Power Macro Modelling for CMOS Inverter of 0.12μm Technology

Shahzad Naseem[1,2], Saira Riaz[1,2], Muhammad Azam[1,2], Syed S. Ali[1,3], Yaseer A. Durrani[3]

[1]Microelectronics Division, Centre of Excellence in Solid State Physics, University of the Punjab, Lahore-54590, Pakistan
[2]School of Physical Sciences, University of the Punjab, Lahore-54590, Pakistan
[3]Electrical Engineering Division, Faculty of Engineering, University of the Punjab, Lahore-54590, Pakistan
shahzad_naseem@yahoo.com

Abstract - Power dissipation of very large scale integrated circuits (VLSI) has emerged as a significant constraint on the semiconductor industry. For dynamic power the voltage, capacitance and frequency are the major components of power dissipation. In this paper, we propose a power macro modelling technique for the CMOS inverter using 0.12μm technology. The dynamic power is directly linked with the load capacitance (C_L), and it is lumped as all internal parasitic capacitances. In our modelling, we take account of the parasitic capacitances with their dependence on channel length and width. Suitable values of other factors (i.e. threshold voltage V_T, gate voltage V_{GS}, drain voltage V_{DD} etc.) are used for power consumption of the CMOS inverter.

Index Terms - CMOS inverter, Power consumption, load capacitance, parasitic capacitance

1. Introduction

The parasitic capacitances are becoming an important issue for designing logic circuits wit h aggressive reduction of MOS transistor dimensions into the deep sub μm regime[1-5]. In digital applications, these parasitic capacitances have a very strong impact on propagation delay and the overall power dissipation of the circuit. For an analog application these capacitances cause a negative feedback, which again has an influence on gain-bandwidth product. Because of their important role in short channel region, the parasitic capacitances are required to be computed accurately to predict the circuit performance.

Over the years, several models have been developed for parasitic capacitances. With precise mathematical iterations, Kamchouchi et al.[6] derived semi-empirical model using Schwartz-Christoffel transformation. By using conformal transformation, Shrivastava et al.[7] developed a simple analytical model with the assumption that the potential near the gate electrode is constant between the silicon substrate and bottom of the gate electrode; this model is prone to errors. Afterwards, Suzuki[8] presented a model based on Shrivastava's model with accurate boundary conditions. Nihar et al[9] developed a model by taking the presence of source/drain electrodes and high K-gate dielectric material into account.

In this paper, we simplify two-capacitance model amongst the available five models[10].The model that we use is shown schematically in Figure 1, from which we compute the total dynamic power of CMOS inverter after getting lumped sum of parasitic capacitances.

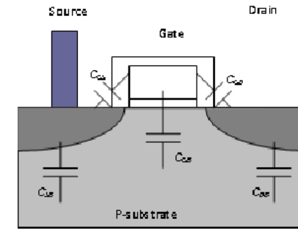

Fig. 1. Proposed MOS model with 3 capacitances

2. Power Macro Model

Power dissipation of a CMOS circuit is comprised of two types: static and dynamic[11,12] that can be expressed as:

$$P_{Total} = P_{Static} + P_{Dynamic} \quad (1)$$

$$P_{Dynamic} = P_{Short-Circuit} + P_{Switch} \quad (2)$$

$$P_{Total} = I_{Leakage} \cdot V_{dd} + I_{Static} \cdot V_{dd} + P_{Short-Circuit} + P_{Switch} \quad (3)$$

Where PTotal is total power dissipation. Static power PStatic is due to the leakage ILeakage and static IStatic currents. The NMOS and PMOS transistors used in a CMOS logic circuit commonly have non-zero reverse leakage and sub-threshold currents.

For the dynamic power dissipation, the first part $P_{Short-Circuit}$ is caused by direct supply-to-ground paths during the signal transitions. It can be controlled to a small portion of the total power dissipation by appropriate sizing of transistors, and by reducing the input rise and fall times to all the gates in the circuit. The second part P_{Switch} is due to charging and discharging of parasitic capacitances in the circuit. This is demonstrated by an inverter driving load capacitor C_L as shown in Figure 2. P_{Switch} can be calculated by:

$$P_{Switch} = V_{DD}{}^2 \cdot f \cdot C_L \quad (4)$$

Where V_{DD} is the supply voltage and f is the switch frequency, while C_L is load capacitance which includes internal parasitic capacitances of inverter. Since C_L is contributed to total power dissipation only when switching occurs in a circuit. In this paper, C_L is also called switch capacitance C_{SW}.

© 2013. The authors - Published by Atlantis Press

According to Chandrakasan et al.[13] in a "well-designed" circuit P_{Switch} accounts for over 90% of the total power dissipation. Thus, the total power dissipation for a CMOS circuit can be approximated by:

$$P_{Total} \approx P_{Switch} = C_L.V_{DD}^{\ 2}.f \qquad (5)$$

Since power is the energy consumed per second, energy E can be stated as:

$$E = P_{Total} \cdot t \approx P_{Switch} \cdot t = C_L \cdot V_{DD}^{\ 2} \cdot f \cdot t \qquad (6)$$

Where, $1/f$ is the time period for each switch. In a synchronized circuit, t is clock cycle and f is clock frequency. C_L is switch capacitance per cycle. At constant frequency power/energy can be used interchangeably. Further, if we assume that the supply voltage V_{DD} is fixed, reducing the power/energy dissipation is equivalent to reducing the switch capacitance.

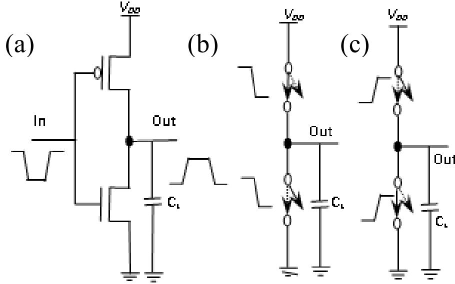

Fig.2 (a) CMOS inverter & eq. circuit for (b) charging and (c) discharging C_L

Charging and discharging occurs due to the PMOS and NMOS transistors respectively. So the dependence of the propagation delay on C_L suggests that getting C_L as small as possible is crucial for the realization of high performance circuits[10].

3. Parasitic Capacitances

A lot of work has been done for parasitic capacitances in CMOS technology. According to Shrivastava et al.[1] the most significant MOS parasitic circuit element is the gate-to-channel capacitances which vary both in magnitude and its division in the following three components:Gate to bulk capacitance $C_{GB,}$ Gate to source capacitance C_{GS}, Gate to drain capacitance C_{GD}

When the transistor is in the cutoff region in which the gate to source voltage is less than the threshold voltage i.e. $V_{GS} < V_T$, no channel exists and the total capacitance C_{GC} appears between gate and the bulk. In the resistive region with $V_{GS} > V_T$ and a small voltage, V_{DS}, is applied between drain and source, an inversion layer is formed which acts as a conductor between source and drain. Consequently C_{GB} =0 as the body electrode is shielded from the gate by the channel. In this region the capacitance is distributed between source and drain evenly.

The SPICE models for the MOS transistor have demonstrated the three built-in MOSFET models[10]. In our power macro modeling approach, we take the LEVEL 3 models and implement the following capacitance models using 0.12 μm technology.

$$C_{GS} = \frac{2}{3} C_i \left[1 - \left(\frac{V_{GS} - V_T - V_{DSAT}}{2(V_{GS} - V_T) - V_{DSAT}} \right)^2 \right] \qquad (7)$$

$$C_{GD} = \frac{2}{3} C_i \left[1 - \left(\frac{V_{GS} - V_T}{2(V_{GS} - V_T) - V_{DSAT}} \right)^2 \right] \qquad (8)$$

$$C_{GB} = 0 \qquad (9)$$

$$C_i = W.L \frac{\epsilon_o \epsilon_r}{T_{ox}}$$

Where W is the channel width, L is the channel length, T_{OX} is the oxide thickness, ϵ_o is the absolute permittivity and ϵ_r is the relative permittivity.

In macro model, we use different values of channel width for both PMOS and NMOS transistors. We considered L=0.12μm as constant for both transistors and T_{OX}=3nm, ϵ_o = 8.85x10^{-12} F/m, ϵ_r = 3.9 in case of SiO_2. In Equations (7-8), the value of V_{DSAT} is also a function of V_{GS} and V_T. So we used V_{DSAT} given below to get a more simplified form of the above capacitances:

$$V_{DSAT} = \frac{V_{GS} - V_T}{1 + \frac{V_{GS} - V_T}{E_C L}} \qquad (10)$$

Where E_C is the critical field at which electron velocity saturation occurs; it is around 1.5x10^6 V/m (or 1.5 V/μm) and the saturation velocity V_{SAT} is approximately 105 m/s. Critical field for holes is -1.95x10^6 V/m. By using above V_{DSAT} and using the approach given in the reference[10], the simplified form of Equations (7-8) can be represented as:

$$C_{GS} = \frac{\epsilon_o \epsilon_r}{T_{OX}} \left[1 - \left\{ \frac{V_{GS} - V_T}{E_c L + 2(V_{GS} - V_T)} \right\}^2 \right] \qquad (11)$$

$$C_{GD} = \frac{2}{3} W.L \frac{\epsilon_o \epsilon_r}{T_{OX}} \qquad (12)$$

In the previous model[10] the junction capacitances C_{DB} and C_{SB} were used in:

$$C_{SB} = W.L_{source} \frac{C_J}{\left(1 - \frac{V_{BD}}{PB}\right)^{MJ}} \qquad (13)$$

$$C_{DB} = W.L_{drain} \frac{C_J}{\left(1 - \frac{V_{BD}}{PB}\right)^{MJ}} \qquad (14)$$

Where W is the channel width L_{drain} and L_{source} are the lengths of drain and source regions respectively taken as 0.42 µm. C_J is around $3x10^{-4}$ F/m². PB is the built in potential of the junction and is around 0.8 V. MJ is called the grading coefficient and equals 1/2 for the abrupt junction and 1/3 for the linear or graded junction.

4. Models Parameters

For any particular model which is built for a particular technology range, there must be well defined parameters that hold well for every experimental iteration. In our technology we used the following parameters:
Threshold Voltage (V_{TO}) for NMOS and PMOS is 0.4 & -0.4V, Carrier Mobility (U_0) is 0.06 & 0.025 m²/V.s, Gate Oxide Thickness (T_{OX}) is 3 nm, surface potential at strong inversion (PHI) is 0.3 V. Bulk Threshold Parameter ($GAMMA$) is 0.4 $V^{0.5}$, Saturation Field Factor ($KAPPA$) is 0.01 V^{-1}. Maximum drift velocity (V_{MAX}) for NMOS and PMOS is 150 & 100 km/s, whereas MOS Channel Length (L) is 0.12 µm respectively.

In deep submicron technology, the integrated circuits (*IC*) with low voltage internal supply and high voltage input/output (*I/O*) interface are common. These models in Equations (11), (12), (13) and (14) are implemented at 1.2V whereas the *I/O* devices operate at standard voltages (2.5, 3.3 or 5V). Another reason for the lower internal voltage operation is thermal breakdown of the oxide layer. The gate oxide thickness is fixed at 3 nm in order to get increased switching performances. Due to the fact that the molecular distance of SiO_2 is 20 Å, i.e. 10 atoms; the oxide may be destroyed by a voltage higher than a maximum limit V_c, called the breakdown voltage.

Table.1. Capacitances for PMOS

W (µm)	C_{GS} (F)	C_{GD} (F)	C_{SB} (F)	C_{DB} (F)
0.4	2.98E-16	2.51E-16	5.04E-17	4.51E-17
0.5	3.73E-16	3.14E-16	6.30E-17	5.63E-17
0.6	4.47E-16	3.77E-16	7.56E-17	6.76E-17
0.7	5.22E-16	4.39E-16	8.82E-17	7.89E-17
0.8	5.96E-16	5.02E-16	1.01E-16	9.02E-17
0.9	6.71E-16	5.65E-16	1.13E-16	1.01E-16
1	7.45E-16	6.28E-16	1.26E-16	1.13E-16

A first order estimation is 0.1 V/Å[14] which is expressed as:

$$V_C = \frac{K}{T_{OX}} \tag{15}$$

Where K is Breakdown Coefficient and V_C is Critical Breakdown Voltage. Another parameter related to gate oxide thickness is the gate oxide capacitance i.e.

$$C_{OX} = \frac{\in_{OX}}{T_{OX}} \tag{16}$$

Increased gate oxide thickness T_{ox} causes gate oxide capacitance C_{ox} to decrease, therefore reducing drain current I_{DSAT}, and increasing the threshold voltage V_T[15].

5. Results and Discussion

We have calculated each component of lumped capacitance. The channel length is considered constant whereas the channel width for both MOS transistors is varied for a specific range. The values of parasitic capacitances are listed in Table.1 for different values of channel widths of PMOS transistors.

The design of an inverter starts with the implementation of one NMOS and one PMOS transistor, but using the same channel width for both transistors is not the best idea since the p-channel MOS transistor switches at half the current of n-channel MOS transistor. The origin of this mismatch can be seen in the general expression of current delivered by n-channel and P-channel MOS devices.

$$I_{DS}(NMOS) = \frac{\in_o \in_r \mu_n W_{NMOS}}{T_{OX} L_{NMOS}} \tag{17}$$

$$I_{DS}(PMOS) = \frac{\in_o \in_r \mu_p W_{PMOS}}{T_{OX} L_{PMOS}} \tag{18}$$

If $W_{NMOS} = W_{PMOS}$ and $L_{NMOS} = L_{PMOS}$, then the current delivered by NMOS and PMOS transistors will be proportional to electrons' and holes' mobility respectively i.e. $I_{DS} (NMOS)\ \alpha\ \mu_n$ and $I_{DS} (PMOS)\ \alpha\ \mu_p$.

Typical mobility values are $\mu_n = 0.068$ m²/V.s and $\mu_P = 0.025$ m²/V.s. Consequently the current delivered by the n-channel MOS device is more than twice that of the p-channel MOS device. Usually, the inverter is designed with balanced current to avoid significant switching discrepancies. Therefore, balanced current and switching performances are required. The best approach consists of enlarging the PMOS channel width as it is directly proportional to the current delivered by the PMOS transistor. In our design we take the channel width of PMOS transistor two times than that of NMOS. Its current becomes doubled and becomes comparable with the NMOS current[10]. The specific channel width values for NMOS and their corresponding parasitic capacitances are given in Table.2.

In the next step all the capacitances of MOS transistors are lumped to get a single load capacitance C_L including the wiring capacitance that is taken to be 0.12fF for this technology. This is a considerable simplification of the actual situation even in the case of a simple inverter. A typical CMOS inverter may drive successive gates and the total gate capacitance C_G is the gate capacitance of transistors being driven by the inverter.

Table.2. Capacitances for NMOS

W (µm)	C_{GS} (F)	C_{GD} (F)	C_{SB} (F)	C_{DB} (F)
0.2	1.47E-16	1.28E-16	2.52E-17	2.91E-17
0.25	1.84E-16	1.60E-16	3.15E-17	3.64E-17
0.3	2.20E-16	1.92E-16	3.78E-17	4.36E-17
0.35	2.57E-16	2.24E-16	4.41E-17	5.09E-17
0.4	2.94E-16	2.56E-16	5.04E-17	5.82E-17
0.45	3.30E-16	2.88E-16	5.67E-17	6.55E-17
0.5	3.67E-16	3.21E-16	6.3E-17	7.27E-17

To calculate accurately the total charge supplied to the load inverters we must perform an analysis with non-linear charge storage elements. However, the total gate capacitance C_G for the two inverters 1 & 2 can be estimated as:

$$C_G = C_{OX}[(W.L)_{p1} + (W.L)_{n1} + (W.L)_{p2} + (W.L)_{n2}] \tag{19}$$

In this designing the channel length is taken to be 0.12 µm as a technology parameter, whereas the channel widths of two successive inverters are similar to the first one. Therefore by using these dimensions Equation (19) can be simplified as:

$$C_G = \frac{2 \in_{0X} L}{T_{OX}} \left[W_p + W_n \right] \tag{20}$$

Where W_p and W_n are channel widths of PMOS and NMOS transistors respectively and $\in_o$ is the absolute permittivity. The simplified model of a CMOS circuit consisting of several gates can be viewed as one large capacitor that is charged and discharged between the power-supply rails. Therefore, the load capacitance, which is often specified as power–dissipation capacitance is used to approximate the dynamic power consumption because CMOS dissipates power only during switching (dynamic power). After computing the lumped capacitance C_L, the dynamic power dissipation is estimated and is given in Table.3. for 1 GHz frequency.

Table.3. Power Calculation for CMOS inverter

C_{TOT} (F) (*NMOS*)	C_{TOT} (F) (*PMOS*)	C_G (F)	C_L (F)	$P=V_{dd}^2 C_L f$ (w)
3.22E-16	6.23E-16	4.25E-16	1.64E-15	2.36E-06
4.02E-16	7.78E-16	5.31E-16	1.99E-15	2.86E-06
4.82E-16	9.34E-16	6.37E-16	2.34E-15	3.37E-06
5.63E-16	1.09E-15	7.43E-16	2.69E-15	3.87E-06
6.43E-16	1.25E-15	8.50E-16	3.04E-15	4.37E-06
7.23E-16	1.40E-15	9.56E-16	3.39E-15	4.88E-06
8.04E-16	1.56E-15	1.06E-15	3.74E-15	5.38E-06

The obtained values of our design clarify the linear relation between the power consumption and the lumped capacitance. A considerable change occurs in dynamic power for a small change in load capacitance. In large memory circuits it will become more significant and play an important role in VLSI circuit performance. Therefore, with a minute change of 0.05µm in channel width of NMOS and a change of 0.1µm in the channel width of PMOS transistors, a change of 0.504 µW occurs in power consumption. Figure 3 shows this change as a linear function.

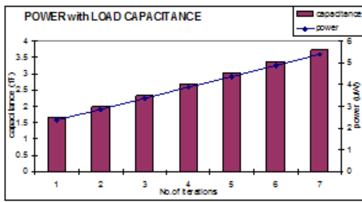

Fig. 3.Variation of power with CMOS width dependent load capacitance

For seven different values of channel width and their corresponding load capacitances, C_L taken from Table.3, variation of power consumption of our designed inverters with applied frequency is plotted in Figure 4.

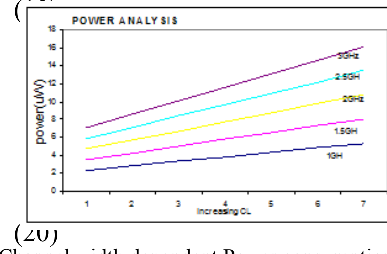

Fig. 4. Channel width dependent Power consumption variation

$$P_{Dynamic} = P_{Short\text{-}Circuit} + P_{Switch}$$

Lower gate delays in digital circuits are essential to ensure the faster rate of data processing and to get improved overall performance. The difference in propagation delays of logic elements is the major contributor to glitches in asynchronous circuits as a result of race conditions. Propagation delay increases with operating temperature, marginal supply voltage as well as an increased output load capacitance C_L. The latter is the largest contributor to the increase of propagation delay. If the output of a logic gate is connected to a long trace or is used to drive many other gates (high fan-out) the propagation delay increases substantially. The overall propagation delay of the inverter is defined as the average of two values i.e.

$$t_p = \frac{t_{pHL} + t_{pLH}}{2} = 0.69 C_L \left(\frac{R_{eqn} + R_{eqp}}{2} \right) \tag{21}$$

Where t_{pHL} & t_{pLH} are delay times from high-to-low and low-to-high propagation respectively and R_{eqn}, R_{eqp} are equivalent on-resistances of NMOS and PMOS transistors respectively. We simplified the on-resistance formula by using I_{DSAT} and after using some algebraic expressions to get the following simplified relation which is directly linked with the channel width of the device. By using the Multisim simulated results of transient response, we get the values of C_L by using Equation (21). For seven different values of channel widths for NMOS from 0.2 to 0.5µm with a change of 0.05µm and for PMOS from 0.4 to 1 µm with a change of 0.1µm, we performed simulations of our designed invertor.

$$R_{eq} = \frac{L V_{DD}}{6K' V_{DSAT}} \left(\frac{9 - 7\lambda V_{DD}}{2V_{DD} - 2V_T - V_{DSAT}} \right) \frac{1}{W} \tag{22}$$

The simulated and estimated values are given in Table.4 with percentage errors. The power analysis is also in good accordance for both cases, with an average percentage error of 12.8%, as given in Figure 5.

Table.4. Comparison of estimated (E) and simulated (S) valuesalong with % Error (Er)

E,C_L (fF)	E,t_p (ps)	S,t_p (ps)	S,C_L (fF)	*E,Power*	*S,Power*	Er,(%)
1.64	4.66	5.34	1.87	2.36	2.69	14
1.98	4.53	5.34	2.34	2.86	3.36	17.5
2.34	4.43	5.34	2.8	3.36	4.04	19.9
2.69	4.37	5.34	3.27	3.87	4.71	21.7
3.04	4.32	4.06	2.84	4.37	4.09	6.4
3.39	4.29	4.06	3.2	4.88	4.6	5.59
3.73	4.25	4.06	3.55	5.38	5.12	4.9

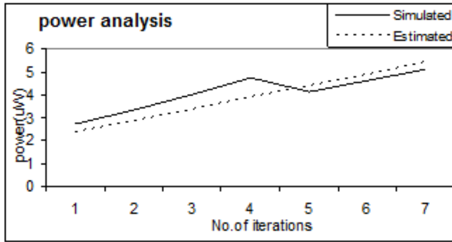

Fig. 5. Comparison between estimated and simulated values

Regression analysis is performed to fit the model's coefficient. Figure 5 illustrates the correlation between the simulated power estimation and the estimated power values. We measured the correlation coefficient that is around 91%. In both the simulated and estimated cases the variation of power with load capacitance remains same as shown in Figure 5. In the simulated case a 1GHz switching of the inverter induces a circuit power dissipation of 4.71 μW while in our work the induced power for 1 GHz switching is 3.87 μW with a maximum error of 21.7%. As we have to take $(W/L)_p$ twice as big as $(W/L)_n$ because $\mu_n = 2\mu_p$, therefore in this case $(W/L)_n$ = (0.35/0.12) and $(W/L)_p = (0.7/0.12)$. As the power consumption is linearly proportional to the clock frequency and increases gradually with the increase in load capacitance, similar behavior can be seen for the estimated case as in Figure 5. In simulations we got the simulated values of load capacitance C_L by using the propagation delay of inverter using Equation (21) and plotted the simulated values of power consumption which have the same linear relation with C_L except a sudden decrease in power after four iterations. The main culprit behind this decrease is the on-resistances R_{eqn} and R_{eqp} for NMOS and PMOS respectively and the propagation delay of inverter. Since it can be extracted from Equation (21) that a decrease in average on-resistance gives an increased value of C_L and a decrease in propagation delay t_p will correspond to a decreased value of C_L.

5. Conclusions

An analytical formulation for the dynamic power consumption of CMOS inverter with respect to transistor sizing has been presented which includes the capacitive dissipation. In this research work we built a mathematical model of parasitic capacitances in order to estimate the dynamic power dissipation of CMOS inverter. A brief description of all internal parasitic capacitances related to channel and junctions is provided with their precise models for the 0.12 μm technology. These models are simplified and each capacitance is calculated for seven different values of channel widths for both NMOS and PMOS transistors. By using these capacitances, the lumped capacitance is computed and this capacitance is then used to estimate the dynamic power dissipation of CMOS inverter. Furthermore, simulations are performed for the same channel dimensions and the transient response of inverter is used to calculate the power consumption. The estimated and simulated results are compared and analysis of power consumption with increasing frequency is also done.

Acknowledgments

This work was supported Higher Education Commission (HEC), Pakistan under Grant given to COE in Solid State Physics.

References

[1] V. M. Shrivastava, A. Priyank, S. P. Singh, G. Singh, Proc. of International Conference on Recent Trends in Info. Telecom. Comp. (2010) March 12-13; Kerala, India.
[2] G. Consentino and G. Ardita, Proc. of IEEE Int. Symp. Ind. Elect. (2009) July 5-8; Seoul, Korea.
[3] J. D. Carey, J. Nanosci. Nanotechnol. 9, 11 (2009).
[4] V. D. Kunz, T. Uchino, C. H. Groot, P. Ashburn, D. C. Donaghy, S. Hall, Y. Wang and P. L. F. Hemment, IEEE Trans. Electron Dev., 50, 1487 (2003).
[5] F. Pregaldiny, C. Lallement and D. Mathiot, Solid State Electron., 46, 2191 (2002).
[6] H. Kamchouchi and A. Zaky, IEEE Trans. Electron Devices, ED-30, 183 (1983).
[7] R. Shrivastava and K. Fitzpatrick, IEEE Trans. Electron Devices, ED-29, 1870 (1982).
[8] K. Suzuki, IEEE Trans. Electron Devices, 46, 1895 (1999).
[9] N. R. Mohaputra, M. P. Desai, S. G. Narendra and V. R. Rao, IEEE Trans. Electron Devices, 50, 959 (2003).
[10] E. Sicard, S. D. Bendhia, Basics of CMOS Cell Design, Mc-Graw Hill, USA (2007).
[11] J. M. Rabaey and M. Pedram, Low Power Design Methodologies, Kluwer Academic Publisher Inc. USA (1996).
[12] J. Frenkil, Proc. Inter. Sym. Low -Pow. Elect. Des. **(1997)** August 18-20, California, USA.
[13] A. P. Chandrakasan, S. Sheng and R. W. Brodersen, IEEE J. Soli-Sta. Cir. 27, 473 (1992).
[14] A. Z. H. Wang, On chip ESO protection for integrated circuits - An IC Design perspective, Kluwer Academic publishers, USA (2002).
[15] X. Chen and D. Velencies, I.R.E.E, 1, 123 (2006).