperature sweep, *VDD* is set to the nominal value of 1 V and *T* is swept across 6 different values: 0, 25, 50, 75, 100 and 125 ^{o}C. For the voltage sweep, *T* is set to the typical value of 70 ^{o}C and *VDD* is swept across 5 different values: 0.8, 0.9, 1.0, 1.1 and 1.2 V. In each case, circuits are set to perform their slowest possible transition during evaluation phase. Delay is calculated between the insertion of the (buffered) clock and the switching of output node *EP255* to the correct value.

The second simulation set is used to calculate the energy consumption and EDP of examined circuits under typical conditions (*V DD* = 1 V, *T* = 70 ^{o}C) with varied input switching probability *a* = 0.125, 0.25 and 0.5. The circuits are set to operate at the maximum frequency (*F*) attainable under such conditions, which is calculated as *F*=1/(2·*max_delay*), where *max_delay* is obtained from the first simulation set. We generated 3 different sets (with varied *Į*) of 256 random 50 -bit sequences, which are fed to the PE for a time period of 50 clock cycles. Inputs *Di*

toggle at the beginning of each cycle, with a probability denoted by *I*. When *a* = 0.125, 0.75 and 0.5, each input has 12.5%, 25% and 50% chance to toggle, respectively. The average power dissipation, including all voltage sources (inputs/clock/supply), is extracted from SPICE and divided by *F* to provide the energy per operation/clock cycle.

C. Simulation Test Bench

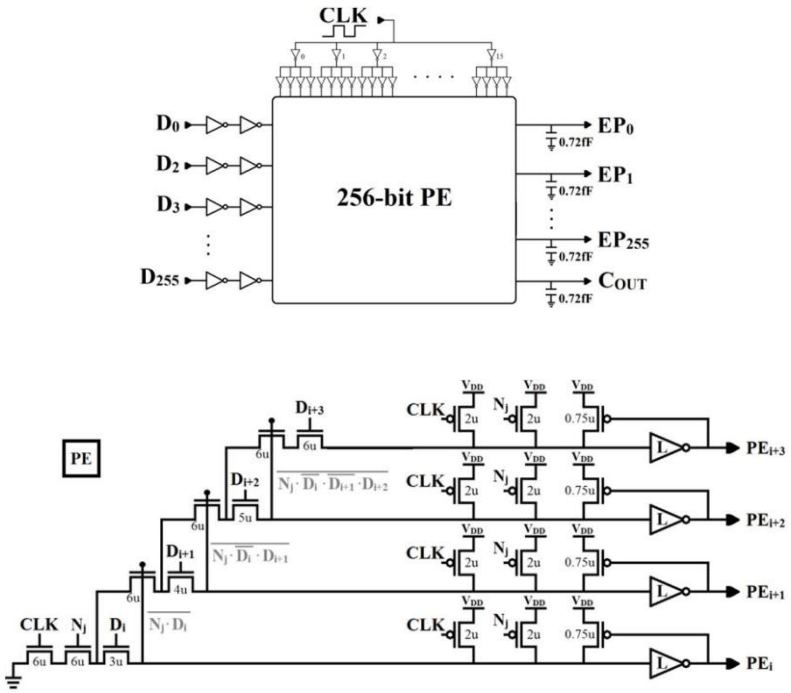

Fig. 8. Simulation test bench

Fig. 8 illustrates the test bench used in all simulations. Inputs *Di* are fed to the PE in the form of voltage sources buffered by a pair of inverters. The inverters are unit-sized for roughly equal rise/fall times and used to provide a sensible input slope. The clock signal (*CLK*) is a single voltage source buffered by a tree of 80 inverters for even distribution to various parts of the circuit. The clock inverters are four times unit-size (*Wn*=240 nm, *Wp*=480 nm) since *CLK* has to drive a large load. Finally, each output (*EPi* and *COUT*) is connected to a capacitor of 0.72fF, which is the FO4 capacitance we measured for the 32 nm model in typical conditions (T = 70^{o} C, *VDD* = 1V).

5. COMPARATIVE RESULTS

This section sums up the results obtained from our comparative analysis, including gate/transistor count, delay, energy and EDP. Simulation data are illustrated in Tables I-IV.

A. Gate and Transistor Count

Table I displays the results regarding gate and transistor count (input/clock buffers not included). The proposed 256-bit PE contains a total of 744 logic gates and 3182 transistors. In terms of gate count, it achieves 10.79% and 16.96% improvement in comparison with [11] and [11]^{k}, respectively, whereas in terms of transistor count, the corresponding improvement is 8.77% and 19.80%. Even without keepers, the original circuit [11] still has 306

more transistors than the proposed one. The main reason is the new compact 4-bit PE cell, which allows for the elimination of inverters on the primary inputs, effectively saving 192 inverter gates and 384 transistors compared to [11]. Additionally, M^2CMOS saves 6 transistors on each decision module, compared to standard MS-CMOS implementation. These results are an indication of the area efficiency achieved by the two novel design styles.

B. Delay

Tables II and III display the simulation results concerning delay under varied temperature and voltage, respectively. Considering all 11 simulation cases, the proposed circuit achieves 22-57% lower delay than $[11]^k$. This significant improvement occurs largely due to the fact that the 256-bit multilevel lookahead/folding architecture uses wide 11-input domino OR gates which require large keepers to match 10% of the pulldown network width.

In the worst case scenario, a single nMOS with *Wn*=120 nm has to overpower a keeper with *Wp*=135 nm (see Fig. 1), resulting in slow transitions especially at low voltage operation. The proposed tree architecture strictly follows a radix-4 structure. in newer technologies. Recent related research at the 22 nm [14] found domino logic w/o keepers to be faster than MS-CMOS and domino with keepers to be of similar performance. We have shown that M^2CMOS can potentially achieve even higher performance and less area than domino with keepers.

C. Energy and EDP

Table IV displays the simulation results concerning energy consumption and EDP with varied *a*. Under typical operating conditions, the proposed circuit achieves 50% higher operating frequency. Regarding energy per operation and for *a*=0.125, 0.25 and 0.5, it achieves 4.65%, 5.56% and 8.23% improvement, respectively.

As for EDP and corresponding values for *a*, it achieves 36.61%, 37.21% and 39.00% improvement, respectively. It is evident that the proposed circuit is increasingly energy-efficient for larger values of *a*, due to the small load on primary inputs *Di*. This is mainly achieved by the new 4-bit PE cell which allows each input signal to drive a single transistor gate and does not require inverters to obtain complementary values. Energy reduction is also achieved (regardless of *a*) due to smaller clock load, since the proposed design uses significantly fewer footer transistors.

6. CONCLUSION

A new 256-bit CMOS PE has been presented, incorporating 2 novel logic styles, namely Multiple-Output Monotonic CMOS (M^2CMOS) and Dynamic Inversion (DI). The new PE architecture offers several advantages independently of technology, including small gate/device count, reduced input/clock load and increasing energy-efficiency in case of high-activity input. In addition, the computational tree follows a fixed radix-4 structure, maintaining performance in the order of O(log2(*N*)) regardless of input size *N*.

A comparative analysis on 32 nm predictive technology concluded that the proposed circuit is faster, more compact and more energy-efficient than the conventional 256-bit design.

256-bit PE	#Gates	%Improv.	#Transistors	%Improv.
Proposed	744	-	3182	-
[11]	832	-10.58	3488	-8.77
[11]*	896	-16.96	3968	-19.81

TABLE I. GATE AND TRANSISTOR COUNT

256-bit PE	Critical Delay (ps)					
	0 °C	25 °C	50 °C	75 °C	100 °C	125 °C
Proposed	265	331	408	493	587	686
[11]*	460	537	631	759	856	982
%Improv.	-42.39	-38.36	-35.34	-35.05	-31.43	-30.14

TABLE II. DELAY VS TEMPERATURE (V_{DD}=1V)

256-bit PE	Critical Delay (ps)				
	0.8 V	0.9 V	1.0 V	1.1 V	1.2 V
Proposed	987	633	476	388	333
[11]*	2304	1107	716	533	428
%Improv.	-57.16	-42.82	-33.52	-27.20	-22.20

TABLE III. DELAY VS VOLTAGE (T=70◦C)

256-bit PE		Critical Delay (ps)	Max Frequency (GHz)	Energy per Operation (pJ)	EDP (ps · pJ)
α=0.72	Proposed	476	1.05	0.82	390.32
	[11]*	716	0.70	0.86	615.76
	%Improv.	-33.52	+50.00	-4.65	-36.61
α=0.74	Proposed	476	1.05	0.85	404.6
	[11]*	716	0.70	0.90	644.4
	%Improv.	-33.52	+50.00	-5.56	-37.21
α=1	Proposed	476	1.05	0.89	423.64
	[11]*	716	0.70	0.97	694.52
	%Improv.	-33.52	+50.00	-8.23	-39.00

TABLE IV. ENERGY AND EDP (T=70◦C,V_{DD}=1V)

REFERENCES

[1] S.W. Huang and Y. J. Chang, "A full parallel priority encoder design used in comparator," *53rd IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*, pp.877-880, Aug. 2010.

[2] J. Mohanraj, P. Balasubramanian and K. Prasad, "Power, delay and area optimized 8-bit CMOS priority encoder for embedded applications," *10th International Conference on Embedded Systems and Applications*,PP111-113, 2012.

[3] J.G. Delgado-Frias and J. Nyathi, "A high-performance encoder with priority lookahead," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol.47, no.9, pp.1390-1393, Sep 2000.

[4] Jinn-Shyan Wang and Chun-Shing Huang, "A high-speed single-phase-clocked CMOS priority encoder," *IEEE International Symposium on Circuits and Systems. Emerging Technologies for the 21st Century*, vol.5, pp. 537-540, May 2000.

[5] J. S.Wang and C. H. Huang, "High-speed and low-power CMOS priority encoders," *IEEE Journal of Solid-State Circuits*, vol.35, no.10, pp.1511-1514, Oct. 2000.

[6] S. K. Maurya and L. T. Clark, "Fast and scalable priority encoding using static CMOS," *IEEE International Symposium on Circuits and Systems (ISCAS)*, pp.433-436, May-June 2010.

[7] D. Balobas and N. Konofaos, "Low-power, high-performance 64-bit CMOS priority encoder using static-dynamic parallel architecture," *5th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, pp. 1-4, May 2016.

[8] C. H. Huang, J. S. Wang and Y. C. Huang, "Design of high-performance CMOS priority encoders and incrementer/decrementers using multilevel lookahead and multilevel folding techniques," *IEEE Journal of Solid-State Circuits*, vol.37, no.1, pp.63-76, Jan 2002.

[9] K. Pagiamtzis and A. Sheikholeslami, "Content-addressable memory (CAM) circuits and architectures: a tutorial and survey," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 3, pp. 712-727, March 2006.