Performance evaluation of Near-Threshold Ultradeep Submicron Digital CMOS Circuits using Approximate Mathematical Drain Current Model

Karan Singh¹, Shruti Kalra²

¹,² Department of Electronics and Communication, Jaypee Institute Of Information Technology, Noida, India
e-mail: vermakaransingh@gmail.com

Abstract— This paper examines CMOS inverter performance near threshold voltage using a tailored drain current model for submicron nodes up to 22nm. Validated via HSPICE simulations, the model incorporates factors like inverse narrow width effect, Miller capacitance, and layout dependencies. Demonstrating continuity and precision across a wide operational spectrum, results indicate deviations within 1-5%, affirming the model’s robustness. These findings underscore its utility for designers grappling with the intricacies of near-threshold voltage operation, offering a reliable framework to optimize CMOS inverter performance in ultra-deep submicron technologies.

Index Terms— Inverse Narrow Width Effect, Layout-Dependent Effects, Near Threshold Computing, Miller Capacitance

I. INTRODUCTION

Operating circuits at a supply voltage near the transistor threshold voltage offers numerous advantages:

(1) **Lower Power Consumption:** As the supply voltage nears the transistor threshold voltage, the transistor has reduced current levels. Consequently, power consumption decreases as the current flowing through the transistor reduces. This sub-threshold operation is advantageous in low-power applications, as it allows for efficient power management while still maintaining necessary functionality in electronic devices and circuits [1].

(2) **Reduced Leakage Current:** Lower supply voltages lead to decreased leakage current through the transistor, enhancing circuit efficiency and prolonging battery life in portable devices [2].

(3) **Increased Sensitivity:** Operating circuits near the transistor threshold voltage offers advantages in increased sensitivity, particularly in low-power or high-gain applications, due to several reasons: Low-Power Applications: In low-power scenarios, minimizing supply voltage is critical to reduce power consumption. Operating near the threshold voltage allows for lower supply voltages, leading to significant power savings. This is advantageous for battery-powered devices and energy-efficient systems where extending battery life is paramount. High-Gain Applications: Near the threshold voltage, transistors exhibit higher transconductance (gm), making them more responsive to small input signals. This increased transconductance enhances the gain of the circuit, making it suitable for applications requiring precise signal amplification, such as sensor interfaces and communication receivers [3].

(4) **Improved Noise Performance:** While operating circuits near the transistor threshold voltage can improve certain aspects of noise performance, such as flicker noise, it can lead to a poorer signal-to-noise ratio (SNR) and noise margin due to reduced headroom and increased susceptibility to variations in transistor characteristics. For instance, the reduced supply voltage can exacerbate the impact of random dopant fluctuations, leading to increased variability in transistor threshold voltage and degraded noise margins [3] (refer table I).

(5) **Smaller Transistor Size:** Although reducing the supply voltage can enable the use of smaller transistors in scaled CMOS processes, the relationship between supply voltage and transistor size is complex and context-dependent. In scaled technologies, lowering the supply voltage is often accompanied by transistor scaling to maintain performance and mitigate short-channel effects [3]. However, these benefits come with drawbacks such as slower processing speeds and increased process variations [5]-[7]. Consequently, design iterations are often necessary to meet desired criteria. Many sources of variation, fortunately, can be systematically modeled at the device or circuit level [8]. Notably, layout-dependent effects (LDE) [9] and the inverse narrow width effect (INWE) [10] are intrinsic to circuits and cannot be avoided. LDEs, primarily caused by strain engineering, result in layout parameters influencing
device performance. Process-induced stress also significantly affects performance, particularly in the near-threshold voltage (NTV) regime [11]. Similarly, INWE makes the threshold voltage of narrow width devices dependent on the device’s width [12]. Neglecting these factors in circuit design can lead to inferior performance and increased post-layout adjustments.

To address these challenges, performance modeling at the logic gate level, accounting for these variations, is beneficial [13]-[19]. For instance, [20] proposes a delay model for subthreshold circuits, while [21] suggests using ON-current ($I_{ON}$) to simplify delay estimates, albeit with reduced accuracy. Integrating complex current equations [22] with existing models remains challenging, prolonging delay estimation efforts. Table I lists the literature survey of current drain current models for MOS devices.

Logic library cell lookup tables offer an alternative for delay estimation based on load and slew. These tables, derived from SPICE simulations on netlists using layout parameters, consider LDEs. However, characterizing logic cell libraries for process-voltage-temperature fluctuations is resource-intensive due to the large number of variables involved.

In optimizing digital circuit performance, customized design methods are popular. Considering LDEs during the initial design phase is crucial for reducing design time. Alternatively, the effective current approach, requiring finite DC current values for accurate delay estimation [33], provides precise results. The effective driving current ($I_{eff}$) serves as a measure of digital circuit performance [33]. Models for $I_{eff}$ for various CMOS logic gates are presented in [34]-[35], but conventional models produce errors in predicting delay for circuits in the NTV regime. Consequently, developing $I_{eff}$ models specific to near-threshold conditions is imperative [36]. Traditionally, a logic gate’s logical effort (LE) has been considered independent of size and layout factors [37]. However, LDEs necessitate considering $I_{eff}$ per unit width as a function of layout factors in sub-90-nm technologies [37]. New LE model-based design approaches address transistor scaling in the subthreshold range [38]-[39]. [40] proposes energy optimization methods considering INWE but ignores LDEs. Additionally, [41] discusses optimal P/N ratios without considering LDEs, while [42] presents reverse short-channel effect techniques which may not be effective for deep nanoscale nodes due to channel stress dependence on channel length.

Notably, neglecting LDEs during critical pathway design allows significant room for improving circuit performance [5]-[37]. While LDEs have been addressed in the superthreshold regime [5]-[37], renewed attention to near-threshold circuits is warranted. Simulation can aid in understanding these impacts, contributing to reduced design time and improved performance.

Key contributions of this work include:

- Performance evaluation of CMOS inverters (static, dynamic, and energy consumption) near threshold voltage using a mathematical approximate drain current model tailored for ultra-deep submicron technology nodes up to 22nm.
- Validation of the suggested model through industry-standard HSPICE simulations to establish its validity.
- Consideration of the inverse narrow width effect, Miller capacitance, and layout-dependent effects, which become significant near threshold voltage.
- Development of a continuous and accurate model applicable across a wide operating range, ranging from several times the thermal voltage to approximately double the threshold voltage in modern technologies.
- Demonstration that the obtained results fall within an acceptable error range of 1-5

### II. Mathematical Modelling of Near Threshold Drain Current $I_{ds}$

**Proposed Analytical Model:**

1. **Tailored Drain Current Model:** The analytical model intricately captures the behavior of the drain current in CMOS inverters operating near threshold voltage. It considers various factors such as channel length modulation, carrier mobility, and threshold volt-
Importance of the Analytical Model:

1. Robustness and Reliability: The analytical model exhibits robustness and reliability, with simulation results indicating deviations within a narrow range of 1-5%. This affirms the model’s accuracy and underscores its utility as a reliable framework for designers working with near-threshold voltage operation [47].

2. Optimization in Submicron Technologies: The model’s ability to accurately capture CMOS inverter performance in ultra-deep submicron technologies is of paramount importance. It provides designers with a powerful tool to optimize circuit performance, minimize power consumption, and enhance overall reliability in advanced semiconductor technologies [47].

3. Insightful Design Exploration: By offering a detailed understanding of device behavior and performance characteristics, the analytical model facilitates insightful design exploration. Designers can leverage this knowledge to identify critical design parameters, optimize circuit layouts, and achieve desired performance targets with greater efficiency [47].

A MOS transistor operates in weak inversion if the gate-to-source voltage \( V_{gs} \) is less than the threshold voltage \( V_{th} \). This definition assumes that for the gate voltage larger than the threshold voltage, the transistor starts to operate in moderate inversion. With further increase of the gate-source voltage, the transistor enters strong inversion. Some of the models available utilize the expression of strong inversion with some fitting constants to model the drain current of moderate and weak inversion. The same argument may be given about the saturation region; i.e., in strong saturation, drain current becomes independent of drain-

---

Table II. Literature Survey of Drain Current Models for MOS Devices

<table>
<thead>
<tr>
<th>Reference (Year)</th>
<th>Model Description</th>
<th>Key Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>[24],(2021)</td>
<td>Compact model considering channel length modulation and hot carrier effects.</td>
<td>Improved scalability for nanoscale MOS devices, validated through TCAD simulations.</td>
</tr>
<tr>
<td>[30],(2021)</td>
<td>Reliability-aware model considering degradation effects in harsh operating conditions.</td>
<td>Robustness in extreme temperature and voltage conditions, ensuring device longevity.</td>
</tr>
<tr>
<td>[31],(2016)</td>
<td>3D IC model considering vertical integration effects in stacked MOSFET structures.</td>
<td>Consideration of interlayer coupling effects, aiding in design optimization for 3D integrated circuits.</td>
</tr>
<tr>
<td>[32],(2020)</td>
<td>Heterojunction model for heterostructure MOS devices, accounting for bandgap engineering effects.</td>
<td>Improved understanding of charge transport across heterojunction interfaces, leading to enhanced device performance.</td>
</tr>
</tbody>
</table>
to-source voltage (Note: idealized MOSFET is considered without channel length modulation). Weak saturation is defined by quadratic dependence of drain-to-source voltage (keeping fixed gate-to-source voltage). In between these two regions occurs a moderate saturation region. Again, the transition equations for weak to moderate and moderate to strong saturation are absent for short channel devices. The model presented in [43] is valid for long channel transistors up to 1 μm technology node. In this paper, an extension of [43] with simple mathematical expansions to model the drain current for short channel MOSFET operating in all the inversion (weak, moderate, and strong) and saturation (weak, moderate, and strong) conditions is presented (Valid up to 22nm technology node). Table III below shows the different drain current symbols used for various inversion/saturation conditions.

### Table III.: Symbols used for different drain current conditions

<table>
<thead>
<tr>
<th>Inversion/Saturation</th>
<th>Deep</th>
<th>Strong</th>
<th>Moderate</th>
<th>Weak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weak</td>
<td>I_{DWD}</td>
<td>I_{DSM}</td>
<td>I_{DSM}</td>
<td>I_{DSW}</td>
</tr>
<tr>
<td>Moderate</td>
<td>I_{DMD}</td>
<td>I_{DMS}</td>
<td>I_{DMM}</td>
<td></td>
</tr>
<tr>
<td>Strong</td>
<td>I_{DSM}</td>
<td>I_{DSS}</td>
<td>I_{DSW}</td>
<td></td>
</tr>
</tbody>
</table>

According to the EKV model, the drain current $I_{ds}$ may be represented as [48]

$$I_{ds} = I_S I_C$$  \hspace{1cm} \text{(1)}

In this case, $I_S$ represents specific current which is equal to $I_S = 2\mu \text{eff} C_{ox} \phi_t^2 W/L$; $n$ represents the sub-threshold slope; $C_{ox}$ is the gate oxide capacitance; $\mu \text{eff}$ signifies the effective mobility; $\phi_t$ represents the thermal voltage; $W$ denotes the channel width; $L$ is the effective channel length; $I_C$ is the symbol for inversion coefficient. The inversion coefficient describes the degree of transistor inversion, and covers both the sub- $V_{th}(IC < 1)$ and above $V_{th}(IC > 1)$ regions of the transistor.

Now [48],

$$I_{ds} = \ln^2 \left[ 1 + \exp \left( \frac{V_p - V_s}{2} \right) \right] - \ln^2 \left[ 1 + \exp \left( \frac{V_p - V_d}{2} \right) \right]$$  \hspace{1cm} \text{(2)}

where $I_C = I_{ds}/I_S = i_{ds}$ is the normalized value of drain current to make it unitless; $V_p = V_{gs} - V_{th}$ is the normalized value of pinch-off voltage and $V_d = V_d/\phi_t$ and $V_s = V_s/\phi_t$ are the normalized value of drain and source voltage respectively. Referencing via the source rather than the bulk and eliminating the normalization results in:

$$I_{ds} = I_S \ln^2 \left[ 1 + \exp \left( \frac{V_{gs} - V_{th}}{2n\phi_t} \right) \right]$$

$$-I_S \ln^2 \left[ 1 + \exp \left( \frac{V_{gs} - V_{th} - V_{ds}}{2\phi_t} \right) \right]$$  \hspace{1cm} \text{(3)}

The model described above is appropriate for devices with large channel lengths. For short channel devices, the ‘2’ component in the previous equation corresponding to velocity saturation is substituted by the factor $\alpha$, and the following drain current model is presented:

$$I_{ds} = I_S \ln^\alpha \left[ 1 + \exp \left( \frac{V_{gs} - V_{th}}{2n\phi_t} \right) \right] - I_S \ln^\alpha \left[ 1 + \exp \left( \frac{V_{gs} - V_{th} - V_{ds}}{2\phi_t} \right) \right]$$  \hspace{1cm} \text{(4)}

Where $\alpha$ is the velocity saturation index. Its value may be determined numerically using interpolation, as done by EKV [48], or analytically by equating it to the drain current of deep saturation derived from BSIM4 and then solving for $\alpha$. When the drain current reaches a critical value, it becomes independent of $V_{ds}$ (neglecting the effect of channel length modulation). Thus, by excluding the terms depending on $V_{ds}$ in Equation 4, the drain current in the deep saturation area may be represented as follows:

$$I_{DD} = I_S \ln^\alpha \left[ 1 + \exp \left( \alpha \frac{V_{gs} - V_{th}}{2n\phi_t} \right) \right]$$  \hspace{1cm} \text{(5)}

By expanding Equation 5 using the series expansion provided in [33], we get:

$$\ln^\alpha (x) = \left( -\sum_{n=1}^{\infty} \frac{(-1)^n(x-1)^n}{n} \right)$$  \hspace{1cm} \text{for} \ |x-1| < 1  \hspace{1cm} \text{(6)}

$$\ln^\alpha (x) = \left( \ln (x-1) - \sum_{n=1}^{\infty} \frac{(-1)^n}{n(x-1)^n} \right)$$  \hspace{1cm} \text{for} \ |x-1| > 1  \hspace{1cm} \text{(7)}

If $x = 1 + \exp \left( \frac{V_{gs} - V_{th}}{2n\phi_t} \right)$ in Equation 5, and it is assumed to have weak inversion ($V_{gs} < V_{th}$), then $|x-1| < 1$. It can be shown that the series in Equation 6 is converging for $0 \leq x < 2$ and is approximated by the first term alone:

$$\ln^\alpha (x) = (x-1)^\alpha$$  \hspace{1cm} \text{(8)}

As a result, the drain current associated with weak inversion and deep saturation may be represented as:

$$I_{DWD} = I_S \exp \left( \frac{\alpha (V_{gs} - V_{th})}{2n\phi_t} \right)$$  \hspace{1cm} \text{(9)}

However, as $x$ approaches 2, the series in Equation 6 begins to diverge from the original series. In order to address this, the following modified form of the first two terms of the series expansion is used:
\[ \ln^a(x) = (x-1)^a (1 - \alpha_1 (x-1)) \quad (10) \]

where \( \alpha_1 = 1 - (\ln 2)^a \).

As a result, the drain current of weak inversion for the aforementioned condition is:

\[ I_{DWD} = I_s \exp \left( \frac{a(V_{gs} - V_{th})}{2n\phi_t} \right) \left( 1 - \alpha_1 \exp \left( \frac{V_{gs} - V_{th}}{2n\phi_t} \right) \right) \quad (11) \]

Inversion ends when the condition \( V_{gs} = V_{th} \) is met. When \( V_{gs} \) exceeds \( V_{th} \), the transistor enters a state of moderate inversion. For this reason, an approximation of Equation 7 is used:

\[ \ln^a(x) = \left( \ln(x-1) + \frac{1}{x-1} \right)^a \quad \text{for } x \geq 2 \quad (12) \]

If \( x = 1 + \exp \left( \frac{V_{gs} - V_{th}}{2n\phi_t} \right) \), then it is provided as \( I_{DMD} \) in Equation 12 for the drain current of deep saturation and moderate inversion:

\[ I_{DMD} = I_s \left( \frac{V_{gs} - V_{th}}{2n\phi_t} + \exp \left( -\frac{V_{gs} - V_{th}}{2n\phi_t} \right) \right)^a \quad (13) \]

For \( x > 2 \), the series shown above is convergent. However, if this approximation is not suitable for the value of \( x \) near 2, the modified version of the first two terms of the series provided by Equation 7, where \( x \) is a positive integer, is used:

\[ \ln^a(x) = \left( \ln(x-1) + \frac{1}{x-1} - \frac{\alpha_2}{(x-1)^2} \right)^a \quad (14) \]

Here, \( \alpha_2 = 1 - \ln 2 \).

The transistor moves closer to the strong inversion zone of operation when \( V_{gs} \) is increased even further. When the first term in Equation 13 dominates the second term, the transistor is said to be in strong inversion. The transistor is considered to be in strong inversion when the second term is reduced to around 5% of the first term. Therefore,

\[ \frac{V_{gs} - V_{th}}{2n\phi_t} = 20 \exp \left( -\frac{V_{gs} - V_{th}}{2n\phi_t} \right) \quad (15) \]

The transition voltage (which marks the beginning of strong inversion) may be calculated as:

\[ V_{gs} - V_{th} = 2.2(2n\phi_t) \quad (16) \]

As a result, the current of deep saturation and strong inversion, \( I_{DSD} \), is:

\[ I_{DSD} = I_s \left( \frac{V_{gs} - V_{th}}{2n\phi_t} \right)^a \quad (17) \]

If the series approximation is performed with the help of Equation 13, the drain current of deep saturation and moderate inversion is:

\[ I_{DMD} = I_s \left( \frac{V_{gs} - V_{th}}{2n\phi_t} + \exp \left( -\frac{V_{gs} - V_{th}}{2n\phi_t} \right) \right)^a \quad (18) \]

Equation 18 describes the transconductance properties of MOSFETs operating in the moderate inversion zone.

The drain current results simulated for deep saturation conditions of n-channel MOSFET from BSIM4 for different inversion levels are equated using our model equations to derive the value of alpha for moderate and strong inversion, as shown in Table IV for different technological nodes.

Figure 1 depicts the verification of the change in drain current of NMOSFET and PMOSFET with regard to \( V_{gs} \) at three technology nodes operating in moderate inversion regions using BSIM4.

The drain current of a MOSFET is composed of two components: drift and diffusion. The temperature coefficients of these two components are different. When the inversion is weak, the diffusion component takes over, which has a positive temperature coefficient, and the threshold voltage decreases as the temperature rises. When there is substantial inversion, the drift component dominates and has a negative temperature coefficient. However, in moderate inversion, both components contribute to the drain current, and the temperature coefficient will be determined by the bigger of the two components. Figure 2 depicts the change of the Zero Temperature Coefficient (ZTC) point voltage (ZTC is considered widely for biasing temperature-compensated circuits) and the threshold voltage at 22nm technology node. The voltage at the zero temperature coefficient point has been computed for NMOSFET and PMOSFET using a mathematical \( a \)-power law-based model for NMOSFET and PMOSFET. It has been shown that the ZTC point for scaled transistors is located in the moderate inversion region. Consequently, as compared to the strong inversion approximation previously employed in the literature, using a continuous drain current model valid in all three operating zones (weak, moderate, and strong) offers a more accurate estimate of the ZTC voltage.

Digital Object Identifier: 10.29292/jics.v19i2.692
## III. CMOS Inverter-DC Characteristics

The CMOS inverter is the heart of any digital design since it converts voltage to current. Once the functioning of the system is well understood, the design of more complicated logic becomes easier. CMOS inverter performance may be used to predict the electrical and thermal behaviour of other digital circuits to an extremely high degree of accuracy. The examination of the CMOS inverter may be expanded to include the performance of fundamental digital building blocks like NAND, NOR, and XOR, which in turn are used to construct modules such as multipliers and processors [50], among others. The static behaviour of the CMOS inverter is investigated in this part at various technology nodes, supply voltages, and temperatures using a suggested $\alpha$-power MOSFET model, and the findings are validated using the industry-standard BSIM simulator.

As illustrated in figure 3, the voltage transfer or DC characteristics may be achieved by superimposing the current characteristics of NMOSFET and PMOSFET acquired by the suggested $\alpha$-power-based model on the voltage transfer or DC characteristics. The calculated and simulated Voltage Transfer Characteristics (VTC) of a CMOS inverter for transistors operating in moderate inversion are shown in figure 4.

## IV. Modeling the Near Threshold Delay

An equation for the delay of a CMOS inverter is found by using the drain presented in section II. of this document. To begin, consider the case of discharging the output capacitance using NMOS. In this case, the propagation $t_p$ of a gate may be defined as follows [14]:

$$ t_p = \frac{C_L V_{dd}}{I_{ON}} \tag{19} $$

Here, $C_L$ is the load capacitance. $I_{ON}$ is the ON current of NMOS. The value of $I_{ON}$ is obtained from Equation 18 by putting $V_{gs} = V_{dd}$.

$$ t_p = \frac{C_L V_{dd}}{2n\mu_{eff} C_{ox} W_i L_i \left(\frac{V_{gs} - V_{th}}{n \phi_i}\right)^2 I_C} \tag{20} $$

Here,

$$ I_C = \left(\frac{V_{gs} - V_{th}}{2n \phi_i} + \exp\left(-\frac{V_{gs} - V_{th}}{2n \phi_i}\right) - \alpha_2 \exp\left(-\frac{V_{gs} - V_{th}}{n \phi_i}\right)^\alpha\right) \tag{21} $$

Capacitance under the load term $C_L$ refers to the sum of the inherent capacitance of the driving stages and the load capacitance of fan-out gates. As a result, $C_L$ is proportional to the following [15]:

$$ C_L \propto C_{ox} L(\xi W_i + W_{i+1}) \tag{22} $$

In this Equation, $\xi$ represents the relationship between the parasitic capacitance of a driving phase and the input gate capacitance of a fanout. The annotations for the driver and loading stage are represented by the numbers $i$ and $i+1$. As a result, $W$ in Equation 20 is
rewritten as \( W_i \). Equivalently, substituting the value \( C_L \) into Equation 20,

\[
t_p = \frac{L^2}{2n\mu_{eff}q_i^2} \frac{V_{dd}}{TC} \frac{\xi_j W_i + W_{i+1}}{W_i}
\]

The route delay may be estimated as follows for an \( N \)-stage path:

\[
t_{p,i} = \frac{L^2}{2n\mu_{eff}q_i^2} \frac{V_{dd}}{TC} \sum_{i=1}^{N} \frac{\xi_j W_i + W_{i+1}}{W_i}
\]

\[
t_{p,i} = k_{pd} \frac{V_{dd}}{TC} t_{p,i}(W)
\]

In this Equation, \( k_{pd} \) is the technological dependent factor, and \( t_{p,i}(W) \) is the transistor size dependent factor. Overall, time for \( N \) stages may be defined in terms of the mean propagation delay of \( t_p \) of a single stage as well as the logic depth \( Ld \), which is a convenient representation [15].

\[
t_{p,i} = t_p Ld
\]

Figure 5 depicts the fluctuation of time with supply voltage, which ranges from weak to moderate to strong inversion. The findings of the analytical calculations are validated using BSIM4 [52] for the 32nm technology node, which is a simulation tool. From the figure, it can be observed that the shortest delay is achieved at higher supply voltages.

V. MILLER EFFECT

Fig. 1 shows a CMOS inverter’s Miller capacitance \( C_M \). It is principally created by the gate-to-drain capacitances of the NMOS and PMOS devices that contribute to this parasitic capacitance. For each transition in the output voltage waveform, the Miller capacitance is drained and subsequently recharged in the opposite direction of the waveform’s transition. Accordingly, the effective parasitic capacitance is \( 2C_M \) when seen from the standpoint of energy. However, from the standpoint of delay, the Miller capacitance created by

Digital Object Identifier: 10.29292/jics.v19i2.692
the shifts in the operational zones of the two transis-
tors results in considerable nonlinearities [53]. When a
step input waveform is assumed, the discharging cur-
rent via the load capacitance \( C_L \) is as follows:

\[
    i_{C_L} = C_L \frac{dV_{out}}{dt} = C_M \left( \frac{dV_{in}}{dt} - \frac{dV_{out}}{dt} \right) - i_n
\]

The results of standard Laplace domain analysis are as follows:

\[
    C_L \left[ sV_{out}(s) - V_{out}(0-) \right] + C_M \left[ sV_{out}(s) - sV_{in}(s) - V_{CM}(0-) \right] = -i_n(s)
\]

\[
    s(C_L + C_M)V_{out}(s) - (C_L + 2C_M)V_{dd} = -i_n(s)
\]

\[
    V_{out}(s) = \frac{C_L + 2C_M V_{dd}}{C_L + C_M} - \frac{i_n(s)}{C_L + C_M} \frac{1}{s}
\]

Traditionally, with deep sub-micrometer technologies,
the saturation voltage \( V_{dsat} \) (drain-to-source) is less
than half the supply voltage \( V_{dd} \) (supply voltage). Con-
sequently, before the output voltage hits \( V_{dd}/2 \), the
NMOS transistor only conducts current in the satura-
tion region, and \( I_{ON} = i_n \) may be represented as a con-
stant average saturation current \( I_{dn} \) [53].

\[
    V_{out}(t) = \frac{C_L + 2C_M}{C_L + C_M} V_{dd} - \frac{i_{dn}}{C_L + C_M} t
\]

At some point in time, \( V_{out}(t) \) approaches half of the
supply voltage given as:

\[
    t_{phl} = \frac{C_L + 3C_M}{2} \frac{V_{dd}}{I_{dn}}
\]

Similarly,

\[
    t_{plh} = \frac{C_L + 3C_M}{2} \frac{V_{dd}}{I_{dp}}
\]

A. **Miller Effect in Conjunction with a Ramp Input Volt-

gage**

In actual integrated circuits, the waveform of the in-
put voltage is different from that of a step function.
Nonzero input transition time has two principal effects: a slow transition of the CMOS inverter from one zone of operation to another, and a reduction in Miller effect with linearly rising input transition time. While just one transistor is responsible for conduction of current to and from $C_L$, both transistors actively engage in the current conduction process and switch through a number of distinct working zones [53].

$$i_{C_L} = C_L \frac{dV_{out}}{dt} = C_M \left( \frac{dV_{in}}{dt} - \frac{dV_{out}}{dt} \right) - i_n + i_p \quad (33)$$

Different transition relations between $V_{in}$ and $V_{out}$ may be used to solve Eq. 33, which can be solved by solving numerous differential equations. [54] proposed a formula for the CMOS inverter fall delay time as:

$$t_{phl} = \frac{1 + 2n}{6} t + \frac{\Delta Q}{I_{dn}} + \frac{C_L V_{dd}}{I_{dn}} \left( \frac{C_M}{C_M + C_L} + f(V_{dsat}) \right) \quad (34)$$

where $V_{dd}$ is the power supply voltage, $n = V_{thn}/V_{dd}$, $I_{dn}$ is the NMOS saturation current, $f(V_{dsat})$ is a function of the NMOS drain-to-source saturation voltage. The charge provided to the output node by the PMOS transistor represents the influence of the PMOS transistor during the discharge transition:

$$\Delta Q = (C_M + C_L)(\Delta V_o - V_{dd}) \quad (35)$$

In this case, $V_o$ is the voltage overshoot. As a result of the increased charge $\Delta Q$, the output delay time is $\Delta Q/I_{dn}$. In Eq. 34, the three factors contribute to the delay imbalance when $C_L, C_M,$ and $f$ are altered, which is due to the differences in NMOS and PMOS device characteristics. The contribution of the second transistor is determined by the combined capacitances $C_L + C_M$, as well as the transistor transconductance, which is determined by the contributed charge $\Delta Q$. Transistor operating modes fluctuate owing to variations in the drain-to-source saturation voltage $V_{dsat}$, which are independent of the voltage difference $V_{thn} - V_{thp}$. According to Eqs. 26 and 33, the Miller effect diminishes as the rate of change of the input-to-output voltage decreases. Furthermore, overshoot and undershoot voltages are reduced. At the same time, the short circuit power dissipation grows in a manner that is essentially linear with the input transition time. Growing the input transition time has contradictory impacts on the output delay time: (1) it reduces the output delay by decreasing the Miller effect; (2) it increases the output delay due to greater short-circuit contention currents. Figure 6 shows the transient response of CMOS inverter with FO4 load. It can be seen from the plot that before the output voltage hits $V_{dd}/2$, the NMOS transistor only conducts current in the saturation region for the transistors operating in moderate inversion. Figure 7 shows the modification of delay for a minimal size CMOS inverter with a FO4 load and supply voltage varying from moderate to strong inversion. It can be seen from the figure that the minimum delay at a lower technology node is greater than the minimum delay at a higher technology node because the load capacitance at the lower technology node is greater than that at the higher technology node.

VI. Modeling the Near Threshold Energy

The total energy $\epsilon_{total}$ used for any logic circuit is the total of the switching energy, i.e., energy related to the charging / discharging of the load capacitance $C_L$, $\epsilon_{switching}$, and the leakage element of energy consumption, $\epsilon_{leakage}$, in one clock period. For a single gate, it may be represented as follows [50]:

Digital Object Identifier: 10.29292/jics.v19i2.692
The energy expended during the course of $N$ steps, or route energy, may be stated as:

$$
\epsilon_{\text{total},i} = C_{\text{ox}} L \sum_{j=1}^{N} (\xi_{i} W_{j(i)} W_{i+1}) V_{dd}^{2} a + (k_{p} d V_{dd}^{2} t_{p}(W)) (2 n \mu_{eff} C_{\text{ox}} \sum_{j=1}^{N} W_{i} \phi_{i}^{2}) \exp \left( \frac{\alpha V_{dd} - V_{th}}{n \phi_{t}} \right) V_{dd}
$$

(38)

The energy used over time as the supply voltage fluctuates from weak to strong inversion. For the 32nm technology node, the analytical results are confirmed using BSIM4. As the plot indicates, the least energy use occurs in low supply voltages. Energy consumption in modern CMOS circuits is mostly caused by charging and discharging load capacitance and may be lowered quadratically by lowering the supply voltage $V_{dd}$, as seen in the figure 8. Voltage scaling has emerged as one of the most convincing strategies for reducing power usage. While it is obvious that CMOS circuits continue to function at low supply voltages below the threshold voltage, referred to as subthreshold voltages, the circuit’s lack of durability and performance restrict the designer’s ability to run the circuit in this area. Due to the quadratic scaling in the upper threshold regime, energy is very sensitive to supply voltage $V_{dd}$. Thus, voltage scaling up to the near threshold area results in about tenfold energy savings at the expense of approximately tenfold performance deterioration, as seen in figure 9. Nevertheless, when voltage is lowered below the threshold voltage $V_{th}$, the dependency of energy on supply voltage $V_{dd}$ becomes more complicated. In the subthreshold region, circuit delay increases exponentially with supply voltage $V_{dd}$, causing leakage energy to increase in a near exponential fashion, i.e. leakage current, delay, and supply voltage. This rise in leakage energy prevents any further energy decrease and establishes a minimum energy point.
Figure 8: Variation of energy consumed over one cycle for minimum size CMOS inverter driving another inverter as a function of supply voltage. The analytical results are verified using BSIM4 at 32nm technology node.

Figure 9: Energy and delay in different supply voltage operating regions for analyzing energy/delay trade-off for minimum size CMOS inverter driving another inverter at 32nm technology node.

VII. Influence Of Layout Dependent Effects on Drain Current and Delay

Device current does not rise linearly with transistor width or number of fingers in strain-engineered technologies, which results in systematic variability [37]. When it comes to the near-threshold regime [7], this variability has a greater influence and should be taken into consideration at the prelayout stage in order to get an optimum design with the least amount of postlayout adjusting. In this part, we demonstrate that our model is accurate in predicting the delay in the presence of a large number of fingers.

The effective current $I_{eff} = I_{ON}$ from $V_{in} = 0.95V_{dd}$ to $V_{in} = 0.5V_{dd}$ variation of an inverter cell while NF is changed from 1 to 10 is seen in the figure 10.

The fluctuation in $I_{eff}$ for 2 distinct width values is seen in the graph. According to our observations, when NF is altered, $I_{eff}$ fluctuates greatly; this is in contrast to the traditional expectation that $I_{eff}/NF$ stays constant [which is indicated as conventional in figure 10. Because of this, our models may be employed in circuit optimization while taking the layout-dependent systematic deviations into consideration. Therefore, the first step in using our methodology is to choose effective width. Following that, the driving capacity of such a logic gate is raised just by increasing the number of fingers. A constant width value is required in order to remove the impacts of INWE, and so this restriction ensures that the design process is kept to a bare minimum. Although this limitation is not difficult to meet, it may be loosened in our technique owing to the fact that the increase in logical effort owing to $I_{eff}$ and INWE could be easily included into our formulas. The logical effort can be calculated for $i_{th}$ stage as:

$$g_i = \frac{C_{in,i}I_{eff,r}}{C_{in,r}I_{eff,i}}$$ (40)

Digital Object Identifier: 10.29292/jics.v19i2.692
Here, $C_{in,i}$ and $C_{in,r}$ are the input capacitance of the $i$th stage and the reference capacitance. $I_{eff,r}$ and $I_{eff,i}$ are the effective current of reference stage and $i$th stage.

Table V shows an evaluation of the proposed methodology in comparison to previous studies. It has been shown that the buffer chain with $C_{out}/C_{in} = 2500$ constructed using the suggested methodology yielded 10.3 percent and 25 percent improvements in latency and energy per operations respectively.

VIII. Conclusion

In this paper, we have thoroughly investigated the effective driving current of a CMOS inverter operating in the near-threshold domain, specifically focusing on Moderate Inversion conditions. Through the utilization of simple DC simulations, we have accurately determined the value of the effective current, providing valuable insights into device behavior in this operating regime. Additionally, we have systematically incorporated the influence of Miller capacitance into our model, resulting in enhanced accuracy compared to previous studies. Furthermore, we have introduced a novel modified Logical Effort (LE) approach, leveraging the proposed effective current model to optimize circuit performance. Notably, our methodology accounts for variations such as Inverse Narrow Width Effect (INWE) and Layout Dependent Effects (LDEs) during the prelayout design phase, a consideration overlooked in prior literature. Experimental results validate the efficacy of our technique, with a buffer chain constructed using our methodology demonstrating significant improvements. Specifically, the buffer chain with a $C_{out}/C_{in}$ ratio of 2500 achieved a remarkable 10.3% reduction in latency and a 25% decrease in energy per operation compared to previous designs. These findings underscore the importance of accurately modeling near-threshold CMOS circuits and highlight the potential benefits of considering layout-dependent variations early in the design process. Overall, this research contributes to advancing the state-of-the-art in CMOS inverter design and optimization, offering valuable insights for future development in low-power, high-performance computing applications.

REFERENCES


<table>
<thead>
<tr>
<th>Circuit under test</th>
<th>Cout/Cin</th>
<th>Cin(FF)</th>
<th>$W$ ($\mu m$)</th>
<th>Energy per operation (fJ)</th>
<th>Delay(ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chain of Inverters</td>
<td>580</td>
<td>0.4</td>
<td>0.13</td>
<td>15</td>
<td>16.21</td>
</tr>
<tr>
<td></td>
<td>2500</td>
<td>0.4</td>
<td>0.13</td>
<td>91</td>
<td>16.54</td>
</tr>
</tbody>
</table>


