Effective Drive Current in CMOS Inverters for Sub-45nm Technologies

Jenny Hu*, Jae-Eun Park†, Greg Freeman† , Richard Wachnik† , H.-S. Philip Wong*
* Department of Electrical Engineering, Stanford University, CA, USA jennyhu@stanford.edu
† IBM, East Fishkill, NY, USA

ABSTRACT

We propose a new model for the effective drive current (I_{eff}) of CMOS inverters, where the maximum FET current obtained during inverter switching (I_{PEAK}) is a key parameter. I_{eff} is commonly defined as the average between I_H and I_L, where $I_H=I_{ds}(V_{gs}=V_{DD}, V_{ds}=0.5V_{DD})$ and $I_L = I_{ds}(V_{gs}=0.5V_{DD}, V_{ds}=V_{DD})$. In the past, this I_{eff} definition has been accurate in modeling the inverter delay. However, we find that as devices are scaled further into the nanoscale regime, the maximum transient current can deviate severely from I_H, in which case, another metric should be used. The deviation of I_{PEAK} from I_H is found to increase as delay decreases or as device overdrive voltage increases. We define $I_{eff} = (I_{PEAK}+I_M+I_L)/ 3$, where $I_M = I_{ds}(V_{gs}=0.75V_{DD}, V_{ds}=0.75V_{DD})$. We evaluate our model against others by comparing the analytical and HSPICE extracted I_{eff} ratios across devices of varying threshold voltages, V_{TH}. Our model is shown to better capture changes in V_{TH}/V_{DD}, which are important since V_{DD} and V_{TH} will be key parameters for optimizing device performances for target applications (low power or high performance) in sub-45nm technologies.

Keywords: CMOS, inverter, delay, performance

1. INTRODUCTION

As we continue to scale CMOS devices further into the nanoscale regime, performance metrics valid for older technologies need to be re-evaluated for their suitable application with the current technology. FET performance is typically measured through the CMOS inverter delay $\tau = CV/I$, where C is the load capacitance, V is the power supply V_{DD}, and I is the saturation on-current $I_{ON}= I_{ds}(V_{gs}= V_{ds}= V_{DD})$. In recent years, the concept of an effective drive current has been proposed as a better representation for calculating the inverter delay, due to the fact that I_{ON} is never reached during switching [1 – 4]. However, as devices are scaled, current-voltage characteristics continue to deviate from the predicted behavior of ideally scaled MOSFET due to minimal scaling of V_{TH} and V_{DD}, and non-idealities such as parasitic series resistances, velocity saturation, and DIBL. To more accurately assess the performance of sub-45nm CMOS, a more representative effective drive current is needed.

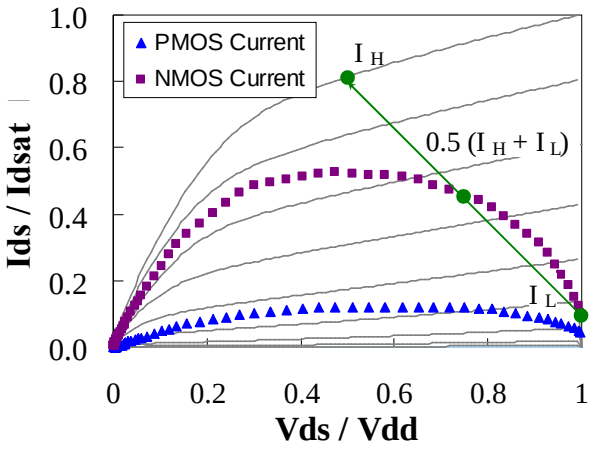

Figure 1. Pull-down switching trajectory of low V_{TH} device, with bias points plotted in equal time intervals. Arrow indicates assumed trajectory in the two-point I_{eff}.

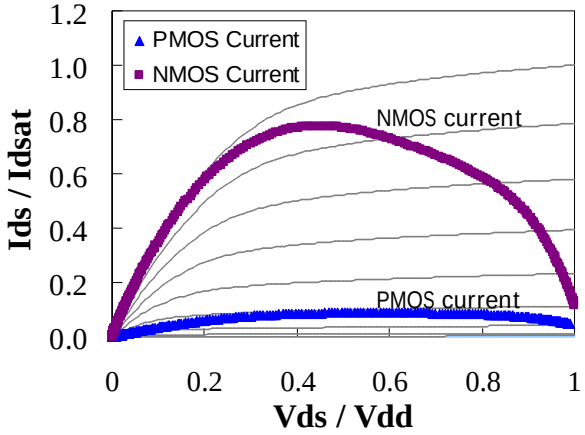

Figure 2. Pull-down switching trajectory of 250 nm technology, where I_{PEAK} occurs closer to I_H.

2. TRANSIENT CURRENT TRAJECTORY

In determining a suitable method to evaluate the inverter delay, it is important to first examine the transient current trajectory to better understand approximations used in I_{eff}. Fig. 1 illustrates the transient inverter current trajectory of a pull-down transition plotted over the DC current characteristics of an NMOS, with current normalized to I_{ON}. It is noted that the peak current (I_{PEAK}) obtained during switching falls far below I_H, and even

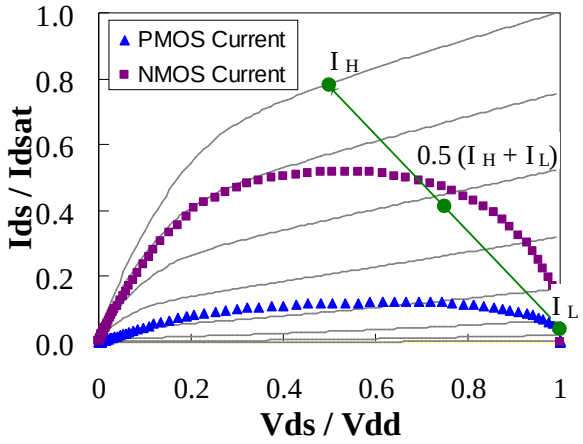

Figure 3. Pull-down switching trajectory of a higher V_{TH} device, showing a larger V_{PEAK} than in Figure 1, but I_{PEAK} is still far from I_H.

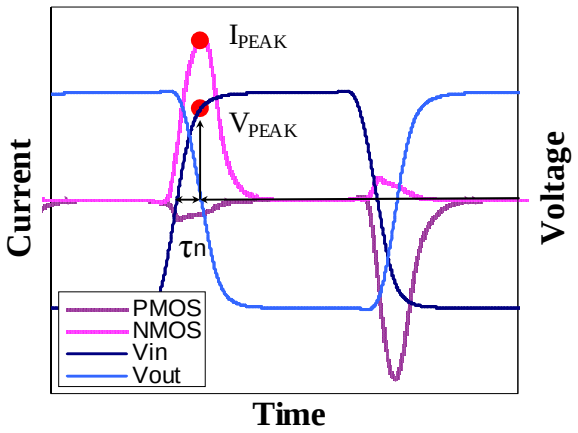

Figure 4. Transient response of both a pull-up and pull-down transition, showing $V_{PEAK} < V_{dd}$ and the dependence upon delay.

further from I_{ON}. This deviates greatly from previous technologies, where I_{PEAK} approached I_H as shown in Fig.2 for a 250nm technology. With the inverter delay defined as the time between $V_{IN}=V_{DD}/2$ to $V_{OUT}=V_{DD}/2$ during switching, the pull-down delay can be viewed as the integration of CV/I , where I ideally is the trajectory current. From Na et al. [2]:

$$\tau_{PD} = \int_{V_{GS}=V_{DD}/2}^{V_{DS}=V_{DD}/2} \frac{C_{LOAD}}{I_{DS}(V_{DS},V_{GS})}\, dV_{DS} \qquad (1)$$

From the equation it is apparent that the more closely the I_{eff} equation follows the current trajectory path, the more accurately it will reflect the device performance. The current trajectory has a parabolic shape, so a minimum of three points would be needed to more precisely capture the trends in current.

In addition to examining the current trajectory across technology nodes, we examine the changes in the trajectory as we scale the threshold voltage V_{TH} of devices in a given technology. Designing devices for a variety of applications, whether low power or high performance, results in devices with a wide range of V_{TH}, thus making it important to study the effect of scaling V_{TH}/V_{DD} on I_{eff}. The current trajectory of a higher V_{TH} device is plotted in Fig. 3. When comparing Fig. 1 and 3, it is significant to note the differences in the current trajectory paths taken by the devices with two separate V_{TH}. As the V_{TH} of a device decreases, the peak current in the trajectory occurs at a lower V_{gs}, further away from I_H. This behavior can be explained by Fig. 4, which illustrates the transient response of a ring oscillator. For a given V_{DD}, as V_{TH} decreases, the current overdrive is greater, thus decreasing the delay. With a shorter delay, the output falls from V_{DD} to V_{DD} /2 faster than the input can rise to V_{DD}, causing I_{PEAK} to occur at a V_{gs} much lower than V_{DD}. Therefore, a decrease in delay also decreases V_{PEAK}, defined as the gate voltage at which I_{PEAK} occurs. This can be equivalently interpreted as the current trajectory following a lower V_{gs} curve.

This change in current trajectories suggests that evaluating I_H at the same bias regardless of the device overdrive is not sufficient. Also, the simplification of the parabolic current trajectory as a linear path from I_L to I_H [1] is no longer a good approximation if the peak current is far below I_H. In previous technologies, larger τ allowed $I_{PEAK} \approx I_H$ and $V_{PEAK} \approx V_{DD}$, but this is changing as devices and power supplies are scaled. We propose an I_{eff} model that includes more than two points of current and takes into account of I_{PEAK}.

3. EFFECTIVE DRIVE CURRENT MODEL

Na et al. [1] defined the effective drive current by a two-point model, $I_{eff} = (I_H + I_L)/2$, where $I_L = I_{ds}(V_{gs}=0.5\ V_{DD}, V_{ds}= V_{DD})$ and $I_H = I_{ds}(V_{gs}= V_{DD}, V_{ds}=0.5\ V_{DD})$. This effectively approximates the current trajectory as a linear path between I_H and I_L. Deng et al. [2] introduced a three-point model of $I_{eff} = (I_H + I_M + I_L)/3$, where $I_M = I_{ds}(V_{gs}=0.75\ V_{DD}, V_{ds}=0.75\ V_{DD})$. A four-point model was also proposed, where current through the device that should be off is accounted for. The four-point model was tested, however, we did not find significant benefit over the three-point model in a complexity and accuracy tradeoff.

We propose another three-point model of $I_{eff} = (I_{PEAK} + I_M + I_L)/3$, where I_{PEAK} is used to replaced I_H. Including I_{PEAK} will allow Ieff to include the effects of changes in switching trajectories between devices.

To analyze these I_{eff} models, the analytical I_{eff} is compared to the simulated I_{eff} extracted from HSPICE simulations of a nine-stage ring oscillator using IBM's 45nm bulk and SOI models for several devices designed for a range of V_{TH}. These models take into account critical doping changes between the various V_{TH} devices.

NSTI-Nanotech 2008, www.nsti.org, ISBN 978-1-4200-8505-1 Vol. 3

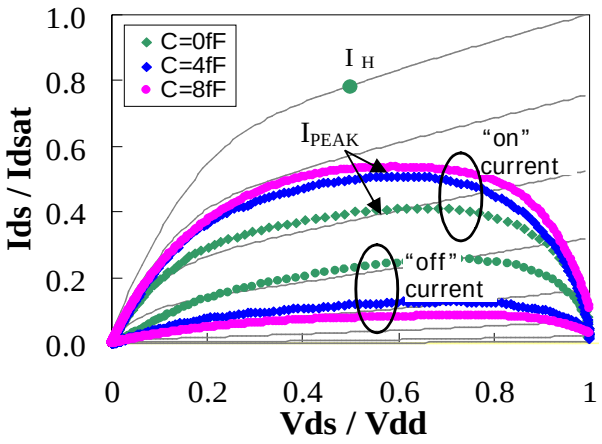

Figure 5. Dependence of I_{PEAK} on the load capacitance. A smaller load capacitance has a smaller I_{PEAK} and greater short circuit current in the device that is turning off.

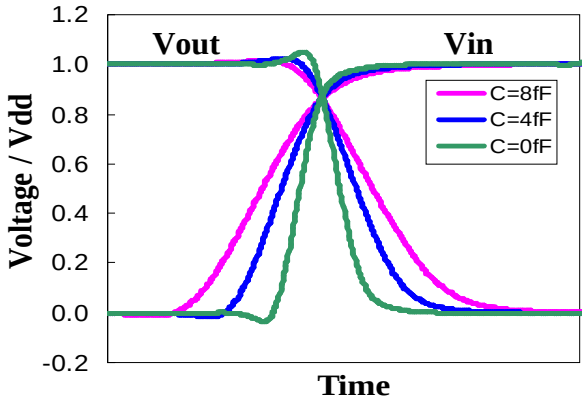

Figure 6. A smaller load capacitance has a larger short circuit current in the device that is shutting off is due to greater overlap of voltage range that both Vgs and Vds are significant. A large load capacitance is used to accentuate the effect for illustration purposes.

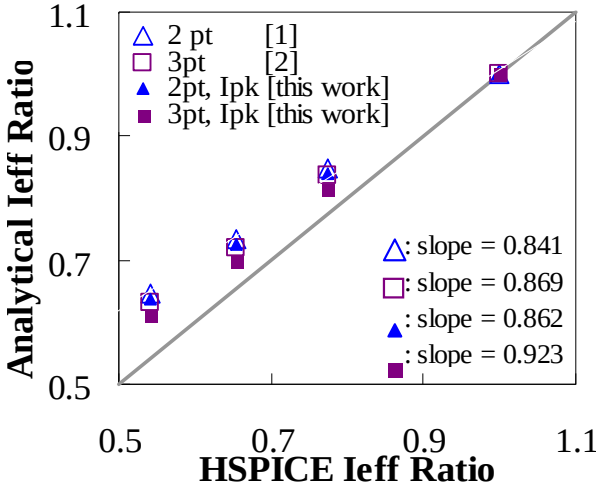

Figure 7. Evaluation of the relation between analytical and HSPICE I_{eff} ratios for several definitions of I_{eff}. Ratios are taken relative to the I_{eff} of the device with the lowest V_{TH}. For each definition of I_{eff}, the slope of the best fit line is compared, with a slope of 1 being ideal. The maximum slope occurs in the three-point definition using I_{PEAK}.

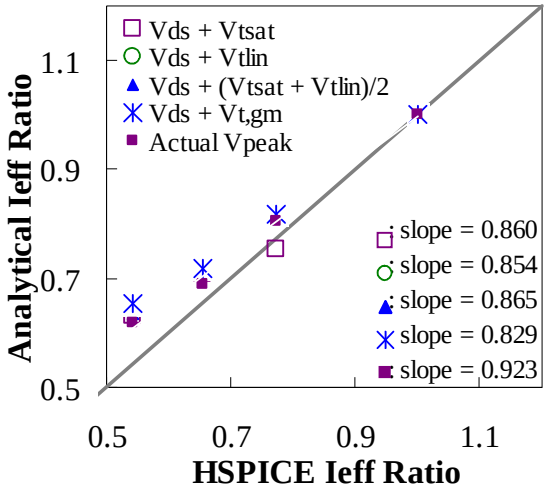

Figure 8. Evaluation of the correlation between analytical and HSPICE I_{eff} ratios for several definitions of V_{PEAK}. Maximum slope is for $V_{PEAK} = 0.5V_{dd} + (V_{TSAT}+V_{TLIN})/2$

4. PEAK CURRENT

From the ring oscillator simulations, we found the dependence I_{PEAK} has on the load capacitance. As illustrated in Fig. 5, in the pull-down case, as the load capacitance decreases, I_{PEAK} of the NMOS also decreases. This can be explained by Fig. 4, due to a decrease in the delay with reduced load capacitance. More notable is the increase in the short circuit current of the device that is turning off, or off-current, which is the PMOS during a pull-down transition. The origin of this off-current is further elaborated in Fig. 6. Consider the waveform of Vin and Vout. We note that for smaller load capacitances, the slope of the voltages is sharper, resulting in greater overlap of voltages in the range where both the NMOS and PMOS are on, thus a larger off current. Although we illustrated the dependence on the load capacitance, the underlying reason is a shorter delay, so this same dependence would appear in any situation where the delay is sufficiently short. In cases where this current approaches I_{PEAK} through smaller load capacitance or more advanced technologies, it will be critical to employ a four-point model to account for this off current. However, in the 45nm technology evaluated in this work, for the reasonable load capacitances considered, a three-point model is found to be adequate.

5. RESULTS

Our model is shown to better capture changes in V_{TH}/ V_{DD}. Fig. 7 summarizes the evaluation of several I_{eff} definitions across four devices of varying V_{TH}. The ratios of both the analytical and simulated I_{eff} are taken with respect to the device with the corresponding largest I_{eff}, or lowest V_{TH}. For each definition of I_{eff}, the slope of the best fit line for the four V_{TH} devices is compared, with a slope of 1 being ideal. It is noted that regardless of I_H or I_{PEAK},

when the two-point definition is replaced with the three-point, there is an improvement in the correlation. I_{eff} = $(I_{PEAK} + I_M + I_L)/3$ results in the best correlation between the analytical and HSPICE extracted I_{eff} , with a slope of 0.923 compared to the 0.841 slope of the two-point I_{eff} definition.

In each case where I_H is replaced with I_{PEAK}, there is a noted improvement in correlation, with an overall 7.2% improvement from the two-point I_{eff}.

However, evaluation of I_{PEAK} becomes useful only when the value can be predicted from DC current characteristics. We propose using a simple estimate of I_{PEAK}, disregarding the effects of the load capacitance mentioned in Fig. 6. From Hamoui et al. [4], the maximum transient current occurs roughly when the NFET leaves saturation and enters triode operation, at $V_{gs} = V_{PEAK} = m*V_{ds} + V_{TH}$, where m is the body factor. Since inverter delay is defined as the time between $V_{gs}= V_{DD}/2$ to $V_{ds}= V_{DD}/2$, it is logical to evaluate I_{PEAK} using $V_{ds}= V_{DD}/2$. Threshold voltage can be defined in many ways, so we evaluate several V_{TH} definitions to determine the most suitable for V_{PEAK}: 1.V_{TSAT}, 2.V_{TLIN}, 3.$(V_{TSAT}+V_{TLIN})/2$, and 4.$V_{TH,PEAK\ gm.}$ Fig. 8 illustrates $(V_{TSAT}+V_{TLIN})/2$ has the most promise, although there is a lower correlation than directly extracting V_{PEAK}, there is still a higher correlation compared to the traditional two-point I_{eff}.

6. CONCLUSION

In summary, we propose a new metric for I_{eff}, using $(I_{PEAK} + I_M + I_L) / 3$. We propose using the peak current attained during an inverter switching trajectory, as an alternative to $I_{H,}$ to better correlate with ring oscillator simulation extracted I_{eff}. Use of I_{PEAK} becomes imperative in cases where the peak current trajectory deviates significantly from I_H, which becomes increasingly important as delay decreases or overdrive voltage increases, often as we scale beyond the 45nm technology node.

ACKNOWLEDGEMENTS

This work is supported in part by the Focus Center Research Program (FCRP) (C2S2) and NSF (ECS-0501096). J. Hu is additionally supported by the Stanford Graduate Fellowship and the National Defense Science and Engineering Graduate (NDSEG) Fellowship. This work was also performed in part at IBM.

REFERENCES

[1] K. K. Ng, C. S. Rafferty, and H. Cong, "Effective On-current of MOSFETS for Large-signal Speed Consideration," IEDM Tech. Dig., pp. 693 – 696, 2001.

[2] M. H. Na, E. J. Nowak, W. Haensch, and J. Cai, "The Effective Drive Current in CMOS Inverters," IEDM Tech. Dig., pp.121–124, 2002.

[3] J. Deng and H.-S. P. Wong, "Metrics for Performance Benchmarking of Nanoscale Si and Carbon Nanotube FETs Including Device Nonidealities," IEEE Trans. Elec. Dev., Vol.53, No.6, pp.1317–1322, June 2006

[4] K. Arnim, C. Pacha, K. Hofmann, T. Schulz, K. Schrufer, and J. Berthold, "An Effective Switching Current Methodology to Predict the Performance of Complex Digital Circuits," IEDM Tech. Dig., pp.483 – 486, 2007.

[5] A.Hamoui and N. Rumin, "An Analytical Model for current, Delay, and Power Analysis of Submicron CMOS Logic Circuits," IEEE Trans. on Circuits and Systems II, Vol.47, No.10, pp.999–1006, October 2000.