

An Investigation of the Impact of Technology Scaling on Power Wasted as Short-Circuit Current in Low Voltage Static CMOS Circuits

Amitava Chatterjee, Mahalingam Nandakumar, and Ih-Chin Chen

Semiconductor Process and Device Center, Texas Instruments
P.O. Box 655012, M/S 461, Dallas, TX 75265

Abstract

In this paper the effects of technology scaling on the fraction of active power P_a wasted as short-circuit power P_s are studied through SPICE simulations. The accuracy of SPICE is verified against experimental data. SPICE simulations show that lowering V_T below 0.1V can increase P_s/P_a significantly beyond what is expected from increased subthreshold leakage. P_s/P_a is typically higher at higher V_{cc} but to first order P_s/P_a is determined by signal slew rates and V_T . It is shown that the input slew rate is constrained by P_s/P_a at low V_T and by performance at higher V_T . We show that P_s increases with increasing gate sheet resistance. A simple analytical model for this effect is verified against the experimental data and used to determine the gate sheet requirements to maintain $P_s/P_a < 10\%$ for sub-0.25 μm technologies.

1. Introduction

In low power CMOS it is common to reduce the power supply voltage V_{cc} to reduce power consumption and simultaneously reduce V_T to enhance performance. The penalty of increased leakage power, $V_{cc} I_{leak}$, with reducing V_T is well recognized, but limited discussion is available on the effect of scaling on short-circuit power dissipation. Here we show that short-circuit power P_s increases significantly as V_T is scaled below 0.1V. Thus, at low V_T , the slew-rate at the input of a signal-driver is constrained by P_s rather than by performance. Another scaling trend is to shrink gate lengths L_g . When L_g is scaled $< 0.25\mu\text{m}$ a rapid increase in gate sheet resistance ρ_{sh} can limit the expected performance gain [1][2]. Here we show that, in addition to increasing gate delay, a high ρ_{sh} increases P_s .

In section 2 we present our results on the effects of scaling V_T and V_{cc} on P_s/P_a (the fraction of active power wasted because of short-current). Experimental measurements of drive currents (I_{dn} and I_{dp}), inverter delays (t_d) and active power are used to verify SPICE and SPICE simulations are used to study the effects of scaling on P_s . In section 3 we present our results on the effect of ρ_{sh} on P_s . A simple model for P_s is proposed and verified against our measurements. Using this model we show the constraints on ρ_{sh} for a roadmap for sub-0.25 μm CMOS. Section 4 summarizes the main conclusions.

2. Short-circuit Power in Inverters for Scaled V_{cc} and V_T

The discussion of short-circuit power is based upon the following terminology. The total power dissipation P_{total} has two components: (a) leakage power P_{leak} , and (b) active power P_a defined as the power dissipated only during logic transitions. A fraction of P_a is wasted as a direct short-circuit current (P_s). We obtain P_a and P_s from SPICE simulations using a short-channel MOSFET model [3] with nominal model parameters extracted from measured I-V and C-V characteristics. P_a for an inverter is obtained by integrating its power supply current over τ the duration of logic transitions in a period $T=1/f$. A fraction of this power is wasted if there is a nonzero direct current through the MOSFET device which is supposed to be switched off. Thus we compute P_s by integrating this calculated quasistatic direct current over the same duration, τ . Since the leakage current I_{leak} is present during a transition there is a leakage component of P_s given by $V_{cc} I_{leak} \tau f$. This implies that P_{leak} the power consumption due to the quiescent current (leakage during the time when the MOSFET voltages are static) is given by $V_{cc} I_{leak} (1-\tau/T)f$. However, it is common practice to define $P_{leak} = V_{cc} I_{leak}$. Thus, consistent with this practice, we have computed P_s and P_a excluding the leakage component in addition to computing P_s and P_a by simple integration of the MOSFET

currents. In the results presented here we have clearly identified the method of calculation in each context.

The accuracy of SPICE is verified by comparison with experiment before using simulations to model the effects of V_{cc} and V_T on P_s . Devices are fabricated using processes described in [4]. *Fig.1* shows a comparison of measured and simulated nMOS and pMOS drive currents. Good agreement over a reasonably wide range of L_g , V_T , and V_{cc} demonstrates the accuracy of SPICE dc simulations. *Fig.2* compares the corresponding inverter delays. Again, good agreement is seen indicating accuracy of the transient simulations. *Fig.3* shows a reasonably good agreement between simulated and measured P_a for the same inverter chains. These SPICE simulations show that P_a (excluding leakage) is in excess of that estimated from CV^2f . *Fig.4* shows that this excess power correlates well with the corresponding P_s computed from SPICE. The impact of P_s on the design of static CMOS circuits is discussed in [5] where it is pointed out that a lower input slew rate S_i and a higher output slew rate S_o increases P_s . For $S_i \simeq S_o$ (such as in an inverter chain) $P_s/P_a \simeq 10\%$. In this paper we focus on the effects of technology scaling.

The effects of scaling V_T at a given V_{cc} on P_s/P_a of inverter chains (representing CMOS logic) and single inverters with a fixed input slew-rate (representing signal-drivers) are shown in *Figs.5-7*. For both type of circuits there is a significant increase in P_s/P_a (to about 30% of P_a at $V_T=0$) as V_T is reduced below 0.1V. Furthermore, by computing the leakage portion of P_s we show that only about half this increase can be attributed to a higher subthreshold leakage. By comparing a 0.25 μ m technology with an earlier 0.4 μ m technology we demonstrate that these results are quite general. *Fig.5* compares P_s/P_a vs V_T for inverter chains of fanout=1 and 4. As expected, P_s/P_a is seen to be insensitive to fanout. At higher fanout the effect of a slower input transition on P_s/P_a is compensated for by a higher P_a and a longer output transition. *Fig.6* shows that, for inverter chains, P_s/P_a is insensitive to V_{cc} . A higher short current and faster output slew-rate when V_{cc} is increased from 1V to 1.8V is compensated by a higher P_a and a faster input slew-rate.

The behavior of a single inverter (with an input signal of fixed slew-rate of 3.3V/ns) for varying load C_L and V_{cc} differs from that of the inverter chains. As we shall see, the differences are because in inverter chains the input and output slew rates are approximately equal. *Fig.7* shows that, as in [5], increasing C_L from 100fF to 400fF reduces P_s/P_a from 37% to 15% at $V_T=0$ V. Likewise, reducing V_{cc} reduces P_s/P_a as the output transition time increases because of lower drive currents. *Fig.8* shows that even if V_T is reduced along with V_{cc} ($V_T/V_{cc}=0.2$), lowering V_{cc} greatly reduces

P_s/P_a for the signal-line driver from 50% at $V_{cc}=2.5$ V to about 8% at $V_{cc}=1$ V. The dramatic reduction in P_s/P_a with reducing V_{cc} shown in *Fig.8* illustrates the importance of optimizing the slew rates for power dissipation ($S_o/S_i=3.2$ at $V_{cc}=2.5$ V and 1.6 at 1V).

A more meaningful comparison to study the effect of scaling V_{cc} for single inverters is to adjust the device sizes (W_n and W_p) to maintain the same drive current. *Fig.9* shows that for $I_{dn}=I_{dp}=5.6$ mA, P_s/P_a reduces from 50% at 2.5V to 24% at 1V. Although the drive currents are the same S_o/S_i reduces from 3.2 at 2.5V to 2.8 at 1V due to the increased self loading at $V_{cc}=1$ V. Thus, the difference in P_s/P_a is further reduced if C_L is adjusted in addition to W_n and W_p to maintain $S_o=10.7$ V/ns. Likewise, the difference in P_s/P_a between $V_{cc}=2.5$ V and 1V is greatly reduced ($P_s/P_a=15\%$ at 2.5V and 8% at 1V) if the input slew rate at 2.5V is increased to keep S_o/S_i constant at 1.6. Thus P_s/P_a shows an increase with increasing V_{cc} but to first order P_s/P_a depends on S_o , S_i and V_T rather than on V_{cc} .

Both P_s and P_{leak} are wasted components of the total power dissipation $P_{total}=P_a+P_s+P_{leak}$. *Fig.10* shows the relative importance of $P_{leak}=V_{cc}I_{leak}$ and P_s (excluding leakage) as a function of frequency for inverter chains with V_T varied from 0V to 0.15V. At high V_T (small P_{leak}) P_s is the main wasted power and P_s/P_{total} is independent of frequency. As V_T is reduced both P_s and P_{leak} increases. At low frequencies P_{leak} dominates while at higher frequencies P_s and P_{leak} are comparable. The maximum value for P_s/P_{total} is P_s/P_a . Thus if $S_o/S_i \simeq 1$ (as for the inverter chains in *Fig.10*) then P_s/P_{total} can be kept small.

One would like to determine the constraints on S_i and S_o as V_T is scaled to increase transistor drives. *Fig.11* shows P_s/P_a as functions of the input transition time for different C_L and V_T . As shown in [5] P_s increases approximately linearly with $1/S_i$. The intersection of these curves with $P_s/P_a=10\%$ (indicated by the dotted line in *Fig.11*) is used to determine the constraints on S_i and S_o . Contours of $P_s/P_a=10\%$, such as determined from *Fig.11*, are shown in *Fig.12* for $V_T=0$ V, 0.1V and 0.2V. The corresponding drive currents are 1.45mA, 1.25mA, and 1.05mA, respectively. At low V_T the input slew rate S_i has to be maintained high simply to limit the short-circuit power to 10% of the active power. At higher V_T a lower S_i can be tolerated to satisfy $P_s/P_a < 10\%$. In designs with relatively high V_T the input slew rate is constrained by performance. If S_i is excessively reduced it can degrade performance because S_o reduces with reducing S_i . *Fig.13* shows the lower bounds of S_i for $P_s/P_a=10\%$ and $S_o=75\%$ of $1\text{mA}/(C_L V_{cc})$ as functions of V_T for various values of C_L . The hatched regions indicate the space where both constraints on P_s and S_o are satisfied. Observe

that at low V_T , S_i needs to be designed to control P_s whereas at high V_T S_i is determined by performance requirements.

3. Effect of Gate Sheet Resistance on Short-circuit Power

Maintaining a low ρ_{sh} to achieve high performance becomes increasingly important [1] as conventional TiSi_2 processes encounter significant difficulties for L_g below $0.25 \mu\text{m}$. In addition to the degradation in performance, a high ρ_{sh} causes enhanced P_s . As shown in the inset of *Fig. 14* the resistive gate is analogous to a distributed RC network. SPICE simulations of the short-currents during a pull-down transition corresponding to $\rho_{sh} = 5, 100, 850 \Omega/\text{sq}$ are shown in *Fig. 14*. This illustrates the higher short-current resulting from the delay in propagating a voltage transition down a resistive gate which delays switching the entire width of the transistor.

Experimental evidence of higher P_s with higher ρ_{sh} is shown in *Fig. 15*. Power estimated from CV^2f is compared with measured power of inverter chains at different L_g and V_{cc} from three different salicide processes. We observe increasing power with reducing L_g and increasing ρ_{sh} . The processes corresponding to $\rho_{sh} = 100$ and $850 \Omega/\text{sq}$ at $L_g = 0.26 \mu\text{m}$ are conventional salicide processes known to yield very high poly sheet resistance at fine linewidths. A rapid thermal silicide process is used to obtain $\rho_{sh} = 5 \Omega/\text{sq}$. Similar to the model for degradation in inverter delay we have derived a simple model for P_s assuming that the propagation delay is proportional to the RC product of the gate:

$$P_s = \alpha V_{cc} f \left[I_{dn} t_n \left\{ 1 - \left(\frac{r_t}{(r_I + r_t)} \right)^2 \right\} + I_{dp} t_p \left\{ 1 - \left(\frac{r_I}{(r_I + r_t)} \right)^2 \right\} \right]$$

where $r_t = \sqrt{t_p/t_n}$, $r_I = I_{dp}/I_{dn}$, and the empirically determined proportionality constant $\alpha = 1/6$. t_n and t_p are the RC products for the nMOS and pMOS, respectively. *Fig. 16* shows that this model gives a good fit to the experimental data.

Using this model the requirements on ρ_{sh} for various technology nodes may be estimated corresponding to a roadmap for I_d , t_{ox} , L_g and V_{cc} . *Fig. 17* shows such a roadmap for technology nodes in the sub- $0.25 \mu\text{m}$ regime. The gate sheet requirements for $P_s/P_a \leq 10\%$ are shown in *Fig. 18*. In *Fig. 18* we assume a symmetric inverter with equal ρ_{sh} for nMOS and pMOS gates and that at each technology node the device widths are

scaled for equal I_d/V_{cc} . The gate sheet requirement is insensitive to technology node because as L_g and V_{cc} are reduced, a smaller width is needed for equal I_d/V_{cc} .

4. Conclusions

In conclusion, we have shown that as V_T is reduced below 0.1V there is a significant increase in P_s/P_a . Not more than half this increase can be explained by an increase in subthreshold leakage. P_{leak} is seen to be a major component of wasted power at very low V_T but P_s is comparable to P_{leak} at high operating frequencies. The effect of scaling to lower V_{cc} is seen to reduce P_s/P_a . For $S_i \simeq S_o$, such as in inverter chains, P_s/P_a is below 10% at both high and low V_{cc} . P_s/P_a depends primarily on input and output slew rates and V_T . We compute the constraints on the slew-rates for V_T s ranging from 0.2V to 0V and show that the minimum input slew rate is determined by short power at low V_T and by performance at relatively higher V_T . Increase in short power with increasing gate sheet resistance is observed. A simple analytical model for this phenomena is derived. Based on this model gate sheet requirements for $P_s/P_a < 10\%$ is calculated for a particular technology roadmap for sub- $0.25 \mu\text{m}$ CMOS. It is seen that the requirement is approximately the same for technology nodes down to $L_g = 0.13 \mu\text{m}$.

Acknowledgment

We would like to thank H. Tran, S.S. Mahant-Shetti, and A. Shah for invaluable technical assistance.

References

- [1] T. Yamazaki, K. Goto, T. Fukano, T. Sugii and T. Ito, *IEDM* 1993, p.906.
- [2] A. Chatterjee, M. Rodder and I.-C. Chen, *Proc. SPIE Microelectronic Device and Multilevel Interconnection Tech.*, pp.115, 1995.
- [3] P. Yang and P.K. Chatterjee, *IEEE Trans. on Computer-Aided Design*, pp.169-182, 1982.
- [4] M. Nandakumar et. al 1995 *IEEE Symp. Low Power Electronics*, p.80
- [5] H.J.M. Veendrick, *J. of Solid State Circuits*, pp.468-473, 1984.

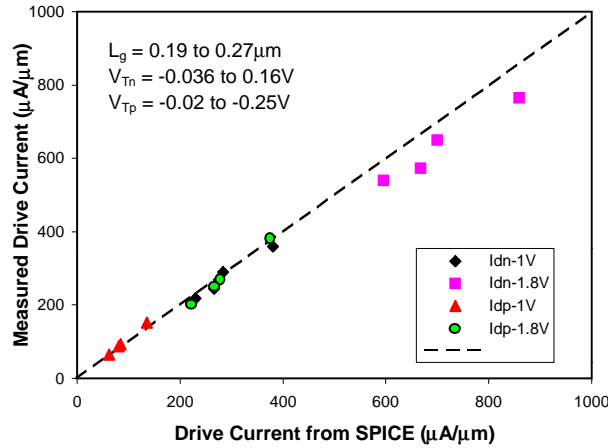


Fig.1 Comparison of measured and simulated drive currents. Good agreement is seen between experiment and SPICE dc simulations.

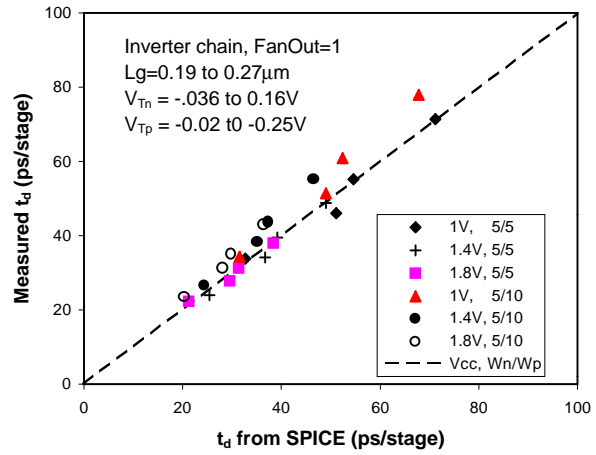


Fig.2 Comparison of measured and simulated inverter delays. Good agreement is seen between experiment and SPICE transient simulations.

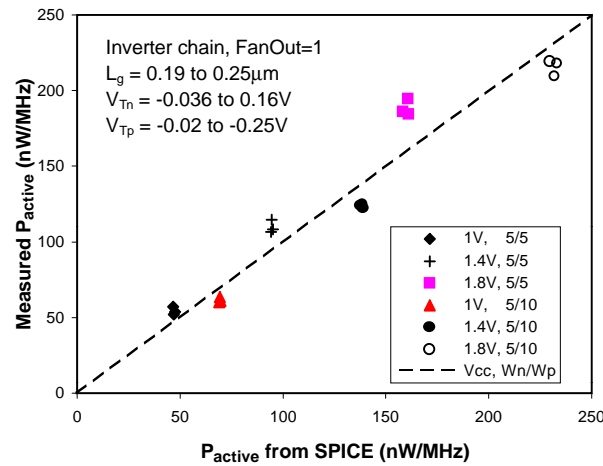


Fig.3 Comparison of measured and simulated active power showing capability of SPICE simulations in predicting power dissipation.

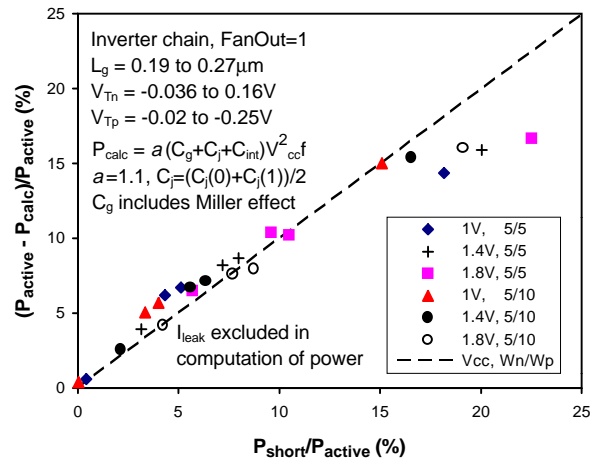


Fig.4 SPICE simulations show fraction of P_{active} in excess of aCv^2f correlates well with P_{short}/P_{active} (fraction wasted as short power). Constant $a=1.1$ is chosen for zero excess power at highest V_T .

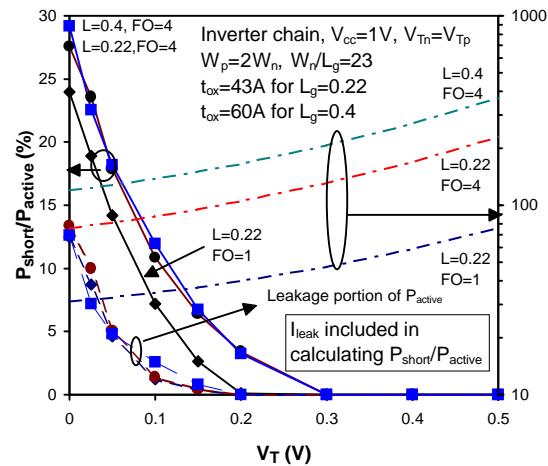


Fig.5 SPICE simulations show P_{short}/P_{active} increase at low V_T . Less than half of P_{short} is due to higher leakage. P_{short}/P_{active} is insensitive to fanout. A slower input slew at $FO=4$ is compensated by slower output slew and higher P_{active} .

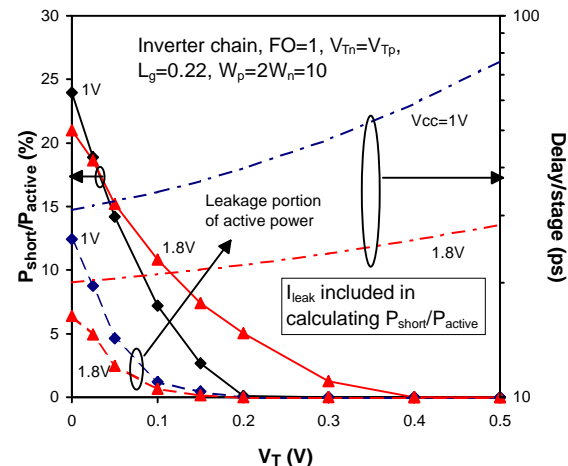


Fig.6 SPICE simulations of P_{short}/P_{active} vs V_T for $V_{cc}=1V$ and $1.8V$. P_{short}/P_{active} is insensitive to V_{cc} . A higher short current and faster output slew at $1.8V$ is compensated by a higher P_{active} and faster input slew rate.

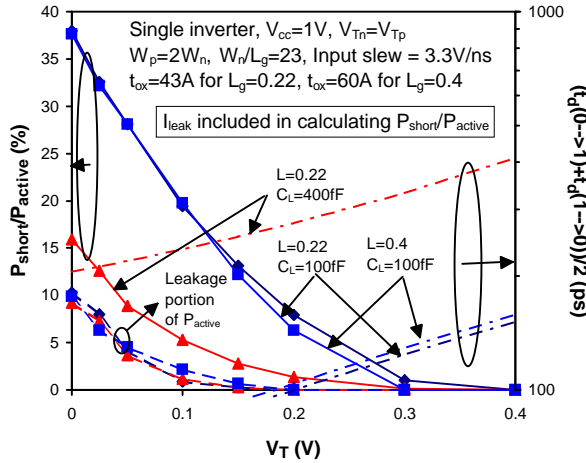


Fig.7 SPICE simulations of a single inverter at a fixed input slew of 3.3V/ns show $P_{\text{short}}/P_{\text{active}}$ increases at low V_T . Note that here $P_{\text{short}}/P_{\text{active}}$ is sensitive to the load C_L . Recall that the inverter chains are insensitive to fanout.

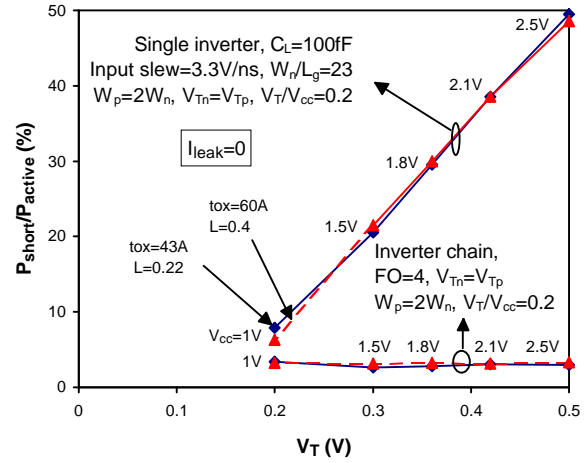


Fig.8 SPICE simulations show that reducing V_{cc} reduces $P_{\text{short}}/P_{\text{active}}$ for single inverters with a fixed input slew rate of 3.3V/ns. In contrast $P_{\text{short}}/P_{\text{active}}$ is roughly constant for corresponding inverter chains. $V_T/V_{cc}=0.2$ for all cases.

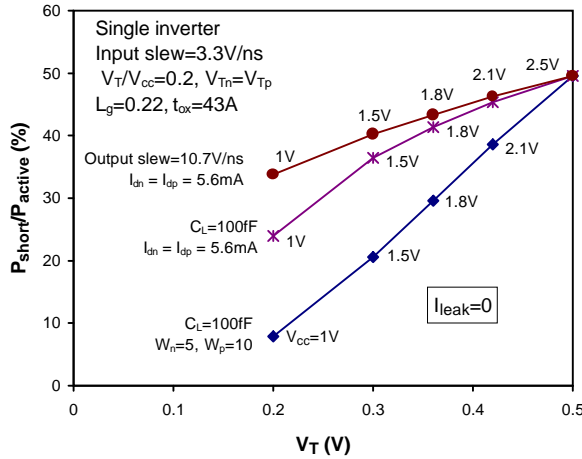


Fig.9 Reduction in $P_{\text{short}}/P_{\text{active}}$ with reducing V_{cc} is smaller if W 's are increased to keep I_{drive} constant. Reduction in $P_{\text{short}}/P_{\text{active}}$ is least if, in addition, C_L is reduced to keep output slew approximately constant. $V_T/V_{cc}=0.2$ for all cases.

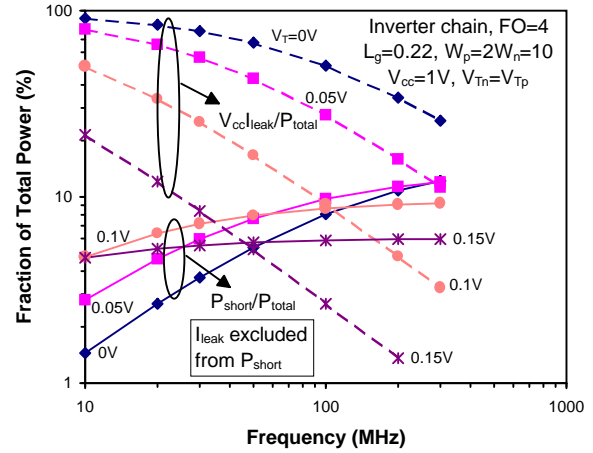


Fig.10 SPICE simulations of P_{short} and leakage power (shown as fractions of total power) vs frequency for $V_T=0, 0.05, 0.1V$. P_{short} is a significant fraction of wasted power at the higher frequencies

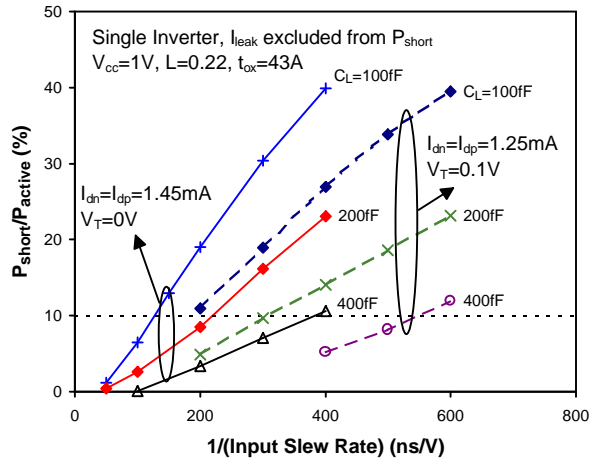


Fig.11 SPICE simulations of $P_{\text{short}}/P_{\text{active}}$ for $V_T=0, 0.1V$ with the leakage component subtracted off. $P_{\text{short}}/P_{\text{active}}$ increases for longer input transitions, lower C_L and lower V_T . For equal input slew rate a faster circuit has higher $P_{\text{short}}/P_{\text{active}}$.

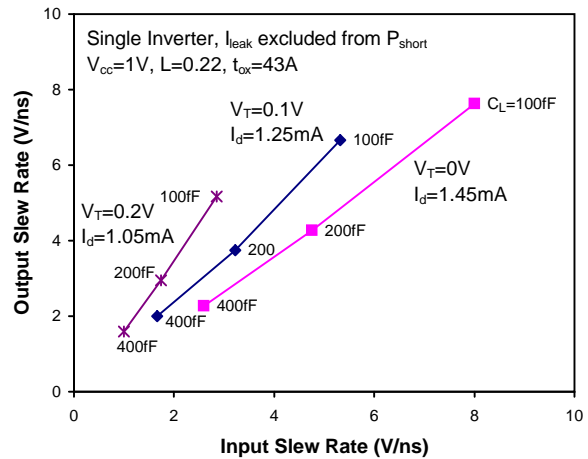


Fig.12 Contours of $P_{\text{short}}/P_{\text{active}}=10\%$ for $V_T=0, 0.1, 0.2V$ corresponding to $I_{\text{dn}}=I_{\text{dp}}=1.05, 1.25, 1.45 \text{ mA}$. At higher V_T a lower input slew can be tolerated for the same output slew. Thus a smaller pre-driver can be used while $P_{\text{short}}/P_{\text{active}} < 10\%$.

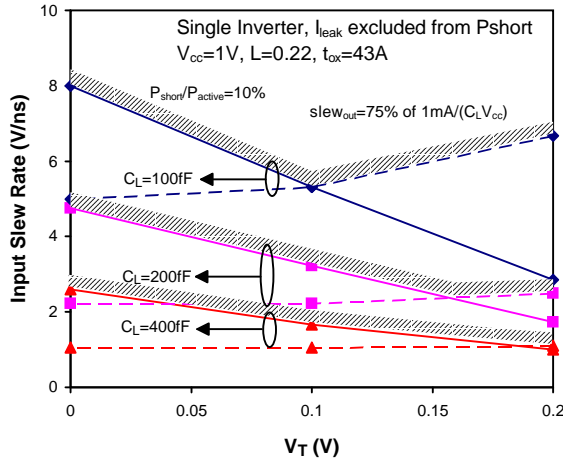


Fig.13 Lower bounds on input slew rate vs V_T for varying C_L . Solid lines correspond to $P_{\text{short}}/P_{\text{active}} < 10\%$ and dashed lines to output slew rate $> 75\%$ of $1\text{mA}/(C_L V_{cc})$. At low V_T input slew is constrained by power and at high V_T by performance.

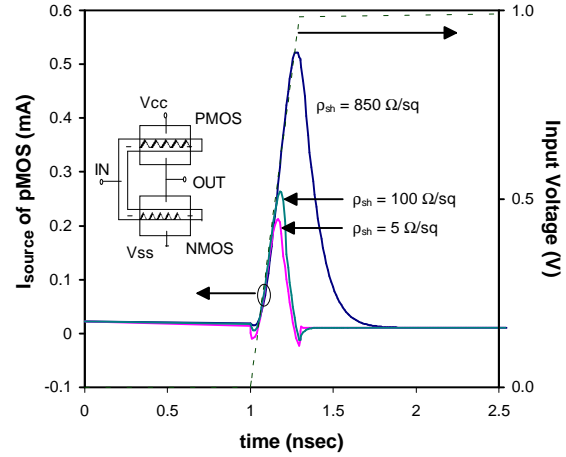


Fig.14 SPICE simulations of short-circuit current during the pull-down transition in an inverter with resistive gates for various gate sheets. Here short-current increases with increasing gate sheet because of the RC delay in turning off the pMOS.

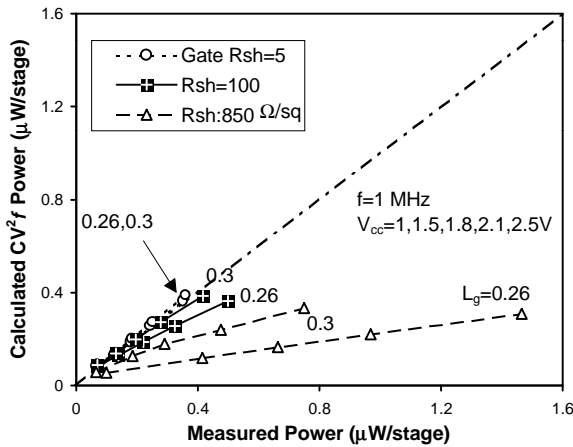


Fig.15 Measured active power vs. calculated CV^2f . Measured power exceeds CV^2f for smaller L_g and larger gate sheet resistance. The sheet resistance increases with decreasing gate length and the number shown corresponds to $L_g = 0.26\mu\text{m}$.

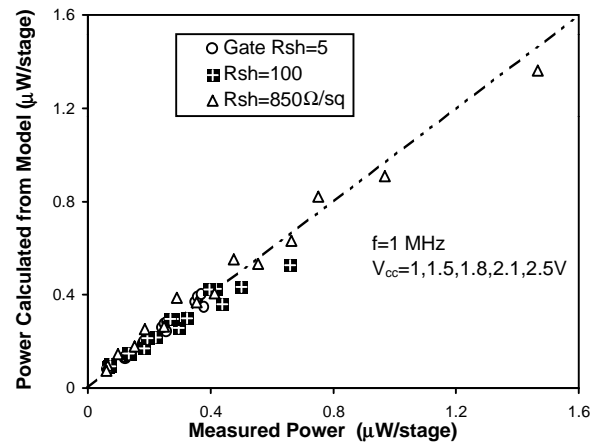


Fig.16 Measured power vs. power calculated using the model for short power. The model shows good agreement with measured power for different gate sheet resistances. Gate sheet resistance shown corresponds to $L_g = 0.26\mu\text{m}$.

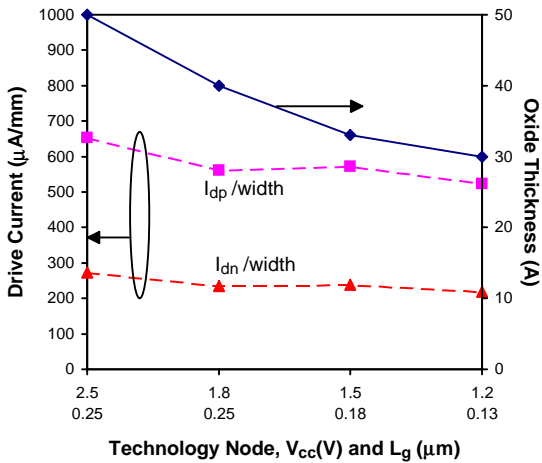


Fig.17 A technology roadmap for drive currents and t_{ox} . A technology node has been defined by gate length and V_{cc} .

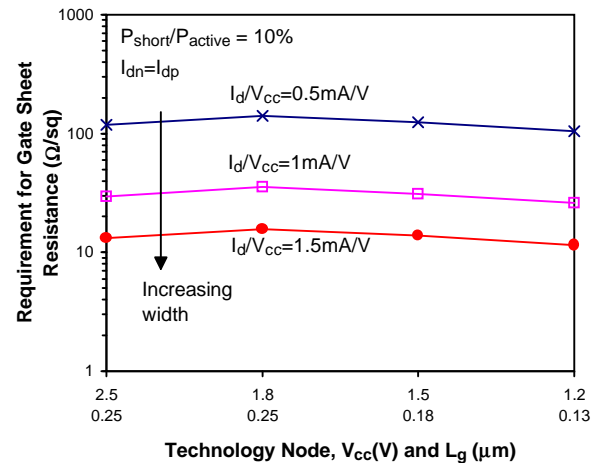


Fig.18 Scaling of gate sheet requirement to maintain $P_{\text{short}}/P_{\text{active}} < 10\%$ for $I_d/V_{cc} = 0.5, 1, 1.5\text{ mA/V}$. Gate sheet requirement is insensitive to technology node. As L_g and V_{cc} are reduced a smaller W is needed for equal I_d/V_{cc} .