# Optimizing Power and Delays Using Novel Multi-Level Transistor Sizing Approach in CMOS Based Circuits

[1]***M.PREMALATHA***, *Assistant Professor, Dept of ECE, GURU NANAK INSTITUTE OF TECHNOLOGY (A), Hyderabad*

[2]***E .SWARNALATHA***, *Assistant. Professor, Dept of ECE, GURU NANAK INSTITUTE OF TECHNOLOGY (A), Hyderabad*

[3]***NILOFER***, *Assistant Professor, Dept of ECE, Malla Reddy College of Engineering and Technology, Hyderabad*

[4]***SATHPUTE DEEPIKA,*** *Assistant Professor, Dept of ECE, Malla Reddy College of Engineering and Technology, Hyderabad*

**ABSTRACT:** Transistor sizes are determined using the analytical equations defining the behavior of the circuit, and a multi-objective optimization strategy based on genetic algorithms (GA) is provided to fit the circuit performance within required specifications. The optimization algorithm is created in MATLAB, and HSPICE simulations based on 0.18µm CMOS technology parameters are used to validate the designed circuit's performance. Utilizing the suggested design tool, several DIGs were effectively designed and validated. As a result, optimizing for operating variances and processes has become crucial for VLSI design. Furthermore, there are difficult trade-offs in many circuit parameters, necessitating the need for proficient optimization techniques. In light of this, this thesis offers a variety of robust transistor sizing calculations based on optimization algorithms for different nanoscale CMOS circuits in digital form. Meta-heuristics like the thermo-based Simulated Annealing algorithm, the swarm-based Artificial Bee Colony, the Particle Swarm Optimization algorithm, and the evolutionary-based Genetic Algorithm are used to maximize the fitness function in conjunction with the set of constraints. To verify the circuit's yield, which is found to be approximately 98.6%, Monte Carlo simulations using 30,000 random statistical samples were run for each logic cell.

**KEYWORDS:** VLSI, CMOS, Low-Power and Optimization

## I. Introduction

Transistor sizing is well established as an effective way to speed up circuits. Numerous studies have been done in this area [1 to 9]. For static CMOS, the delay of a transition can be modelled as dependent on RC, where R is the effective resistance of the transistors in the pull-up or pull-down circuitry and C is the capacitive load driven by these transistors. R is inversely proportional to the width of the transistors while C is proportional to the size of the transistors in the next stage. Hence by increasing the width of the transistor at the current stage delay can be

reduced. The effects of transistor sizing were studied by custom sizing a four-bit adder [38]. A speedup of 2.3:1 was achieved. The transistors were arranged such that the large p-transistors were above the small n-transistors and vice versa, so that no increase in rectangular area resulted from sizing. In most of the literature on sizing, the length of each transistor is kept at a set value while the width is treated as a continuous variable. Linear Programming methods or other numerical simulation methods are then applied to find the optimal size for each transistor in a circuit [1, 17]. Delay has been shown to be reduced by 60% [2] and in certain cases up to 73% [1], hence proving that transistor sizing is a useful tool in delay optimization.

Recent developments of probabilistic techniques have produced a fast and efficient way of estimating power [13], which is proportional to the average switching probability of a node. The power dissipation of a gate is approximated by the change in energy for charging and discharging the output capacitance of the gate. Since a gate does not necessarily switch at every clock cycle, the frequency of switching is estimated by the clock frequency multiplied by the expected number of switches per cycle.

However, this delay model does not take input transition time into account. To verify this delay model and to investigate the effects of input transition time on output delay, cells from a standard cell library were simulated. Cells were laid out with different output loads to examine the dependence of delay on output load, as well as with different input loads (or the load the fanin node sees), to vary the input transition time

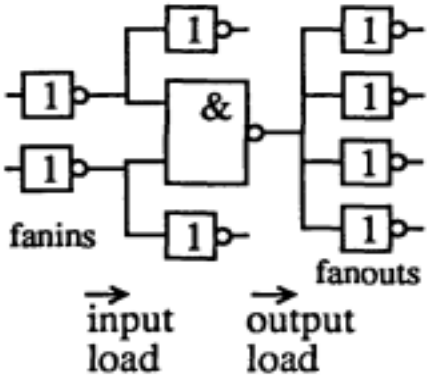



FIG 1: Input and output of different loads

Optimization of VLSI circuits relies heavily on efficient implementation of arithmetic operations considering signal delay, power consumption, and chip area. Such important design considerations and trade-offs lead to a general approach towards transistor sizing that will prove to be extremely useful. In fact, transistor sizing, that is, the operation of enlarging or reducing the channel width of transistors, is a powerful and effective performance optimization tool in the hands of the designer.

Although VLSI circuits can be optimized in a number of ways, such as circuit style selection, structural optimizations, and transistor sizing [1–9], various reasons exist as to why transistor sizing is an important issue. First of all, the power dissipation is a strong function of transistor sizing which affects physical capacitance. Sources of power consumption such as glitches and short-circuit currents can be minimized by careful circuit design and transistor sizing. Second, transistor sizing affects not only the resistance of devices and time constant but also propagation delay of the gate due to the parasitic capacitors. Third, careful transistor sizing is necessary to maintain sufficient noise margins. Even more important is the observation that transistor sizing becomes critical in ensuring proper functionality of a circuit. For all the above reasons, transistor sizing is an essential means of implementing high-performance circuits [10].

Furthermore, with the continued scaling of technology and reduced transistor sizes, the behavior and performance of a circuit could not be investigated without transistor sizing. From simulation results, it has been noticed that a small change in the transistor size for a given technology leads to a remarkable change in the characteristics of a circuit. Therefore, using an appropriate transistor sizing method is necessary for a circuit prior to measuring its parameters.

Since the aim is to have low power dissipation, circuits are first mapped with minimum-sized gates and then changed to larger gates as necessary to satisfy delay constraints. In this way, we start with minimum power, and gates that are not dominant in determining the delay of the circuit remain minimum-sized, thereby ensuring low-power dissipation. Having mapped the circuit, the switching probability of each node can be calculated, and this information, as well as the delay parameters of each gate, can be utilized in optimization routines that select the version that is best suited for each node, such that a given delay constraint is satisfied with minimum power.

## II. Literature Survey

Classical numerical methods, such as the conjugate gradient descent method, have been applied to the transistor-sizing problem: there exist several transistor sizing programs that minimize power consumption while maintaining performance specifications [5, 6, 7]. More recently, several specialized numerical techniques have been proposed [8, 9, 10]. On the analytical side, Cong and Koh have studied the related problem of simultaneous gate and wire optimization for optimal delay and power [13]. Cong and Koh's solution space and optimization metric are different from what we shall see in the present paper. A different analytic approach to the transistor sizing problem, for the performance metric, is given by Hu [11] and another by Horowitz, Indermaur, and Gonzalez [12]. Both Hu and Horowitz et al. present qualitative results; they only analyze basic inverter gates. To the best of the authors' knowledge, the present paper is the first one that goes beyond such a qualitative approach, both in terms of the generality of the optimization metric and in terms of the generality of the considered circuits

Transistor sizing is somewhat complicated especially within complex circuits. Sizing a transistor to speed up one signal path may slow down another due to the capacitive loading effect of path interactions. This technique should therefore be used with caution. The algorithms which

are presented in [12, 13] illustrate a linear method to make a trade-off between power, area, and delay in CMOS circuits. In [14], the relationship between transistor sizes and total circuit delay has been considered as nonlinear. It has been shown by Fishburn and Dunlop [14] that the transistor sizing problem is a convex under the simple lumped RC model.

Hence, a transistor sizing tool to meet our requirements in the performance optimization of VLSI circuits is crucial. An appropriate algorithm is the one that is simple to analyze and implement, decreases the execution time, increases the optimization rate of the goal parameter, and offers flexibility in choosing different optimization parameters.

Selecting the optimization factor is dependent on our needs. At present, for many researches, Power-Delay Product (PDP) which is metric for energy consumption of a circuit, is vitally important and the transistors has been sized to meet the minimum PDP [1, 2, 6, 15]. Several approaches have been applied in transistor sizing, one being a Mathematical optimization method [17–21]. The transistor sizing problem is formulated as a constrained nonlinear mathematical program of optimization factors. MDE (Minimum Delay Estimation) and ADC (Area-Delay Curve) algorithms [22] are within this group. Another approach is a Heuristic approach which was proposed for the first time in TILOS [14]. In this method, the transistor sizes have been changed iteratively until optimization is reached. In another approach, the combination of these two methodologies has been used. In this way a two-stage approach to combine the advantages of the heuristic and the mathematical programming techniques have been proposed. After using a heuristic method to perform an initial sizing, a timing analyzer and a mathematical optimizer are utilized to optimize the design. In this paper, through the review of the pros and cons of various transistor sizing approaches, a systematic and effective algorithm to size the transistors of various full adder cells for minimal energy consumption is suggested.

## III. Implementation Work

This method has been used in previous work to find delay. Brocco modelled gate delay as the time to reach 20% of final value after the input from the previous stage has reached 20% of its final value [12]. Kayssi used a similar method, using Vil as the data point [27]. Weste and Eshraghian modelled the time taken for a logic transition to pass from input to output as the time difference between the 50% level of the input transition to the 50% output level [35]. To investigate the best data point with which to obtain delay values, voltage transition graphs of several gates were analyzed. The input drive, block delay and output drive were obtained with different data points. The voltage transition graph of a node is found to be initially very dependent on the rate of change of the input. Taking delay values from 25% (close to threshold voltage) of input transition to 25% of output transition results in a small input drive value, comparable with that of the output drive. As the data point is moved upwards from 25% to 50% to 70%, the input drive increases, while the output drive decreases. This indicates that the last portion of the transition graph is heavily dependent on the output load.

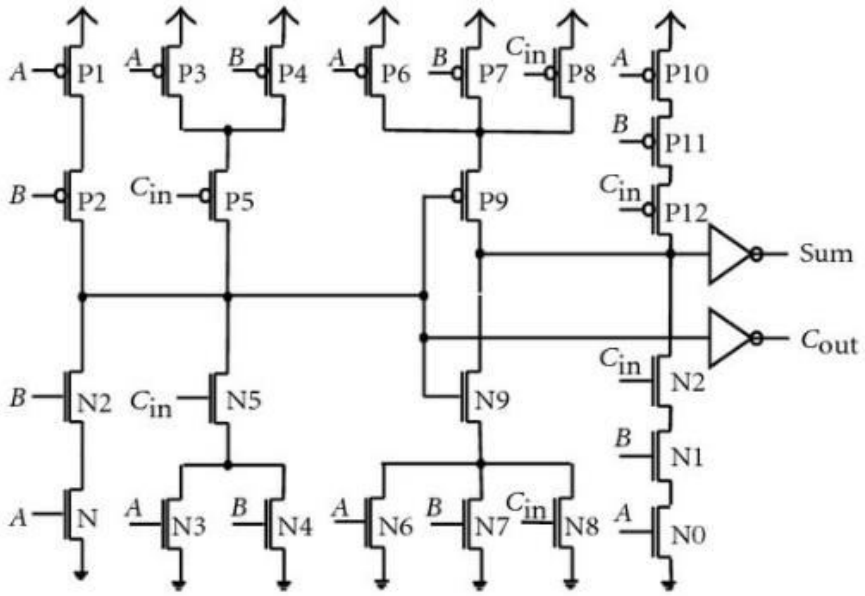



FIG 2: CMOS Design

In a multi-objective optimization problem, such as the full-adder circuit design, different and sometimes conflicting objective functions such as power consumption and propagation delay must be optimized simultaneously [8, 10, 11]. One solution to such optimization problems with multiple objective functions is to combine the values resulted for each objective with different weights (Wi) to obtain a proper fitness value. In this approach, the multi-objective problem converts to a single-objective optimization problem, but to use this method properly, the values of the weights should be selected by the designer. Another approach is to obtain a set of solutions that can optimize the maximum possible number of objective functions. This set of the optimal solutions titled as “Pareto”. Firstly, the initial values for WN, WP are randomly selected and the performance parameters (objectives) of propagation delay and power consumption are calculated regarding the mentioned analytical equations. Then, two separate functions in the Pareto front for propagation delay and power consumption should be defined. These two functions have two variables Wn and Wp, which were initially calculated through the fitness function. The resulting values of Wn and Wp are then replaced into two functions.

A change of gate could have effects on many other gates, since the change in delay could propagate throughout the circuit. To update the delay of the whole circuit at every change of gate version would be computationally intensive. To avoid such continuous updates while keeping accurate knowledge of delay values as gate versions are changed, the critical path is traversed sequentially and delay values on the path updated as versions are changed. This ensures a fairly accurate update of critical delay values needed in the optimization process. The process is described in detail below. After mapping the circuit with minimum-sized gates, a power estimation routine is run on the circuit to determine the switching probabilities of each node. Static timing analysis is also performed to determine the required, arrival and slack times at each node.

**Transistor Sizing Based on CMOS:**

One of the transistor sizing approaches is to use mathematical techniques. Logical effort is a technique to solve the transistor sizing problem in this way. Two algorithms, MDE and ADC, which are dealt with, are based on logical effort. The work [22] shows how these algorithms can size the transistors without running a heuristic sizing tool by calculating the minimum achievable delay and the cost of achieving a target delay.

This approach estimates the size of transistors without incurring the overhead of running a sizing tool. In this manner, different implementations are evaluated based on two metrics. First, the problem of estimating the minimum delay is considered. This metric allows a designer to determine whether an implementation can meet a given delay specification. The delay of a circuit is the maximum delay of all *Input* to *Output* paths of the circuit. In order to meet design goals, transistor sizing is applied to the circuit to reduce this delay. The smallest delay value that can be obtained in this way is referred to as the minimum achievable delay. Due to the associated high area overheads, most circuits other than their critical paths are rarely sized in order to meet this minimum delay value. In addition, the minimum achievable delay along with the unsized circuit delay helps to determine the range of delay value over which an implementation can be used. In short, these two algorithms present a technique that estimates the minimum achievable delay of a circuit which then goes to trace the area-delay curve.

**Heuristic Transistor Sizing Algorithms:**

By optimizing the transistor sizes of the circuits, it is possible to reduce the delay without significantly increasing the power consumption and transistor sizes can be set to achieve minimum PDP [1]. To provide a fair and insightful evaluation of circuits, a systematic and effective way of sizing the transistors for optimal performance is necessary. To provide a good trade-off between the conflicting sizing requirements for power and delay performances, the goal of optimization is to minimize the power-delay product, that is, the energy consumption. Chang's algorithm is an appropriate algorithm for sizing the transistors of a circuit and it is suitable for optimizing PDP [2]. The characteristics and functionality of this convergent algorithm are similar to our proposed algorithm.

Initially in Chang's algorithm, the sizes of the transistors in the circuit are reasonably set. The scaling operations are carried out in several iterations transistor by transistor. In [2], $(Ti)$ is the width of the $i$th transistor at step $j$ and $\Theta k$ is the PDP of the circuit of the $k$th iteration. For every optimization iteration, one transistor at a time is tuned for minimal PDP in $2\times m$ steps with a step resolution of $\pm\psi$. The optimization stops when the performance difference in two successive iterations is smaller than a given error$\varepsilon$. More than one iteration may be necessary because each time a new transistor is sized in the current run, the other transistors sized in the previous run may no longer maintain their optimality [2].

**Transistor Sizing in Multi-level Space:**

Transistor sizing is not a linear issue on the basis that modification of a transistor size within a circuit influences the circuit performances. In this space, we encounter different sizes and different parameters. Therefore, we must focus on an $n$-dimensional space and in general one of the parameters in this space will be optimized, meaning that; $n-1$ dimensions are related to $Kis$. In this paper, $Ki$ is the coefficient which is assigned for $i$th group of transistors in a circuit. The initial value of each $K$ is in fact the initial coefficient of channel width of transistors in a group.

In other words $Kis$ are known as the sizing factors of transistors and the $n$th parameter is the actual target which we desire to optimize. This target can be any of the circuit specifications including that of power consumption, delay, power-delay product, chip area, or their combination with a given weight. If transistor sizing is used to optimize only one target parameter, then the other parameters of the circuit may be neglected and may even lead to weak designs. For this reason, we can apply a product of two or more target parameters with determined weights in forming a new parameter. Then transistor sizing is performed for this new parameter [6]. As an example, the $K1$ to $K4$ are sizing optimization coefficients of transistors in illustrated gray groups which is shown in the XOR/XNOR circuit in Figure 1. These coefficients can optimize parameters such as the power-delay product. Therefore, transistor sizing must be performed in a five dimensional space.

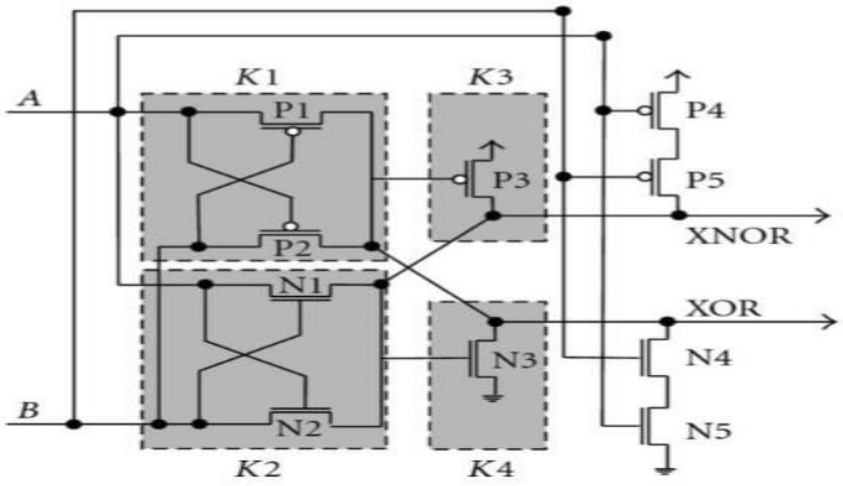
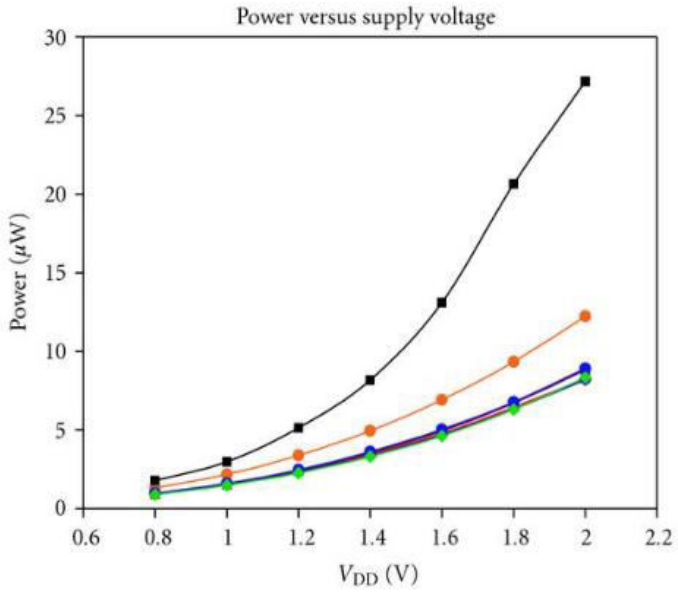



FIG 3: The XOR/XNOR part of NEW HPSC Full Adder.

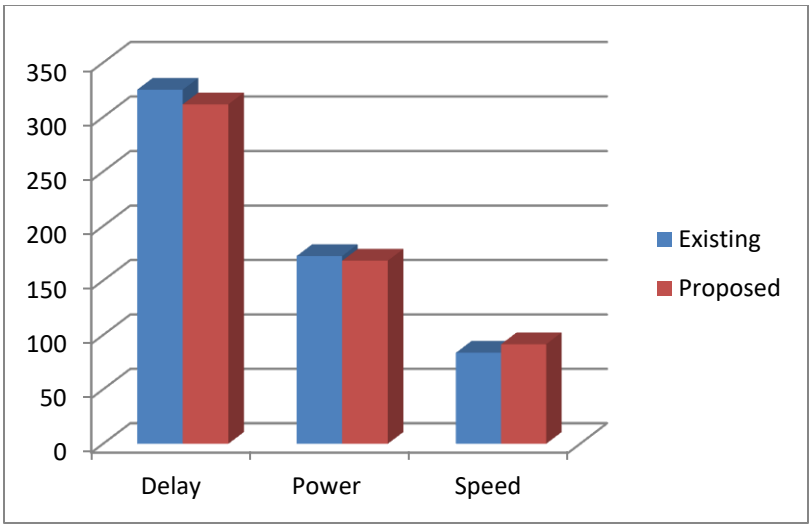
Hence, in an $n$-dimensional space, the $Kis$ are sequentially modified until the target parameter arrives at the best possible position. In fact, changing the size of a transistor in an $n$-dimensional space can modify the circuit specifications. With this in mind, changing the size of transistor $Ti$ in an $n$-dimensional space, initiates the process of transistor sizing.

## IV. Results

The seven full adder circuits of Figure 5 are all simulated to achieve the optimum power-delay product (PDP) using Chang's algorithm and the proposed transistor sizing algorithms. Optimization of the transistor sizing is carried out at seven different voltages, 0.8 V, 1.0 V, 1.2 V, 1.4 V, 1.6 V, 1.8 V, and 2.0 V. The step size of the subsequent iterations of these algorithms is set to 0.1 $\mu$m. Thus, the final transistor sizes have the precision of 77% of the channel length, which is 0.13-$\mu$m for our targeted technology. The final transistor widths for the New HPSC full adder cell which is optimized by both of the algorithms at different supply. The power, delay and power-delay product at supply voltage ranges from 0.8 V to 2.0 V of these full adder

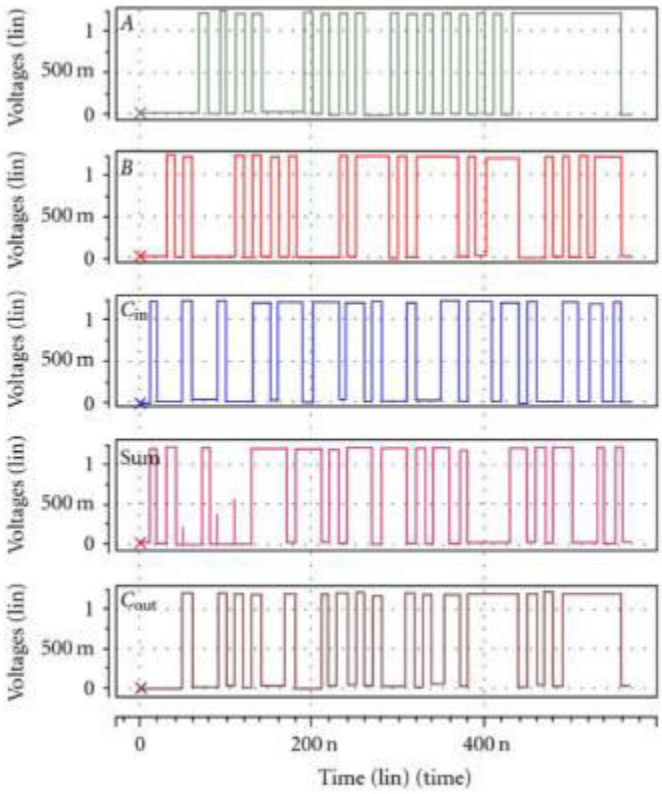



| | Existing | Proposed |
|---|---|---|
| **Delay** | 325.22 | 311.74 |
| **Power** | 172.52 | 168.24 |
| **Speed** | 83.68 | 91.54 |




| gate name | pin | input pin load | rising output (ns) | | | falling output (ns) | | |
|---|---|---|---|---|---|---|---|---|
| | | | 1/ input drive | block delay | 1/ output drive | 1/ input drive | block delay | 1/ output drive |
| nand2 | a1 | 1.12 | 0.105 | 0.755 | 0.448 | 0.031 | 0.694 | 0.307 |
| | a2 | 0.99 | 0.094 | 0.700 | 0.415 | 0.068 | 0.799 | 0.316 |

In order to experience all the possible transitions as inputs, an input test pattern with 56 transitions must be applied to a full adder circuit. An input transition may or may not result in a change at the output node. Even if there is no switching activity at the output node, some internal nodes may be switching. This switching activity results in some power dissipation. Thus, for an accurate result, all the possible input combinations are considered for all the test circuits [1]. Therefore, a group of input test patterns which offers all the 56 different transitions from one input combination to another are used as the input vectors for the full adder cell. Figure 4 shows the input stimulus, and Sum and *C*out outputs




## V. Conclusion

This study presents a novel approach to transistor sizing optimization for automated digital integrated circuit design. The approach utilizes a multi-objective genetic optimization algorithm, which is based on the behavioural equations of the circuit. An accurate design and optimization tool for digital integrated circuit design was created by combining analytical equations with circuit simulations. To achieve the design goals, the transistor sizes can be optimized by the suggested design algorithm. Using 0.18 µm CMOS technology model parameters, a CMOS full-adder circuit was successfully constructed and simulated in MATLAB and HSPICE to confirm the effectiveness of the suggested algorithm. As shown in this paper, the proposed transistor sizing algorithm not only reduces the power consumption of the full-adder circuit, but also optimizes the transistor dimensions, which reduces the circuit's internal parasitic capacitances and resistances, propagation delay, and PDP. the performances of the full adder circuit are effectively improved. Finally, digital integrated circuit designers can benefit from the proposed approach discussed in this paper to design high-performance VLSI circuits in future manufacturing technologies.

## References


[1] RA Vural, O Der, T Yildirim. Particle swarm optimization based inverter design considering transient performance, Digital Signal Processing. 2010; 20: 1215-1220.

[2] BV Krishna, K Dhanunjaya. Analysis and Modeling of Low Power Array Multipliers using Cadence Virtuoso Simulator in 45nm, Development (IJECIERD). 2013; 3: 49-64.

[3] F Kashfi, S Hatami, M Pedram. Multi-objective optimization techniques for VLSI circuits, 12th International Symposium on Quality Electronic Design. 2011; 1-8.

[4] P Gupta, H Mandadapu, S Gourishetty, Z Abbas. Robust Transistor Sizing for Improved Performances in Digital Circuits using Optimization Algorithms, 20th International Symposium on Quality Electronic Design (ISQED). 2019; 85-91.

[5] D. Radhakrishnan, "Low-voltage low-power CMOS full adder," *IEE Proceedings: Circuits, Devices and Systems*, vol. 148, no. 1, pp. 19–24, 2001.

[6] T. Nikoubin, S. Pouri, P. Bahrebar, and K. Navi, "A new transistor sizing algorithm for balanced XOR/XNOR circuits," in *Proceedings of the 12th Annual International CSI Computer Conference*, Tehran, Iran, February 2006.

[7] M. Aguirre and M. Linares, "An alternative logic approach to implement high-speed low-power full adder cells," in *Proceedings of the 18th Annual Symposium on Integrated Circuits and Systems Design*, pp. 166–171, Florianopolis, Brazil, September 2005.

[8] K. Navi, O. Kavehie, M. Rouholamini, A. Sahafi, and S. Mehrabi, "A novel CMOS full adder," in *Proceedings of the 20th IEEE International Conference on VLSI Design*, pp. 303–307, January 2007.

[9] J. M. Rabaey, A. Chandrakasan, and B. Nikolic, *Digital Integrated Circuits (A Design Perspective)*, Prentice-Hall, Upper Saddle River, NJ, USA, 2nd edition, 2002.

[10] S. S. Sapatnekar and W. Chuang, "Power-delay optimizations in gate sizing," *ACM Transactions on Design Automation of Electronic Systems*, vol. 5, no. 1, pp. 98–114, 2000.

[11] M. R. Berkelaar and J. A. Jess, "Gate sizing in MOS digital circuits with linear programming," in *Proceedings of the European Design Automation Conference*, pp. 217–221, 1990.

[12] M. R. C. M. Berkelaar, P. H. W. Buurman, and J. A. G. Jess, "Computing the entire active area/power consumption versus delay trade-off curve for gate sizing with a piecewise linear simulator," in *Proceedings of the IEEE/ACM International Conference on Computer-Aided*

[13] J. P. Fishburn and A. E. Dunlop, "TILOS: a posynomial programming approach to transistor sizing," in *Proceedings of the International Conference on Computer Aided Design (ICCAD '85)*, 1985.

[14] M. Haghparast and K. Navi, "Design of a novel fault tolerant reversible 12 full adder for nanotechnology based systems," *World Applied Sciences Journal*, pp. 114–118, 2008.

[15] K. Navi, M. Maeen, V. Foroutan, S. Timarchi, and O. Kavehei, "A novel low-power full-adder cell for low voltage," *Integration, the VLSI Journal*, vol. 42, no. 4, pp. 457–467, 2009.