A comprehensive comparison between LE and LM-based methodologies for optimisation of digital circuits

Kunwar Singh*

Department of Electrical Engineering,
Delhi Technological University,
Room No. FW1-SF1, EED, DTU,
New Delhi-110042, India
E-mail: kunwarsingh@dce.ac.in
*Corresponding author

Satish Chandra Tiwari and Maneesha Gupta

Division of ECE, Netaji Subhas Institute of Technology,
University of Delhi,
Room No. 16, Division of ECE, NSIT, Sector-3,
Dwarka, New Delhi-110078, India
E-mail: stiwari@cadence.com
E-mail: maneeshapub@gmail.com

Abstract: This paper presents a comprehensive comparison between Levenberg-Marquardt (LM) and logical effort (LE) theory-based optimisation techniques. While LM is a classical approach for optimisation and is embedded in SPICE, logical effort-based approach is contemporary, design specific and needs simple back of the envelope calculations for optimisation of digital circuits. Both the approaches have been used by digital circuit designers in the literature for comparing a proposed digital circuit with the existing designs while optimising any given design for timing, power and area parameters. The goal of writing this paper is to make digital system designers gain an insight into the procedures used for optimising digital circuits while at the same time a well-defined approach using LM algorithm is provided that can be easily automated with the current generation CAD tools. For the purpose of comparison some standard circuits were chosen and optimised for minimum PDP and PDAP. SPICE simulations have been extensively used for comparing the two methodologies in a 180 nm\1.8 V CMOS technology.

Keywords: VLSI; digital CMOS circuits; logical effort theory; Levenberg-Marquardt algorithm; power-delay product; power-delay-area product.

Biographical notes: Kunwar Singh received his BTech in Electronics and Communication Engineering and MTech in VLSI Design from GGSIP University, New Delhi, in 2006 and 2009. He is currently serving as an Assistant Professor in the Department of Electrical Engineering, Delhi Technological University and is working towards his PhD degree in the University of Delhi. His research interests include low power high performance digital circuits. He has authored and co-authored over 13 research papers in the low power high performance VLSI domain in various international/national journals and conferences.

Satish Chandra Tiwari received his BSc and MSc in Electronics from University of Delhi and Bundelkhand University, Jhansi in 2005 and 2007 respectively. He received his MTech in VLSI Design from Indraprastha University, New Delhi, in 2009. He is currently pursuing his PhD in Electronics Engineering at the Division of ECE, NSIT, University of Delhi. He has authored and co-authored over 11 research papers in the low power VLSI domain in various international/national journals and conferences.

Maneesha Gupta received her BE in Electronics and Communication Engineering from Government Engineering College, Jabalpur in 1981, ME in Electronics and Communication Engineering from Government Engineering College, Jabalpur in 1983 and PhD in Electronics Engineering (Analysis, Synthesis and Applications of Switched Capacitor Circuits) from, IIT Delhi in 1990. She is currently working as a Professor in Division of ECE, NSIT, New Delhi from 2008. Her teaching and research interests are switched capacitors circuits, low voltage/power design techniques and analogue signal processing. She has authored and co-authored over 35 research papers in the above areas in various international/national journals and conferences.

1   Introduction

The overwhelming demand for portable electronic devices has led to a revolution in digital system design in recent times. The large-scale production of these battery operated systems is associated with the need for designing increasingly power efficient IC design without compromising on performance. However, with the VLSI technology scaling down to transistors with channel length of few nanometres the circuit complexity has increased manifolds. Furthermore, it has been a challenge for the digital system designers to optimise circuits with such high complexity in a three dimensional design space viz. power, performance and area.

Digital systems are mainly composed of combinational and sequential components. Several attempts have been made in the past to realise and compare the existing combinational and sequential elements. While propagation delay (worst delay out of rising and falling delay) remains the timing parameter, power dissipation (static and dynamic) is considered for power-related comparisons as far as comparative analysis of combinational circuits is concerned. However, for sequential circuits flip-flop is the basic component and there remained ambiguity regarding the selection of appropriate timing

parameters of flip-flop configurations till the last decade. The correct definition of flip-flop timing parameters was presented by Stojanovic and Oklobdzija (1999) which proposed data-to-output delay as the performance parameter and not clock-to-output delay unlike the previous works (Pedram et al., 1998). Moreover, power dissipation was also divided into three components internal power, clock power and data power. Accordingly, power-delay product was considered as the figure of merit (FOM) by most designers and it was extensively used in the literature for comparing the various combinational and sequential circuit designs (Chung et al. 2002; Aezinia et al., 2006; Nedovic et al., 2002; Tschanz et al., 2001; Strollo et al., 2005).

This paper also addressed the problem of optimising a FF for minimum power-delay product using Levenberg-Marquardt (LM) algorithm embedded in SPICE. However, the approach seemingly correct suffered from three major flaws:

1 in terms of the fixation of upper bound on transistor widths

2 the delay and power characteristics change significantly with the capacitive load

3 finally computational effort is more since no fixed methodology for obtaining the upper bounds is present.

Nearly a decade after this study, another generalised approach towards optimising a flip-flop was introduced by Alioto et al. (2010a, 2010b, 2010c, 2011) based on logical effort (LE) theory. The major contribution of this paper is in terms of a well-defined methodology to fix the upper bound on transistor width for optimisation in a three dimensional space viz. energy (power), delay, and area. This became possible due to introduction of delay sensitivity factor which in turn determines practical design constraints and can be effectively applied to other digital circuits (combinational in nature) apart from the flip-flops as shown in this paper. The second problem was addressed and an analysis of load sensitive flip-flop characterisation was provided in (Heo and Asanovic, 2001). The solution to the third problem is provided in this paper by utilising the upper bounds obtained from the LE approach and reducing the computational effort of the LM methodology. This paper presents a detailed comparison of these two methodologies clearly stating the existing differences using simulation results of standard digital circuit configurations.

The rest of this paper is organised as follows. Section 2 provides an insight into LE theory-based approach for optimisation of digital circuits. It also describes the proposed LM algorithm-based methodology for optimisation of digital circuits. Section 3 describes the simulation parameters and test bench. It also highlights techniques used for transistor sizing and methodology adopted for optimisation of timing, power-delay product and power-delay-area product. Section 4 compares the two methodologies using some standard circuits while optimising the designs for minimum PDP and PDAP and simultaneously highlighting and discussing some major issues with both the methodologies based on the simulation results. Finally, the conclusion is summarised in Section 5. An Appendix is added to show calibration of parameters for delay calculations using LE theory.

2   Comparison of LE and LM-based methodologies

2.1   LE theory

LE theory is based on the topology of a logic gate and the magnitude of the capacitive load that the gate drives. As the load increases, the delay increases; this delay is also dependent on logical function of gate. Hence, the LE theory utilises both the input and output capacitances as well as topological variations in its derived equations. LE theory uses inverter as basic block and derives the LE of any other gate with reference to the basic inverter (Sutherland et al., 1998). LE theory expresses the absolute delay as the product of unit less delay (d) of the gate and the basic delay unit (τ) characterised by particular fabrication process node.

$$d_{abs} = d \cdot \tau \tag{1}$$

Typically, t is about 12 ps for 180 nm/1.8 V CMOS technology. Now again d can be divided into two parts:

1   Fixed delay, i.e., parasitic delay.

2   Delay due to load on gates output (called effort delay or stage effort 'f').

$$d = f + p \tag{2}$$

The effort delay 'f' is dependent on load and properties of logic gate driving that load.

$$f = g \cdot h \tag{3}$$

where 'g' is the LE and 'h' is the electrical effort.

The LE 'g', hence represents how much worse the gate is in producing output current as compared to inverter, given that all other parameters are same. Electrical effort 'f' defines the effect of electrical environment of logic gate on performance and effects of size of transistors on load driving capability.

$$h = c_{out} / c_{in} \tag{4}$$

where c_{out} is the output load capacitance and c_{in} is the capacitance presented by logic gate at one of its input terminal. Hence,

$$d = g \cdot h + p \tag{5}$$

This equation holds true for single stage gates.

2.1.1   LE method for gates with multiple stages

LE method states that the optimised delay D of a path of N cascaded stages is

$$D = N\sqrt[N]{GBH} + P \tag{6}$$

$$D = N\sqrt[N]{F} + P \tag{7}$$

where G, B, H (= C_L/C_{in}), P, F(= GBH) and C_L are the LE, branching effort, electrical effort, parasitic delay, path effort and final load capacitance respectively.

$$D = P(1 + t) \tag{8}$$

From (6) and (8)

$$t = \frac{N\sqrt[N]{GB}\sqrt[N]{C_L}}{P\sqrt[N]{C_{in}}} \tag{9}$$

where t is the relative delay increment with respect to parasitic delay. Equations (8) and (9) indicate that increasing C_{in} to larger values leads to a reduction in the optimised delay and increasing C_{in} beyond a particular value has no significant effect on the flip-flop latency. At this point, delay for the circuit is considered to be saturated. Based on the above analysis, the delay sensitivity factor introduced by Alioto et al. (2010c) is used to obtain the upper bound on the transistor widths to explore the power-delay design space with minimal computational effort.

$$S_D^{C_{in}} = \frac{\partial D}{\partial C_{in}} \frac{C_{in}}{D} = -\frac{1}{N} \frac{t}{t+1} \tag{10}$$

where $S_D^{c_{in}}$ is the delay sensitivity factor and can be obtained from equations (7) to (9).

Optimisation is done under a varying FF input capacitance C_{in} using LE theory. For a fixed C_L and C_{in}, LE theory optimises the delay in the critical path. Since the transistors in the critical path affect the flip-flop speed hence these widths are determined for different C_{in} values to optimise the FF for minimum delay using equation (11).

The upper bounds on the normalised transistor widths w (normalised with respect to W_{min}) are obtained such that the delay sensitivity tends to a minimum value S_{min} which indicates the delay saturation with respect to a particular flip-flop topology. The S_{min} value for our analysis is selected to be −2%.

The input capacitance C_{in} (expressed in femtofarad) of the flip-flop is determined from normalised width w1 (absolute width W1 optimised with respect to W_{min}) of transistors in the first stage of the D-Q path as shown below:

$$C_{in} = (w1 \cdot 360 + 2 \cdot w1 \cdot 360)(1.15 \cdot 10^{-3}) \tag{11}$$

Figure 1 shows the conventional TGFF design. The size of transistors in the critical path are assumed to be independent design variables (IDVs) and optimised for maximum performance in accordance with LE theory. An inverter is added in the first stage before TG in the critical path to protect the input terminal from noise variations.

Table 1 shows the variation of delay with increasing C_{in} values. It is to be noted that the delay saturates at 153 ps for C_{in} = 24.8 fF. This leads to the determination of upper bound on transistor widths early in the design phase and hence defines the limits of

power (energy)-delay design space. The table also lists the corresponding power measured along with the power-delay product. It is readily observed that minimum power-delay product is obtained at $C_{in} = 9.92$ fF.

Figure 1   Transistor sizing methodology for TGFF using LE theory
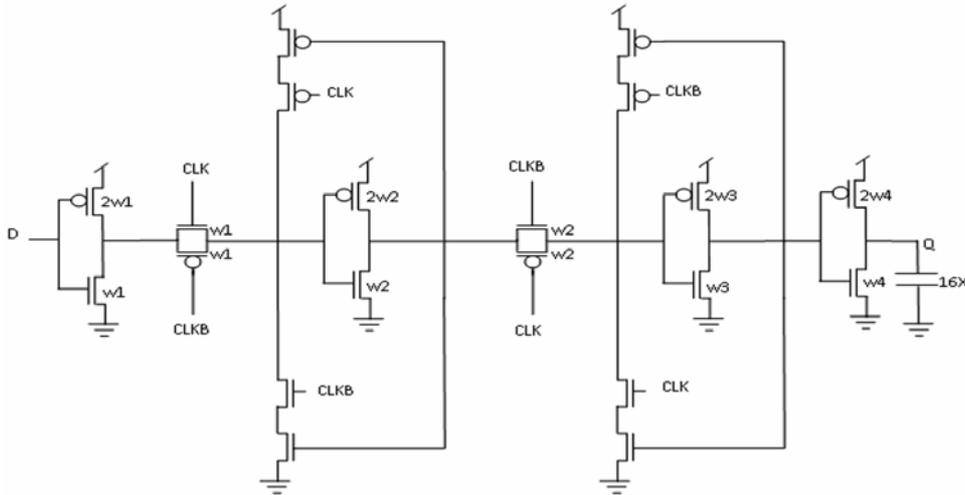
Table 1   Conventional TGFF at 19.92fF load (16×)

C_{in} (fF)	w1	w2	w3	w4	$T_{DQ,min}$ (ps)	Power (uW)	PDP (fJ)	PDAP (fJ.um)
2.48	2	2.35	2.79	6.65	226	554	125.2	21,852
4.96	4	3.95	3.95	7.91	191	585	111.7	20,400
7.44	6	5.35	4.84	8.76	173	599	103.6	19,708
9.92	8	6.65	5.59	9.41	166	615	102	20,140
12.4	10	7.86	6.25	9.95	162	632	102.3	20,891
14.8	12	9.01	6.85	10.4	159	648	103	20,876
17.3	14	10.1	7.40	10.8	157	665	104.4	21,916
19.8	16	11.1	7.91	11.2	155	675	104.6	22,634
22.3	18	12.2	8.39	11.5	154	682	105	23,387
24.8	20	13.2	8.84	11.8	153	689	105.4	24,135

Figure 2 LE-based optimisation methodology in flow chart format
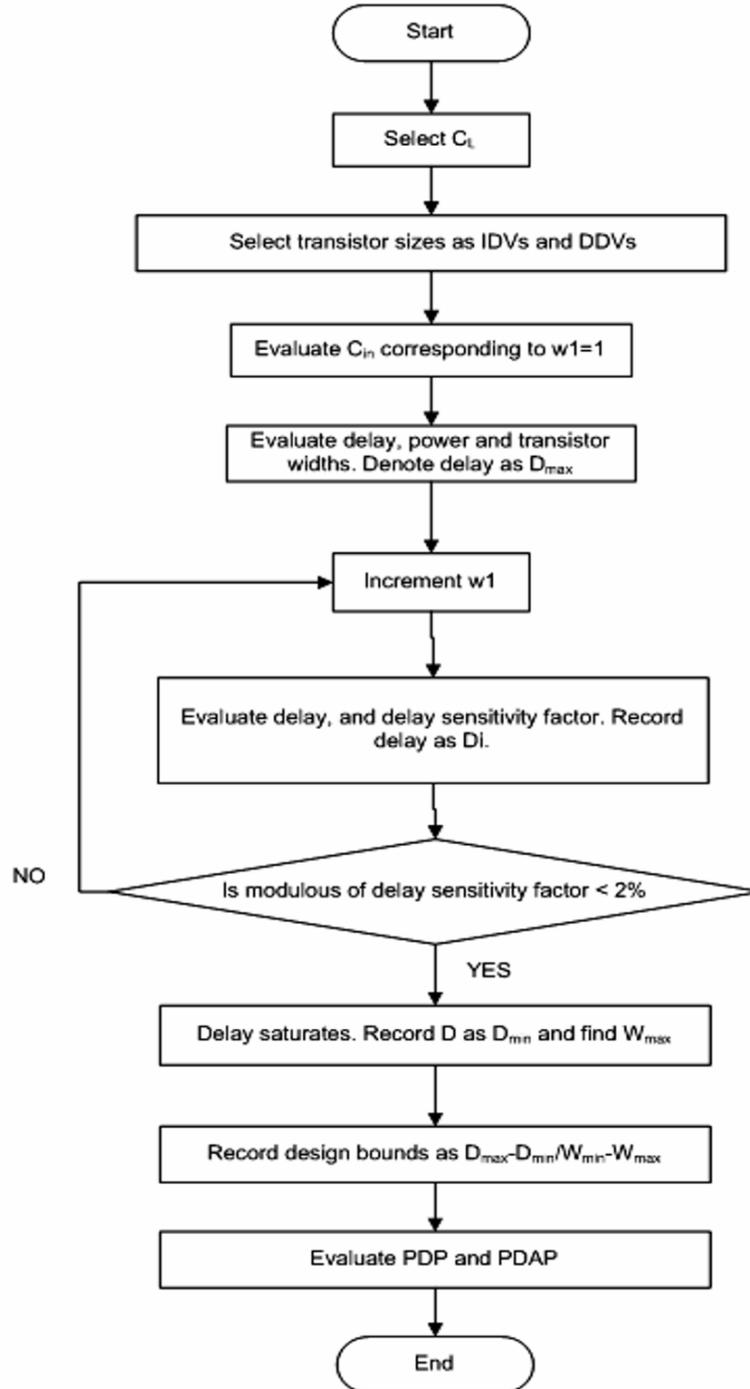
2.2   *LE-based methodology*

Figure 2 illustrates the optimisation methodology based on LE theory a description for which is as follows:

Step 1   To start with the capacitive load C_L is selected.

Step 2   The transistors in the circuit topology are identified as IDVs and DDVs. IDVs lie on the critical path determine the performance of the circuit and need to be optimised whereas DDVs (generally including feedback transistors, keepers, gated keepers) are kept at minimum possible widths which ensure correct circuit functionality.

Step 3   Set C_{in} to its minimum value corresponding to w1 = 1, evaluate transistor sizes and corresponding delay and power, denote the delay value as D_{max}.

Step 4   Repeat the above process by incrementing w1 by 1 in each step while simultaneously checking whether modulus of delay sensitivity factor is less than 2%. For each step record, the delay values as Di (Di values correspond to the worst-case delay out of rising and falling delays) and power and area values as P_i and A_i, respectively. If $S_D^{c_{in}}$ is less than 2% record the corresponding delay value as Di = D_{min} and determine transistor sizes and the maximum transistor width in the circuit denoted as W_{max}.

Step 5   Design space is fixed in terms of delay as $D_{min} - D_{max}$ and as $W_{max} - W_{min}$ in terms of transistor widths.

Step 6   Evaluate PDP_i and $PDAP_i$ (area is evaluated as the sum of transistor widths corresponding to each design point).

2.3   *LM algorithm*

A detailed description of the operating characteristics of LM algorithm is beyond the scope of this work, however, a brief overview is presented to aid the researchers gain some basic insight into the functioning of this popular optimisation approach.

The LM algorithm is used in the domain of mathematics and computing for locating minima of a function expressed in terms of non-linear functions. This approach is widely utilised for a broad range of engineering disciplines while the most frequent application is in the least squares curve fitting problem. It is considered to be a combination of Gauss Newton and gradient descent method (Levenberg, 1944; Marquardt, 1963; Lourakis, 2013). The algorithm operates based on the steepest descent approach while the fast convergence towards the minima is attributed to the Gauss Newton method.

Let us assume that the sum of the squares of the deviations S(a) is to be minimised:

$$S(a) = \sum_{i=1}^{n} \left[v_i - f(u_i, a) \right]^2 \tag{12}$$

while optimising the parameters 'a' of the curve f(u, a) for a set of n datum pairs of independent and dependent variables (u_i, v_i).

LM algorithm is an iterative technique and an initial guess is needed for the parameter vector 'a' to be determined to achieve the minimisation. The basis of the LM algorithm is

that during each iteration step, replacement of the parameter vector 'a' is done by a new estimate, a + t, while t can be determined using the following equation:

$$f(u_i, a + t) = f(u_i, a) + J_i t \tag{13}$$

where J is the Jacobian matrix $\dfrac{\partial f(u_i, a)}{\partial a}$.

The set of linear equations finally solved to determine t are given as:

$$(J^T J) t = J^T [v - f(a)] \tag{14}$$

2.4 Proposed LM-based methodology for optimisation

Figure 3 illustrates the optimisation methodology based on LM theory a description for which is as follows:

Step 1   As an initial step C_L is fixed.

Step 2   The circuit topology is analysed for critical path. The width of the transistors lying on the critical path is set to variable Si while the transistor sizes not belonging to the critical path are fixed at minimum possible widths ensuring correct circuit operation.

Step 3   Select design space Di as $D_{max} - D_{min}$ while providing the input width range as $W_{min} - W_{max}$, obtained from LE theory.

Step 4   The circuit functionality is checked by setting all the transistor widths to minimum. If correct circuital operation is not obtained the transistor widths are incremented by W_{min} and the process is repeated till correct operation is obtained and finally the lower limit of the transistor width range is fixed as the value for which appropriate functionality is achieved.

Step 5   If however, correct functionality is achieved for W_{min} then target delay is set to D_{min} and first run of optimisation is invoked to check if convergence is achieved for width range $W_{min} - W_{max}$. If not, the value of W_{max} is repetitively increased by W_{min} till the target delay converges.

Step 6   As convergence is obtained, transistor widths and corresponding power dissipation is also measured and the process is repeated for all points in the delay design space varying from $D_{max} - D_{min}$.

Step 7   PDP and PDAP corresponding to each design points are obtained.

The most significant disadvantage of LM approach is the lack of a fixed approach which specifies the lower and the upper bound. This leaves the designer with no idea for limiting the design space. In this work, a solution is provided for the problem by defining both the upper and the lower limit of transistor sizes based on minimum delay value obtained through delay sensitivity factor of LE-based design approach. A unique feature of the proposed optimisation process has been the balancing of rising and falling delays for each delay point in the design space.

To gain a better insight into the optimisation methodologies the following designs were optimised for minimum PDP and PDAP.

1   Three input NAND gate.

2   CMOS buffer chain using four inverters.

3   A NAND-based two-bit multiplexer.

4   Transmission gate flip-flop (TGFF).

5   8-bit asynchronous counter design.

Figure 3   LM-based optimisation methodology in flow chart format
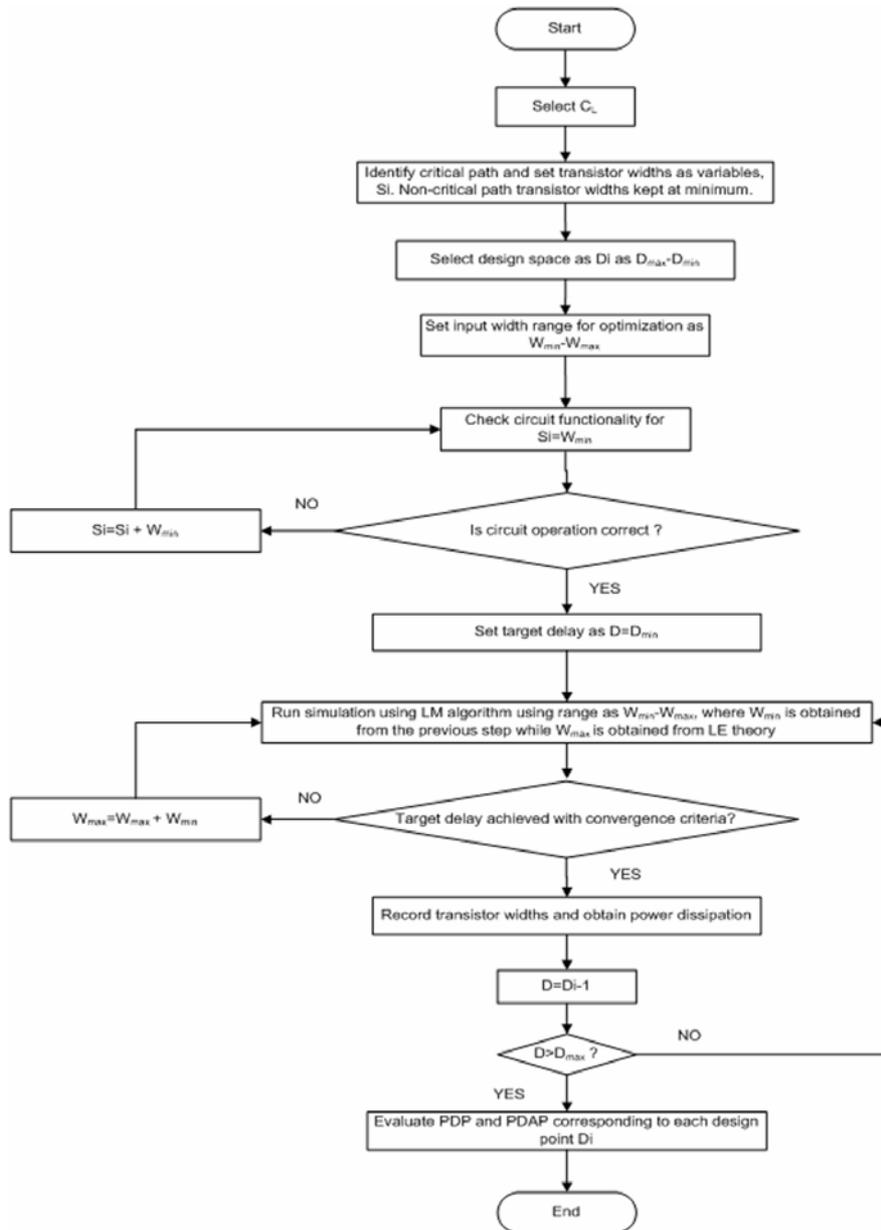
3 Simulation parameters and test bench

The CMOS parameters used for simulation are listed in Table 2. All the circuits were designed using 180 nm CMOS process technology and a supply voltage of 1.8 V. The minimum technology width in case of feedback transistors was fixed at 360 nm (W_{min}) while the rising and falling transitions of data and clock signals were limited by a slope of 100 ps in case of flip-flops and CMOS buffer and 1 ns for three input NAND gate.

Table 2 CMOS simulation parameters

W_{min}	L_{min}	C_{min}	V_{dd}	Frequency	Risetime/falltime
360 nm	140 nm	1.2 fF	1.8 V	250 MHz	100 ps

Figure 4 shows the setup used to characterise and compare the FF designs. Data and clock buffers have been used to provide realistic clock and data signals. Data-to-output delay (T_{dq}) has been adopted as the performance parameter (Oklobdzija et al., 2003). The delay sensitivity factor introduced by Alioto et al. (2010c) based on LE theory has been used for speed optimisation.

Figure 4 Test bench
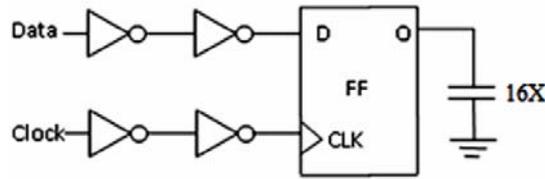
A 16 cycle long pseudorandom sequence with an activity factor of 50% is supplied at the data input for average power measurements in case of flip-flops and static CMOS buffer chain (Stojanovic and Oklobdzija, 1999). Moreover, power dissipation of three input NAND gate is measured over 100ns producing all the eight input bit combinations over the specified period. Transistor sizing methodology adopted is the same as that in Alioto et al. (2010c). Power-delay product PDP and power-delay-area product PDAP have been chosen as the FOM.

The relation between absolute gate capacitance (C_{GATE}) in terms of fF (femtofarads) and absolute transistor width (W) in terms of nanometres (nm) has been derived for Berkeley Predictive Technology Models at 180 nm by fitting simulation data and was found to be

$$C_{GATE} = (1.15 \cdot 10^{-9}) \cdot W \qquad (15)$$

Absolute delay based on LE theory, D_{abs} is obtained by multiplying parameter D with parameter τ as shown in the equation below.

$$D_{abs} = D \cdot \tau \qquad (16)$$

The value of process dependent parameter τ is determined as approximately 12 ps using the calibration technique as mentioned by Sutherland et al. (1998). The detailed procedure is discussed in the Appendix. The measured delays based on LE method

obtained using SPICE are in close agreement with the mathematical results (typically within 10% error).

4   Simulation results and discussion

4.1   *Three input NAND gate*

The first circuit optimised is a three input NAND gate. NAND gate is the most frequently used logic gate in digital circuits and any logic can be realised using NAND gate because it is a universal gate. Although the design of a NAND gate might seem to be trivial, but even a small reduction in the power dissipation of NAND gate is quite significant mainly because large number of NAND gates are used for designing digital systems and low power design of such systems is a major bottleneck for designers. The transistor sizing methodology for three input NAND gate based on LE approach is shown in Figure 5 while Figure 6 shows the transistor sizes for LM-based methodology.

Figure 5   Transistor sizing of three input NAND gate using LE
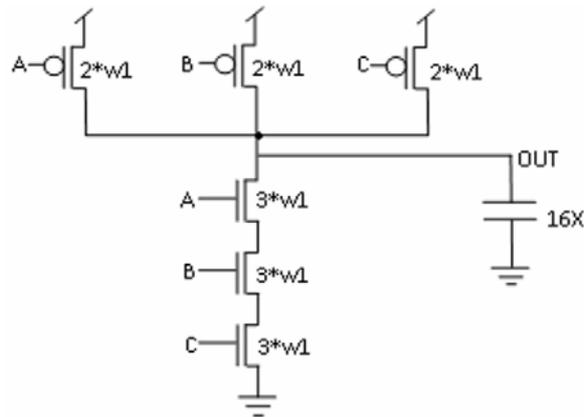
Figure 6   Transistor sizing of three input NAND gate using LM
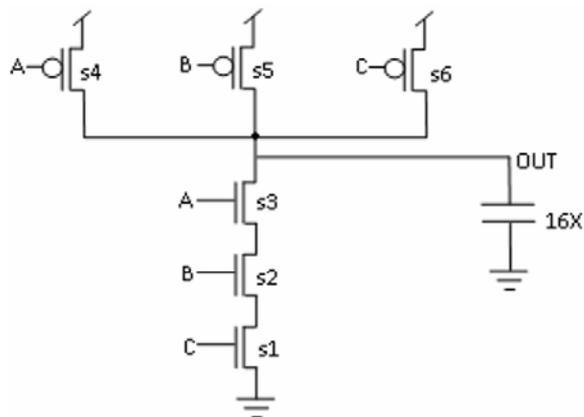
Table 3 shows the results obtained in terms of delay, power, area, PDP and PDAP obtained after optimising the circuit using both the methodologies. In case of LE-based approach C_{in} is progressively increased till the delay is saturated. Once the delay saturates, it can be concluded that increasing the transistor sizes further leads to unnecessary increase in the circuit area and power consumption. Thus, upper bound on the transistor size as well as minimum achievable delay is obtained. Now, the LM approach can be limited to a design space where delay varies between 145 ps and 192 ps (192 ps corresponds to the minimum C_{in}). The upper limit on transistor size in accordance with the LE-based approach was determined to be 6.48 um. However, due to convergence issues the range was finally fixed to 0.36 um–10 um while optimising the design with LM methodology. Those delay points were selected for analysis which converged successfully with respect to the specified criteria within the above mentioned delay limits and optimised widths s1–s6 corresponding to each delay point were obtained. Thereafter, the corresponding power dissipation was measured for 100 ns using the optimised set of widths.

Table 3 Simulation results for three input NAND gate

C_{in} (fF)	Delay (ps)		Power (uW)		Area (um)		PDP (fJ)		PDAP (fJ.um)	
	LE	LM	LE	LM	LE	LM	LE	LM	LE	LM
2.48	192	192	3.61	3.05	5.4	3.8	0.69	0.58	3.74	2.22
4.96	164	180	6.04	3.44	10.8	4.54	0.99	0.61	10.6	2.81
7.44	154	165	8.42	4.07	16.2	5.93	1.29	0.67	20.9	3.98
9.92	150	160	10.8	4.96	21.6	8.12	1.63	0.79	35.2	6.43
12.4	147	153	13.3	4.86	27.0	7.82	1.95	0.74	52.8	5.81
14.8	145	142	15.6	6.41	32.4	11.7	2.26	0.91	73.5	10.6

It can be clearly observed from Table 3 that the power and area requirements continuously scale to higher levels with a corresponding increase in C_{in}, using LE method in comparison to LM-based approach. As a result, three input NAND gate shows 20% improvement in optimal PDP whereas optimal PDAP is also reduced by 40.6%.

The area calculation is based on the sum of widths of optimised transistor sizes. Accordingly, the area corresponding to LE-based approach is

$$\text{Area (LE approach)} = 3w1 + 3w1 + 3w1 + 2w1 + 2w1 + 2w1 = 15w1$$

where wi represent normalised transistor widths (normalised with respect to w_{min}) obtained using LE theory.

$$\text{Area (LM approach)} = \sum_{i=1}^{15} si$$

where si represents the optimised transistor widths obtained from application of LM algorithm.

4.2 Static CMOS buffer with four inverters

Following similar procedures as mentioned in Section 2, the simulation results obtained for a static CMOS buffer chain consisting of four inverters are presented in Table 4. This circuit is mainly used to drive capacitive loads in large circuits. Figure 7 shows the strategy used for transistor sizing using LE theory while Figure 8 shows the transistor sizes in terms of appropriate variables in accordance with the LM approach. The upper limit of transistor size as obtained from LE was 3.75 um whereas the range of transistor was finally fixed to 1 um–5.4 um due to non-convergence of delay points. It can be readily observed from Table 4 that in case of static CMOS buffer, LE outperforms LM approach in the design space as improvement in optimum PDP and PDAP are 18.9% and 43.2%, respectively.

$$\text{Area (LE approach)} = w1 + 2w1 + w2 + 2w2 + w3 + 2w3 + w4 + 2w4$$
$$= 3(w1 + w2 + w3 + w4)$$

$$\text{Area (LM approach)} = \sum_{i=1}^{8} si$$

Figure 7   Transistor sizing of static CMOS buffer using LE
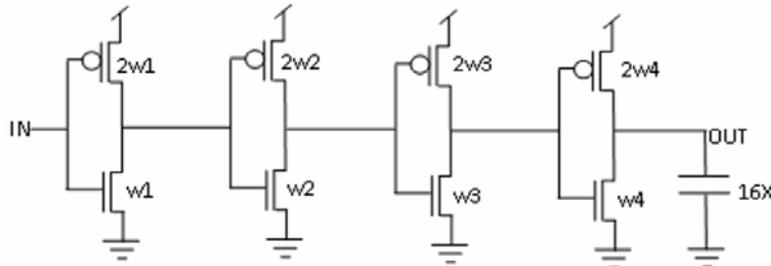
Figure 8   Transistor sizing of static CMOS buffer using LM
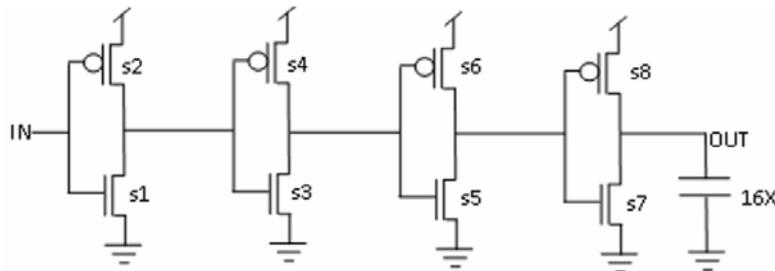
Table 4   Simulation results for static CMOS buffer

C_{in} (fF)	Delay (ps)		Power (uW)		Area (um)		PDP (fJ)		PDAP (fJ.um)	
	LE	LM	LE	LM	LE	LM	LE	LM	LE	LM
2.48	147	147	28	36	15.81	23.89	4.11	5.29	64.9	126.3
7.44	145	145	43	35	26.73	22.56	6.23	5.07	166.5	114.3

4.3 A NAND-based two-bit multiplexer

The third circuit is NAND-based two-bit multiplexer which belongs to the class of combinational circuits. A multiplexer is also one of the most frequently employed components while designing ASICs and digital full custom integrated circuits. The transistor sizing technique for optimising the multiplexer using LE approach has been shown in Figure 9. It is worth pointing out that the circuit consists of two paths Path_1 and Path_2 from select line input 'S' to the output. The longest path consists of three stages, where the first stage NAND_4 of Path_2 lies in parallel with the first two stages INV_1 and NAND_2 of Path_1. The delays for both the paths are balanced by equalising the delays of first two stages of Path_1 with the first stage of Path_2. Figure 10 shows the transistor sizing methodology using the LM approach.

Figure 9 Transistor sizing of NAND-based two-bit multiplexer using LE
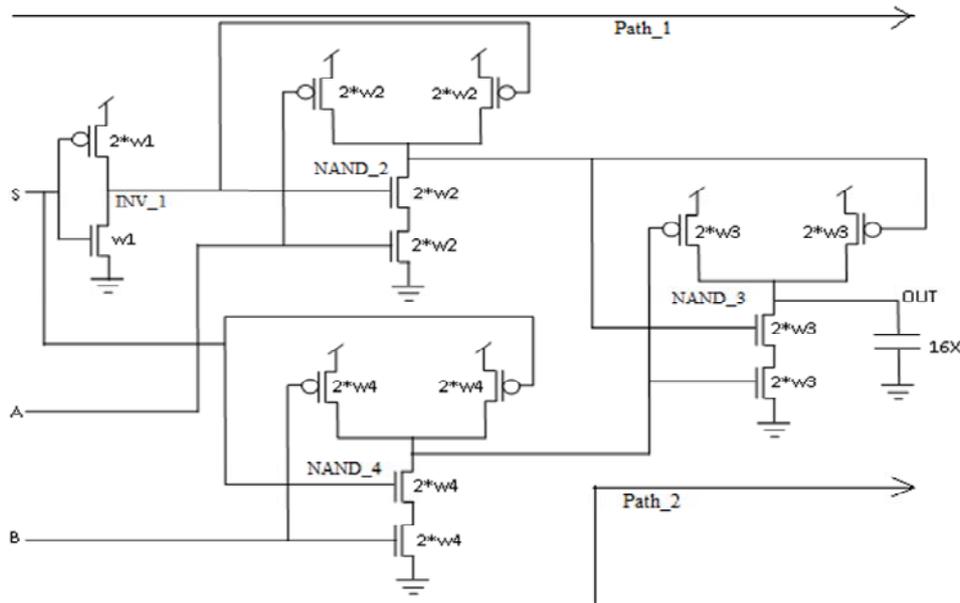
Table 5 Simulation results for NAND-based two-bit multiplexer

C_{in} (fF)	Delay (ps)		Power (uW)		Area (um)		PDP (fJ)		PDAP (fJ.um)	
	LE	LM	LE	LM	LE	LM	LE	LM	LE	LM
1.24	133	133	87	71	22.84	31.63	11.57	9.44	264.25	298.58
2.48	118	118	110	65.4	35.84	31.75	12.98	7.71	465.20	244.79
3.72	110	110	126	66.2	43.96	32.38	13.86	7.28	609.28	235.72
4.96	105	105	141	66.9	50.90	32.84	14.80	7.02	753.32	230.53
6.20	102	102	154	67.3	57.72	33.20	15.70	6.86	906.20	227.75

Figure 10   Transistor sizing of NAND-based two-bit multiplexer using LM
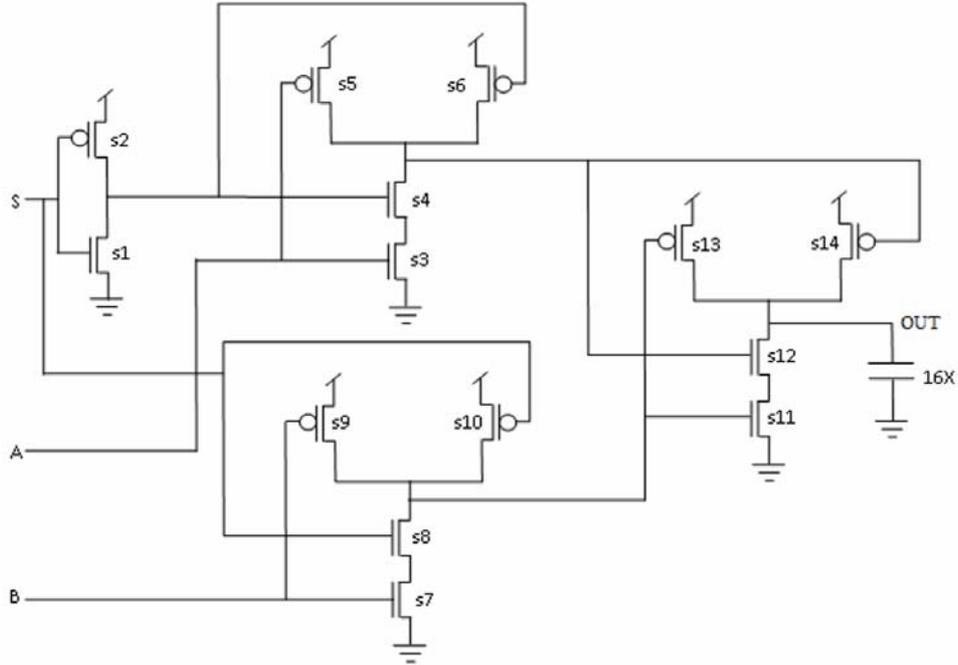
The maximum transistor size corresponding to the saturated delay of 102 ps obtained using LE theory was 6.44 um. This was fixed as the upper bound of LM-based optimisation algorithm while the correct circuit functionality is observed at a lower limit of 1.08 um. Finally, the range for optimising the multiplexer using LM approach was fixed as 1.08 um–6.44 um. Again, the trend continues and it is observed that the power and area increase by significant amounts when delay is reduced from D_{max} to D_{min} in case of LE-based methodology whereas the variations in area and power are marginal when LM-based optimisation is used. The improvements in PDP and PDAP are reported as 40.7% and 13.8%, respectively in Table 5.

$$\begin{aligned}
\text{Area (LE approach)} &= w1 + 2w1 + (2w2 + 2w2 + 2w2 + 2w2) \\
&\quad + (2w3 + 2w3 + 2w3 + 2w3) \\
&\quad + (2w4 + 2w4 + 2w4 + 2w4) \\
&= 3w1 + 8(w2 + w3 + w4)
\end{aligned}$$

$$\text{Area (LM approach)} = \sum_{i=1}^{14} si$$

4.4   *Transmission gate flip-flop*

The third circuit optimised is a TGFF. A flip-flop is an indispensable component for design of synchronous sequential systems and flip-flops and latches are responsible for 30% to 70% power dissipation of a digital system along with the clocking network. The

maximum operating frequency of a system is determined by the latency of flip-flops as they are present at the starting and end points of signal delay paths (Yeap, 1998). TGFF belongs to the master-slave class of flip-flops which are generally used for low power applications (Gerosa et al., 1994). Moreover, TGFF is the best known flip-flop in terms of power-delay trade-off among all flip-flop categories. The transistor sizing methodology for the TGFF using LE technique is shown in Figure 11. The primary task to perform LE-based optimisation is to identify the critical path. The transistor sizes on the critical path are assumed to be IDVs and need to be optimised for highest performance while assuming the rest as dependent design variables (DDVs) and keeping the aspect ratio of rest of these transistors at minimum. Figure 12 demonstrates the transistor sizing methodology using the LM approach, again transistors in the input-to-output path need to be optimised whereas the size of rest of the transistors are fixed at minimum technology widths.

Figure 11   Transistor sizing of TGFF using LE
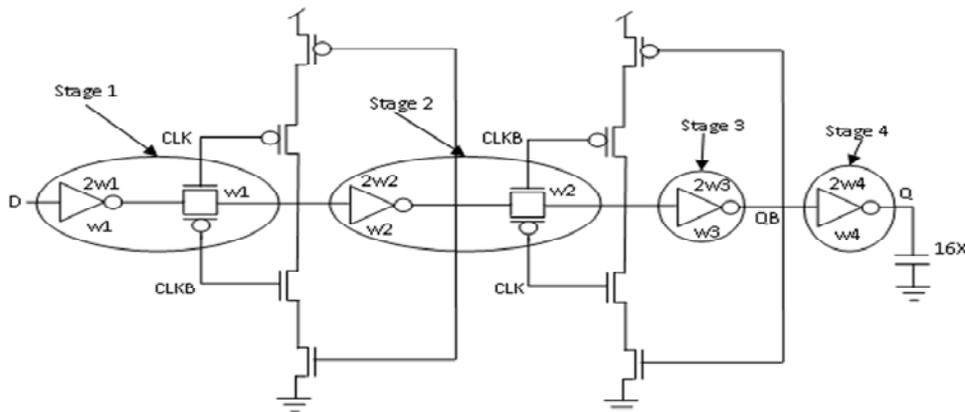
Figure 12   Transistor sizing of TGFF using LM
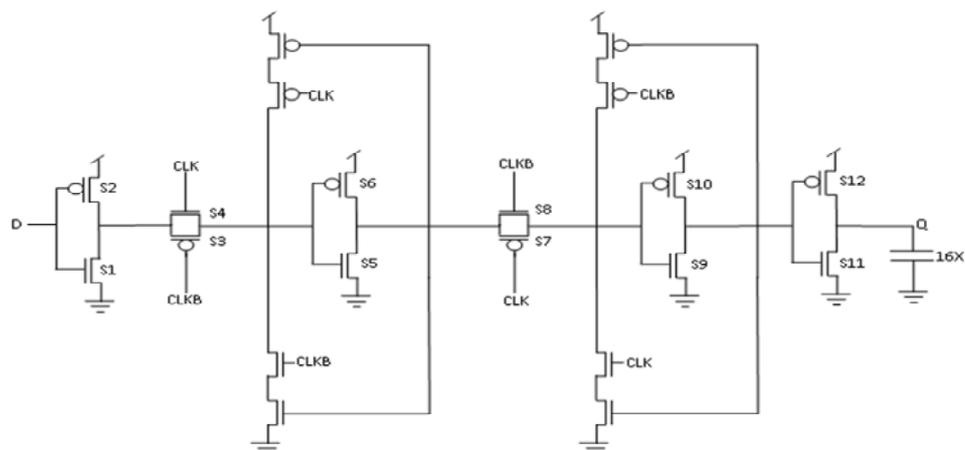
Table 6   Simulation results for TGFF

C_{in} (fF)	Delay (ps)		Power (uW)		Area (um)		PDP (fJ)		PDAP (fJ.um)	
	LE	LM	LE	LM	LE	LM	LE	LM	LE	LM
2.48	226	196	554	511	176.8	181.6	125.2	100.1	21852	18160
4.96	191	159	585	502	185.8	176.6	111.7	79.8	20400	14092
7.44	173	151	599	550	193.6	182.3	103.6	83.05	19708	15140
9.92	166	175	615	552	201.4	185.8	102	96.6	20140	17948
12.4	162	146	632	548	208.5	191.6	102.3	80	20891	15328

The saturated value of delay as obtained from the LE theory is 162 ps. Accordingly, the upper limit on transistor size in accordance with the LE-based approach was determined to be 6.58 um. However, due to convergence issues the range was finally fixed to 0.72 um–10 um.

It is to be noted here that the lower limit was increased to 0.72 um because with the initial transistor size of 0.36 um for each transistor as executed in the first iteration of the LM algorithm, no correct functionality was obtained and this resulted in failure of the LM approach.

Again significant improvements are observed in terms of optimum PDP and PDAP. Table 6 indicates that while the delay is assumed to be saturated at 162 ps in accordance with the LE theory the circuit can still be optimised for lower delay values using the LM approach. Hence, the performance of a system can be upgraded by 10% using the LM approach as compared to the LE approach. Also, note that the PDP and PDAP at highest operating frequency are respectively 21.7% and 26.6% lesser for LM-based approach as compared to the LE methodology.

$$\begin{aligned} \text{Area (LE approach)} = &\{w1 + 2w1 + w1 + w1 + w2 + 2w2 + w2 + w2 \\ &+ w3 + 2w3 + w4 + 2w4\} \\ &+ \{8(0.36u)\} + \{4(12u + 24u)\} \\ = &5w1 + 5w2 + 3w3 + 3w4 + 8(0.36u) + 4(12u + 24u) \end{aligned}$$

{transistors in the critical path} + {transistors in the non-critical path}

+{data and clock buffers}

$$\text{Area (LM approach)}: \sum_{i=1}^{12} si + 8(0.36u) + 4(12u + 24u)$$

{transistors in the critical path} + {transistors in the non-critical path}

+{data and clock buffers}

4.5   Design of 8-bit asynchronous counter

An 8-bit asynchronous counter was implemented by converting the D flip-flop configuration to a T flip-flop configuration using an EXOR gate as illustrated in the Figure 13.

The T flip-flop designed using TGFF is shown in Figure 14. It is considered to be a five stage design and optimised for highest speed using both LE theory and the classical

LM approach. The EXOR gate was realised using transmission gates as revealed in Stage 1 of Figure 14. The transistor sizing of toggle flip-flop based on LM methodology is shown in Figure 15.

Figure 13 Conversion of D flip-flop to toggle flip-flop
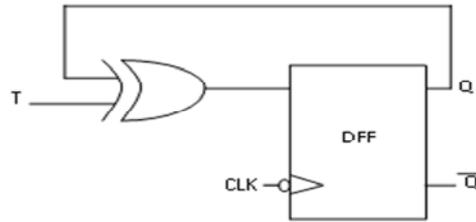
Figure 14 Transistor sizing of toggle flip-flop using LE
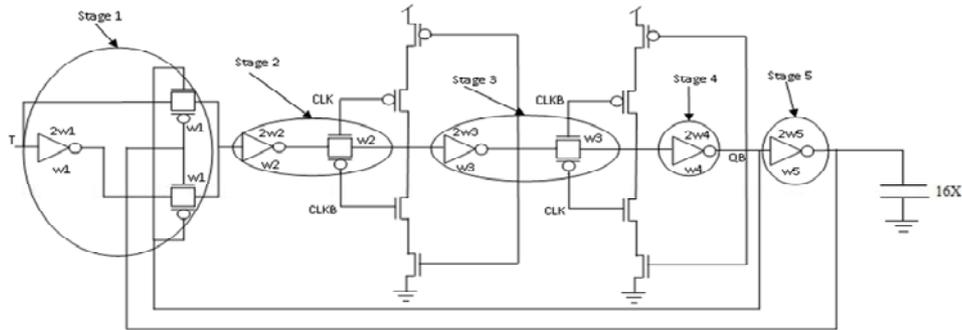
Figure 15 Transistor sizing of toggle flip-flop using LM
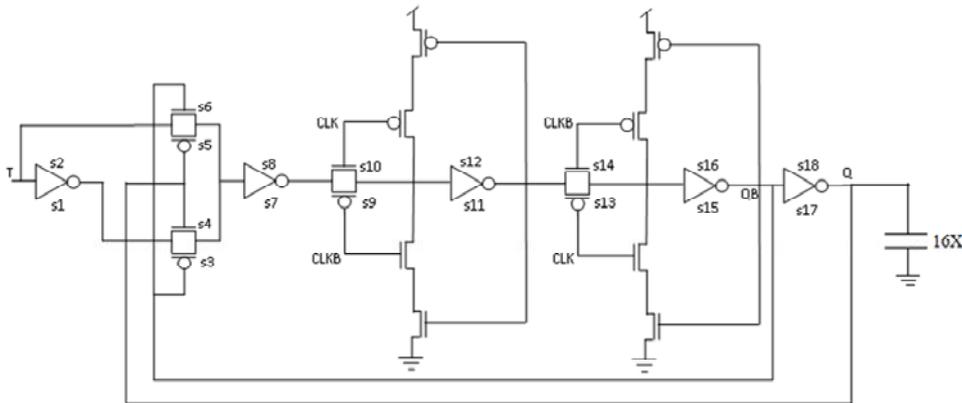
$$\begin{aligned}
\text{Area (LE approach)} &= \{w1 + 2w1 + w1 + w1 + w2 + 2w2 + w2 + w2 \\
&\quad + w3 + 2w3 + w3 + w3 + w4 + 2w4 + w5 + 2w5\} \\
&\quad + \{8(0.36u)\} + \{4(12u + 24u)\} \\
&= 5w1 + 5w2 + 5w3 + 3w4 + 3w5 + 8(0.36u) + 4(12u + 24u)
\end{aligned}$$

{transistors in the critical path} + {transistors in the non-critical path}

+{data and clock buffers}

$$\text{Area (LM approach)}: \sum_{i=1}^{18} si + 8(0.36u) + 4(12u + 24u)$$

{transistors in the critical path} + {transistors in the non-critical path}

+{data and clock buffers}

Table 7   Simulation results for toggle flip-flop

C_{in} (fF)	Delay (ps)		Power (uW)		Area (um)		PDP (fJ)		PDAP (fF.um)	
	LE	LM	LE	LM	LE	LM	LE	LM	LE	LM
24.8	170	170	542	438	208.44	166.56	92.14	74.46	19,205.6	12,402

Table 8   Optimised widths for toggle flip-flop for a target delay of 170ps at 16× load

Optimised transistor widths using LE approach (m)	Optimised transistor widths using LM approach (m)
w1 = 3.60e-006	s1 = 7.2000e-007
w2 = 2.96e-006	s2 = 7.2000e-007
w3 = 2.47e-006	s3 = 7.2000e-007
w4 = 2.05e-006	s4 = 7.2000e-007
w5 = 3.42e-006	s5 = 7.2000e-007
	s6 = 7.2000e-007
	s7 = 6.0413e-006
	s8 = 8.5821e-007
	s9 = 7.7523e-007
	s10 = 7.2000e-007
	s11 = 7.9791e-007
	s12 = 1.2513e-006
	s13 = 9.3030e-007
	s14 = 7.2348e-007
	s15 = 1.0320e-006
	s16 = 7.3646e-007
	s17 = 7.2378e-007
	s18 = 8.4023e-007

The minimum latency for the toggle flip-flop was obtained using the LE theory as 170 ps. Now, the toggle flip-flop was designed using 170 ps as the delay target and corresponding power dissipation, PDP and PDAP were obtained for both LE and LM-based approaches. Table 7 indicates that power consumption of a TGFF-based toggle flip-flop is 19.1% lower while PDAP shows improvement by 35.4%.

For designing the modulo 256 counter, the output Q of each stage is connected to the clock terminal of the next stage through two intermediate inverters (acting as a buffer) sized (W_p = 11.52 u, W_n = 5.76 u) such that the input capacitance of the first inverter acts as the load capacitance for the corresponding flip-flop configuration as depicted in Figure 16. As a result, the load at the output terminal of each flip-flop is uniformly fixed at 19.92 fF. Since the toggle flip-flops are already optimised for maximum speed at 16× (19.92 fF) load, the optimised transistor widths from Table 8 are used to realise the counter and are used for power measurements. The average power dissipation of each counter is estimated over 256 clock cycles at varying frequencies.

Figure 16 Schematic of 8-bit asynchronous flip-flop
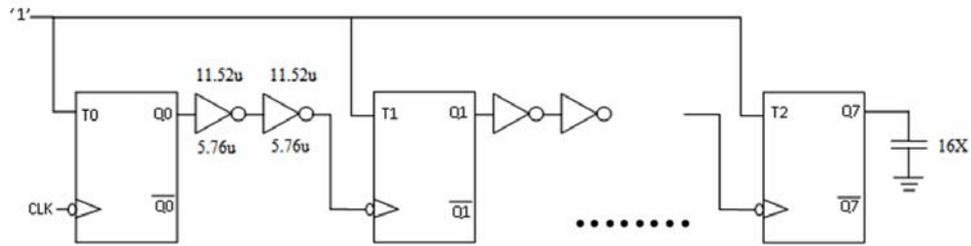
Figure 17 Power dissipation of LE and LM-based counters with varying frequencies
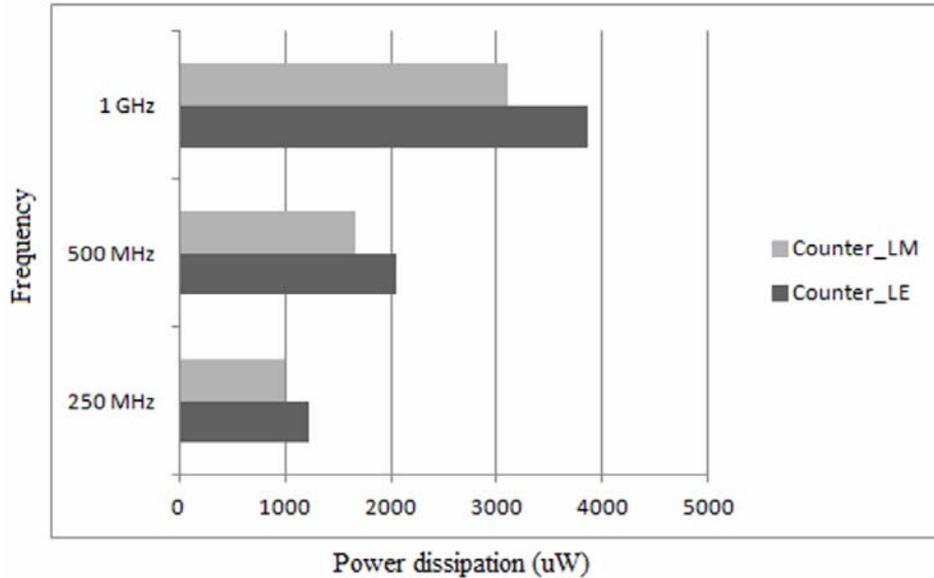
Figure 17 shows the variation in power dissipation at different frequencies. The counter realised using LM approach dissipates 18.5%, 18.8% and 19.6% lesser power at 250 MHz, 500 MHz, and 1 GHz, respectively.

Figure 18 Comparison of PDP for different circuits using LE and LM methodology*
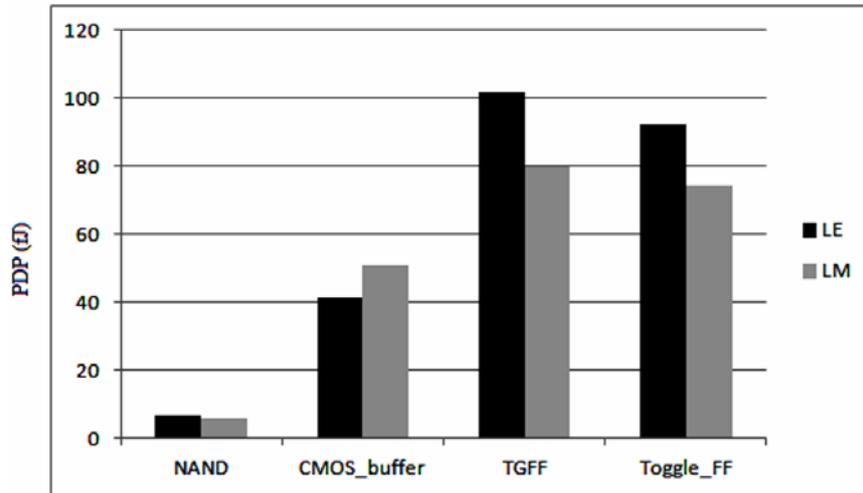
Note: *The PDPs of three input NAND gate and static CMOS buffer are scaled by
10 times.

Figure 19 Comparison of PDAP for different circuits using LE and LM methodology*
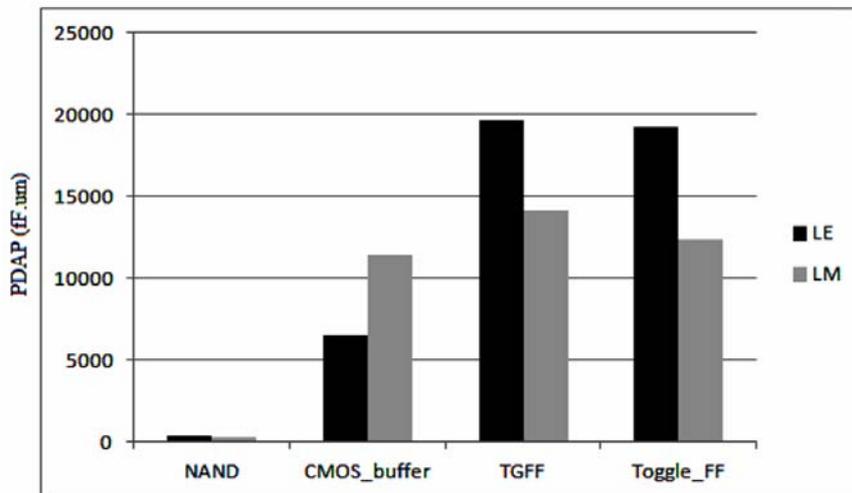
Note: *The PDAPs of three input NAND gate and static CMOS buffer are scaled by
100 times.

By inspection of Figure 18 and Figure 19 it is clear that LM-based methodology results in better optimisation of PDP and PDAP for three input NAND gate, TGFF and toggle flip-flop whereas LE-based methodology provided better optimisation of static CMOS buffer both in terms of PDP and PDAP.

5 Conclusions and future scope

LM and LE-based methodologies used for optimisation of digital circuits have been thoroughly investigated in this paper. The problem of fixing the upper bounds in case of LM-based optimisation approach is emphasised and a solution is provided on the basis of delay sensitivity factor belonging to the LE theory which leads to a reduction in the computational effort. The LM-based methodology is drafted in a manner such that it can be easily automated and embedded into the modern day CAD tools. Some standard digital circuits are used for comparing the two methodologies in terms of optimal PDP and PDAP. Interestingly, simulation results have revealed that the performance of a system utilising TGFF can be upgraded by up to 10.4% with a simultaneous improvement in PDP and PDAP of 21.7% and 26.6% when compared to the optimisation for maximum performance in case of LE-based methodology. Moreover, it can be easily concluded that specifying the upper bounds on transistor widths using LE theory for GA, ACO, PSO-based algorithms or any other emerging optimisation algorithms can also result in significant reduction in the computational effort for determining the optimum transistor sizes in the power(energy)-delay-area design space.

References

Aezinia, F., Najafzadeh, S. and Afzali-kusha, A. (2006) 'Novel high speed and low power single and double edge-triggered flip-flops', *IEEE Asia-Pacific Conference on Circuits and Systems*, pp.1383–1386.

Alioto, M., Consoli, E. and Palumbo, G. (2010a) 'General strategies to design nanometer flip-flops in the energy-delay space', *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 57, No. 7, pp.1583–1596.

Alioto, M., Consoli, E. and Palumbo, G. (2010b) 'Analysis and comparison in the energy-delay-area domain of nanometer CMOS flip-flops: part I – methodology and design strategies', *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, No. 19, No. 5, pp.725–736.

Alioto, M., Consoli, E. and Palumbo, G. (2010c) 'Analysis and comparison in the energy-delay-area domain of nanometer CMOS flip-flops: part II – results and figures of merit', *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 19, No. 5, pp.737–750.

Alioto, M., Consoli, E. and Palumbo, G. (2011) 'DET FF topologies: a detailed investigation in the energy-delay-area domain', *IEEE International Symposium on Circuits and Systems*, pp.563–566.

Chung, W., Lo, T. and Sachdev M. (2002) 'A comparative analysis of low power low voltage dual edge triggered flip-flops', *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 10, No. 6, pp.913–918.

Gerosa, G., Gary, S., Dietz, C., Pham, D., Hoover, K., Alvarez, J., Sanchez, H., Ippolito, P., Tai, N., Litch, S., Eno, J., Golab, J., Vanderschaaf, N. and Kahle, J. (1994) 'A 2.2 W, 80 MHz superscalar RISC microprocessor', *IEEE Journal of Solid State Circuits*, Vol. 29, No. 12, pp.1440–1454.

Heo, S. and Asanovic, K. (2001) 'Load-sensitive flip-flop characterization', *Proceedings of IEEE Computer Society Workshop on VLSI*, pp.87–92.

Levenberg, K. (1944) 'A Method for the solution of certain non-linear problems in least squares', *Quarterly of Applied Mathematics*, Vol. 2, No. 2, pp.164–168.

Lourakis, M.I.A. (2013) 'A brief description of the Levenberg-Marquardt algorithm implemented by Levmar' [online] http://www.ics.forth.gr/lourakis/levmar/levmar.pdf/ (accessed 10 January 2013).

Marquardt D.W. (1963) 'An algorithm for the least-squares estimation of nonlinear parameters', *SIAM Journal of Applied Mathematics*, Vol. 11, No. 2, pp.431–441.

Nedovic, N., Aleksic, M. and Oklobdzija, V.G. (2002) 'Comparative analysis of double-edge versus single-edge triggered clocked storage elements', *IEEE International Symposium on Circuits and Systems*, Vol. 5, pp.105–108.

Oklobdzija, V., Stojanovic, V., Markovic, D. and Nedovic, N. (2003) *Digital System Clocking: High-Performance and Low-Power Aspects*, Wiley-IEEE Press, New York.

Pedram, M., Wu Q. and Wu X. (1998) 'A new design of double edge triggered flip-flops', *Proceedings of ASPDAC, Asia and South Pacific*, pp.417–421.

Stojanovic, V. and Oklobdzija V.G. (1999) 'Comparative analysis of master-slave latches and flip-flops for high-performance and low-power systems', *IEEE J. Solid-State Circuits*, Vol. 34, No. 4, pp.536–548.

Strollo A.G.M., Caro, D.D., Napoli, E. and Petra, N. (2005) 'A novel high-speed sense-amplifier-based flip-flop', *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 13, No. 11, pp.1266–1274.

Sutherland, I., Sproull, B. and Harris D. (1998) *Logical Effort: Designing Fast CMOS Circuits*, Morgan Kaufmann Publishers, San Francisco.

Tschanz, J., Narendra, S., Chen, Z., Borkar, S., Sachdev, M. and De V. (2001) 'Comparative delay and energy of single edge-triggered & dual edge triggered pulsed flip-flops for high performance microprocessors', *International Symposium on Low Power Electronics and Design*, pp.147–152.

Yeap, G.K. (1998) *Practical Low power Digital VLSI Design*, Springer, London.

Appendix

Calibration of Parameters for modelling delays using LE theory

The initial step in modelling delays using LE theory is to express all delays in terms of a basic delay unit τ which is a process dependent parameter. Thus, we define absolute delay as the product of a unit less delay of the gate given by equation (2), and the delay unit τ. Accordingly,

$$D_{abs} = D \cdot \tau$$

as explained earlier.

It needs to be observed that while D represents the delay for a multistage path, d expresses the delay of a single stage logic gate. To determine the absolute delays using LE theory process parameter τ needs to be determined. For this purpose, we calibrate by measuring the delay of a logic gate as a function of its load (electrical effort) using SPICE and fitting a straight line to the simulation results. Figure 20 shows simulated data for an inverter design. Since the LE of an inverter is 1, we expect from d = gh + p that the absolute delay will be d = (h + p)τ. The straight line that connects the points will have slope τ. In our case, the value of τ is estimated to be approximately 12 ps by fitting simulation data.

Figure 20   Simulated delay of inverters driving various loads
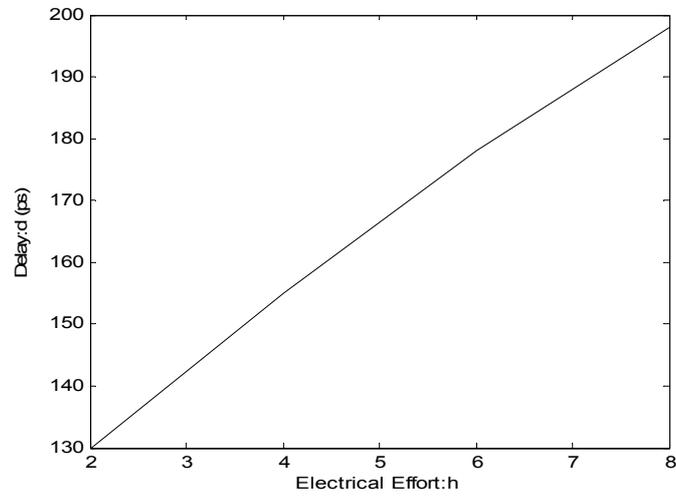
Note: Results from 180 nm, 1.8 V process.