

# CMOS Metal Replacement Gate Transistors using Tantalum Pentoxide Gate Insulator

A. Chatterjee, R.A. Chapman, K. Joyner, M. Otake\*, S. Hattangady, M. Bevan, G.A. Brown, H. Yang, Q. He, D. Rogers, S.J. Fang, R. Kraft, A.L.P. Rotondaro, M. Terry, K. Brennan, S.-W. Aur, J.C. Hu, H.-L. Tsai, P. Jones, G. Wilk, M. Aoki\*, M. Rodder, and I.-C. Chen

Silicon Technology Development, MS 3702 Texas Instruments, P.O.Box 650311, Dallas TX 75265.

\*Tsukuba Research and Development Center, Texas Instruments, Japan

## Abstract

This paper reports a full CMOS process using a combination of a TiN/W Metal Replacement Gate Transistor design with a high dielectric constant gate insulator of tantalum pentoxide over thin remote plasma nitrided gate oxide. MOS devices with high gate capacitances equivalent to that for  $< 2$  nm  $\text{SiO}_2$  but having relatively low gate leakage are reported. Transistors with gate lengths near or below  $0.1 \mu\text{m}$  have good characteristics. Working CMOS circuits using  $\text{Ta}_2\text{O}_5$  gate insulator are demonstrated for the first time.

## 1. Introduction

Tantalum pentoxide gate insulator has been applied to  $3 \mu\text{m}$  Al gate NMOS [1], to  $0.35 \mu\text{m}$  CMOS using WSi/n+ poly gate [2], and to  $0.35 \mu\text{m}$  NMOS using TiN gate [3]. The Replacement Gate design [4] has the unique advantage that the anneal of the source/drain implants is performed before the gate insulator is formed. Thus, here a high anneal temperature of  $1000^\circ\text{C}$  can be used to fully activate the source/drain implants. In conventional designs [1]-[3] the S/D anneal temperature must be limited to  $800^\circ\text{C}$  to prevent degradation of the properties of the tantalum pentoxide or further growth of the interfacial oxide that would result in a reduction of effective capacitance.

## 2. Fabrication

Figure 1 explains the Replacement Gate process as applied to the use of  $\text{Ta}_2\text{O}_5$  gate insulator. First, standard poly-Si gate transistors are built using a twin well CMOS process with STI isolation. The poly-Si gate and  $\text{SiO}_2$  gate insulator are then replaced by an  $\text{SiO}_2/\text{Ta}_2\text{O}_5/\text{TiN}/\text{W}$  insulator/metal-gate stack. Figure 2 shows a cross sectional TEM of a transistor with approximately  $0.1 \mu\text{m}$  gate length. A layered PVD/CVD/PVD TiN film is used unlike the single CVD TiN film used for the NMOS transistors described in [4]. The multi-layered TiN film enables us to utilize the superior electrical qualities of the PVD TiN film [5] without causing any discontinuities along the sides of the gate slot. The conformal CVD TiN film ensures continuity along the sides of the slot. The poor conformality of the PVD TiN deposit causes a narrowing of the top of the gate slot. The TEM image shows that despite the breadloafing of the PVD TiN the CVD W forms a continuous layer over TiN film.

Figure 3 shows a high resolution TEM of the gate and channel interface for a transistor with  $0.1 \mu\text{m}$  gate length. For

this sample, the  $\text{Ta}_2\text{O}_5$  thickness is estimated to be about  $5.5$  nm. Note that it is difficult to obtain this thickness very accurately because the  $\text{Ta}_2\text{O}_5/\text{TiN}$  interface is not very distinct in the TEM image. The interfacial  $\text{SiO}_x\text{N}_y$  is  $2$  nm thick in the TEM as measured using the Si lattice fringes for calibration. Before nitridation by Remote Plasma Nitridation (RPN) the  $\text{SiO}_2$  was  $1.5$  nm thick measured by ellipsometry. It is not clear whether the  $0.5$  nm increase in thickness can be attributed solely to the  $\text{Ta}_2\text{O}_5$  anneal or whether some of the observed increase is due to the RPN process or even simply an artifact of estimating the position of the interfaces from a TEM image.

The disposable poly-Si gate is formed using DUV lithography and a linewidth reduction etch to obtain sub- $0.1 \mu\text{m}$  linewidths [6]. Figure 4 shows that the linewidth reduction is  $\sim 0.05 \mu\text{m}$ . Continuous poly-Si linewidths are obtained down to  $0.05 \mu\text{m}$  linewidths. The linewidth  $L_{\text{gate}}$  of the replacement metal gate MOSFET at bottom of the groove is decreased by the thickness of the  $\text{Ta}_2\text{O}_5$  film deposited on each side of the slot, as illustrated in Fig. 1 c), d), and e).

## 3. Results

Devices with  $5.3$  nm, and  $6.5$  nm  $\text{Ta}_2\text{O}_5$  are fabricated. The NMOS gate capacitances and currents are shown in Figs. 5 and 6 and compared to those for TiN/W- $\text{SiO}_2$  devices from [4]. From Fig. 5,  $C_{\text{gate}}$  for the  $5.3$  nm and  $6.5$  nm  $\text{Ta}_2\text{O}_5$  NMOS capacitors at  $V_G = -3\text{V}$  are  $17.3 \text{ fF}/\mu\text{m}^2$  and  $16 \text{ fF}/\mu\text{m}^2$ , respectively. The value of  $17.3 \text{ fF}/\mu\text{m}^2$  is derived by extrapolation because at the larger  $V_G$ 's the LCR meter fails to compensate for the higher leakage current through the thinner ( $5.3$  nm)  $\text{Ta}_2\text{O}_5$  film resulting in an apparent fall off of  $C_{\text{gate}}$  with increasing electric field. The effective oxide thicknesses corresponding to  $17.3 \text{ fF}/\mu\text{m}^2$  and  $16 \text{ fF}/\mu\text{m}^2$  are  $2.00$  nm and  $2.15$  nm ( $t_{\text{eff}} \equiv \kappa_{\text{SiO}_2}/C_{\text{gate}}$ ). Note that the  $t_{\text{eff}}$  values do not refer to physical  $\text{SiO}_2$  film thicknesses. The pure  $\text{SiO}_2$  film thicknesses would have to be approximately  $1.4$  nm and  $1.55$  nm to obtain capacitance values of  $17.3 \text{ fF}/\mu\text{m}^2$  and  $16 \text{ fF}/\mu\text{m}^2$ , respectively. This difference between  $t_{\text{eff}}$  values and physical film thicknesses is due to the known effects of Fermi-Dirac statistics and quantization of electronic states in the inversion layer.

The gate capacitances and currents for PMOS are shown in Fig. 7 and Fig. 8. The gate currents are particularly small for  $V_G < 0$  V using  $6.5$  nm  $\text{Ta}_2\text{O}_5$  for both NMOS and PMOS.

## 29.1.1

Fig. 9 illustrates that the use of Ta<sub>2</sub>O<sub>5</sub> results in lower J<sub>gate</sub> for a given target value of C<sub>gate</sub> at the probable value of supply voltage of 1.5V.

Figure 10 shows a plot of  $\epsilon_0/C_{g-acc}$ , versus  $t_{Ta2O5}$  for NMOS capacitors.  $\epsilon_0$  is the vacuum permittivity,  $C_{g-acc}$  is the measured capacitance per unit area of the gate insulator stack with the device biased in accumulation, and  $t_{Ta2O5}$  is the thickness of Ta<sub>2</sub>O<sub>5</sub> measured by TEM. The symbols show the experimental data, the solid curve shows the extrapolation to  $t_{Ta2O5} = 0$ , and the dashed lines indicate the upper and lower confidence limit of this extrapolation based on estimates of errors in the measured values of  $C_{g-acc}$  and  $t_{Ta2O5}$ . The intercept on the vertical axis represents the ratio of thickness/dielectric-constant for the interfacial layer. The TEM of Figure 3 shows that the thickness of the interfacial layer is 2 nm. This implies that the dielectric constant of the interfacial layer is 6.1. Such an increase dielectric constant of the interfacial layer is critical for achieving high gate capacitance and is consistent with earlier observations [4] of increased effective dielectric constant of thin oxides after remote plasma nitridation. Note that if the dielectric constant of the interfacial layer were 3.9 (pure SiO<sub>2</sub>), the film thickness would have to be 1.3 nm, in contradiction with the TEM.

Fig. 11 shows J<sub>gate</sub> vs. 1/T for Ta<sub>2</sub>O<sub>5</sub> and SiO<sub>2</sub>. The high activation energies for Ta<sub>2</sub>O<sub>5</sub> conduction are consistent with the results of Matsuhashi [7]. The gate current appears to be a mixed conduction mechanism involving Poole-Frenkel emission in the Ta<sub>2</sub>O<sub>5</sub>.

The thin oxide between the Ta<sub>2</sub>O<sub>5</sub> and the Si channel should limit interface and trap density. The fast interface density estimated from the charge pumping current (Fig 12) is high (2.8E11/cm<sup>2</sup>) relative to pure SiO<sub>2</sub> (2E10/cm<sup>2</sup>) but similar to previous studies for Ta<sub>2</sub>O<sub>5</sub> [2]. Electron and hole mobilities are shown in Fig. 13;  $\mu_p$  is similar and  $\mu_n$  is low compared to standard design W/TiN/SiO<sub>2</sub> gate [8].

Both NMOS and PMOS designs use TiN/W gates, shallow drain extenders and angled pocket implants. The mid-gap workfunction of TiN raises V<sub>TN</sub> and V<sub>TP</sub> by ~0.5V relative to n<sup>+</sup>/p<sup>+</sup> poly gates. Thus, compensating As and BF<sub>2</sub> implants are used to lower V<sub>TN</sub> and V<sub>TP</sub>. Figures 14 and 15 show V<sub>T</sub> and subthreshold slopes versus L<sub>gate</sub>. The NMOSFET's show good characteristics down to L<sub>gate</sub> = 0.06  $\mu$ m, but the PMOS V<sub>T</sub>'s are too low for L<sub>gate</sub> < 0.13  $\mu$ m. This difference may be due to several reasons: the NMOS uses a larger tilt-angle pocket implant, the As buried channel is shallower than the BF<sub>2</sub> buried channel, and the PMOS drain extension is more graded and has a larger overlap with gate. Figure 16 shows the drive current characteristics and Fig. 17 shows the subthreshold and gate current characteristics for a L<sub>gate</sub>=0.06  $\mu$ m NMOSFET. The PMOS characteristics are similarly good, but at L<sub>gate</sub> > 0.1  $\mu$ m.

Figures 18 and 19 show the NMOS and PMOS I<sub>OFF</sub> versus I<sub>ON</sub> characteristics. NMOSFET's have a reasonably high I<sub>ON</sub> of 940  $\mu$ A/ $\mu$ m at I<sub>OFF</sub>=1 nA/ $\mu$ m for V<sub>DD</sub> = 2.0 V, but

for V<sub>DD</sub>=1.5V I<sub>ON</sub> is low (520  $\mu$ A/ $\mu$ m) because of high V<sub>TN</sub> values at short L<sub>gate</sub>'s resulting from the large tilt-angle pocket implant. The corresponding PMOS I<sub>ON</sub>'s are reasonable for both V<sub>DD</sub>=1.5V (330  $\mu$ A/ $\mu$ m) and V<sub>DD</sub>=2.0V (540  $\mu$ A/ $\mu$ m).

CMOS inverter chains with fan-out=1 show low standby current, I(standby), combined with high effective capacitance even at V<sub>cc</sub> = 2.0V. If SiO<sub>2</sub> were used then the gate dielectric film for such high capacitance would be so thin that the gate current would cause a significant increase in I(standby). Fig. 20 shows a plot of I(standby) and active current, I(active), at 1 MHz clock frequency versus delay time per inverter stage with L<sub>gate</sub> as the fundamental variable. The Ta<sub>2</sub>O<sub>5</sub> thickness is 6.5 nm for this data. This type of plot can be used to determine a "maximum figure of merit" defined from the inverter delay of a chain with a gate length for which I(active) = I(standby), (FOM[max]=1/delay at intersection) [9]. The increase in I(standby) at short delays (and gate lengths) is due to the rapid fall-off of V<sub>TP</sub> with decreasing gate length. The estimated gate length for equal standby and active currents is 0.13  $\mu$ m (limited by PMOS V<sub>T</sub> roll-off). The delay times at this intersection are shown in Fig. 21 as a function of V<sub>cc</sub>. The results are similar for 5.3 nm and 6.5 nm Ta<sub>2</sub>O<sub>5</sub>. However, the 6.5 nm thickness results in lower I(standby).

#### 4. Conclusions

The metal Replacement Gate transistor design permits leaving the high-k gate insulator in its optimum state. Very high gate capacitance can be obtained with low gate current using Ta<sub>2</sub>O<sub>5</sub> even at 2.0V. Inverter chains with 2500 stages have been demonstrated operating at 2.0V with acceptable standby current. The Ta<sub>2</sub>O<sub>5</sub> gate insulator process can be improved to even lower gate currents [10].

#### References

- [1] J.-L. Autran, *et al.*, "Fabrication and Characterization of Si-MOSFET's with PECVD Amorphous Ta<sub>2</sub>O<sub>5</sub> Gate Insulator," *IEEE Electron Dev. Lett.* vol. 18, pp. 447-449, 1997.
- [2] I.C.Kiziyalli, *et al.*, "Stacked Gate Dielectrics with TaO for Future CMOS Technologies," *Symp. VLSI Tech.*, pp. 216-217, 1998.
- [3] Y. Momiyama, *et al.*, "Ultra-Thin Ta<sub>2</sub>O<sub>5</sub> Gate Insulator with TiN Gate Technology for 0.1 $\mu$ m MOSFETs," pp. 135-136, *Symp. VLSI Tech.*, 1997.
- [4] A. Chatterjee, *et al.*, "Sub-100 nm Gate Length Metal Gate Transistors Fabricated by a Replacement Gate Process," *IEDM Technical Digest.*, pp. 821-824, 1997.
- [5] H. Yang, *et al.*, "A Comparison of TiN Processes for CVD W/TiN Gate Electrode on 3 nm Gate Oxide," *IEDM Technical Digest.*, pp.459-462,1997.
- [6] W.W. Lee, *et al.*, "Fabrication of 0.06  $\mu$ m Poly-Si Gate Using DUV Lithography with a Designed Si<sub>x</sub>O<sub>y</sub>N<sub>z</sub> Film as an ARC and Hardmask," *Symp. VLSI Tech.*, pp. 131-132, 1997.
- [7] H.Matsuhashi, *et al.*, "Optimum Electrode Materials for Ta<sub>2</sub>O<sub>5</sub> Capacitors for High- and Low-Temperature Processes," *Jpn. J. Appl. Phys.*, vol. 33, pp. 1293-1297, 1994.
- [8] J.C.Hu, *et al.*, "Feasibility of W/TiN Metal-Gate for Conventional 0.13  $\mu$ m CMOS Technology and Beyond," *IEDM Technical Digest.*, pp. 825-828, 1997.
- [9] R.A. Chapman, *et al.*, "A Standby Current Limited Performance Figure of Merit for Deep-Submicron CMOS," *IEEE Transactions on Electron Devices*, vol. ED-44, pp. 1888-1895, 1997.
- [10] S. Kamiyama, *et al.*, "Highly Reliable 2.5nm Ta<sub>2</sub>O<sub>5</sub> Capacitor Process Technology for 256 Mbit DRAMs," *IEDM Technical Digest.*, pp. 827-830, 1991.

## 29.1.2

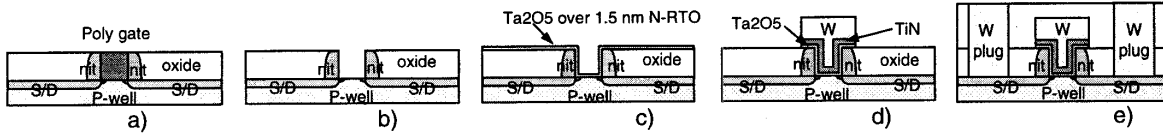


Figure 1. Steps in the fabrication of a metal replacement gate transistor. a) after CMP planarization of polysilicon disposable gate transistor with nitride sidewalls, b) after wet etch removal of polysilicon gate and gate oxide, c) after 1.5nm RPN-RTO then conformal deposition of Ta<sub>2</sub>O<sub>5</sub> followed by an 800C RTA in N<sub>2</sub>, d) after deposition of TiN and 80nm W metal layers and pattern/etch to form t-gate, e) after completion of the transistor with W-plug contacts to the source/drains.

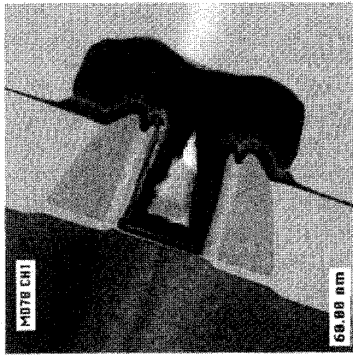


Figure 2. TEM cross section of 0.1µm long W/TiN Replacement T-gate with 5nm Ta<sub>2</sub>O<sub>5</sub> gate insulator. Nitride sidewalls before and after shallow drain extender shown. TiN layer is a composite of CVD (conformal) and PVD (non-conformal) depositions. The larger thickness of PVD TiN at the top of the slot narrows the slot and contributes to the formation of a void inside the tungsten gate.

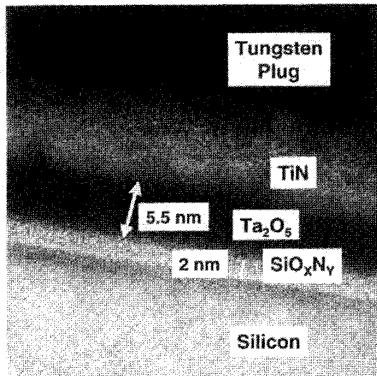


Figure 3. High resolution TEM of gate insulator at the bottom of a 0.1µm gate. The Ta<sub>2</sub>O<sub>5</sub> thickness is approximately the same in large area capacitors, but the TiN thickness is much larger in wide area capacitors than seen in this TEM.

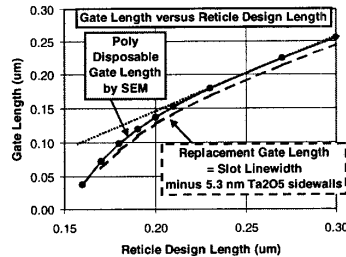


Figure 4. Polysilicon disposable gate linewidth vs. design length. Also shown is the calculated gate length for the metal replacement gate after correcting for the decrease in the slot width due to the 5.3nm thick conformal Ta<sub>2</sub>O<sub>5</sub> film. Linewidth control along a 0.05µm line is +/-0.015µm (3\*σ).

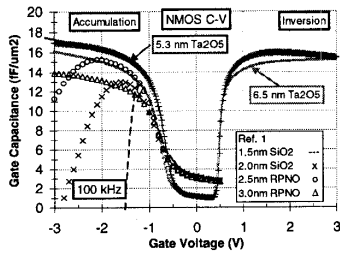


Figure 5. NMOS Gate Capacitance vs. Gate Voltage with Ta<sub>2</sub>O<sub>5</sub> gate insulator of 5.3nm, and 6.5nm thicknesses with pure SiO<sub>2</sub> and nitrided-RTO (RPNO) NMOS from Ref. 4. All cases are with TiN/W gates over p-type silicon. Note that case with 6.5 nm Ta<sub>2</sub>O<sub>5</sub> has C<sub>gate</sub> similar to case with 2.5nm RPNO. Dashed lines at upper left show extrapolation of capacitance values to -3V.

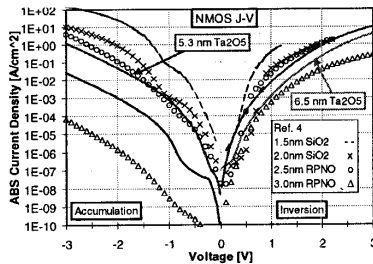


Figure 6. NMOS Gate current vs Gate Voltage with p-type substrate for Ta<sub>2</sub>O<sub>5</sub> cases of 5.3nm and 6.5nm thicknesses. Note 6.5nm Ta<sub>2</sub>O<sub>5</sub> case has considerably lower gate current than Ref. 4 case using 2.5nm nitrided-RTQ (same capacitance).

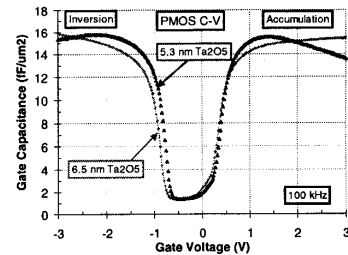


Figure 7. PMOS Gate Capacitance vs. Gate Voltage with Ta<sub>2</sub>O<sub>5</sub> gate insulator of 5.3nm and 6.5nm thicknesses. Both cases are with TiN/W gates over n-type silicon and have BF<sub>2</sub> compensating channel implants.

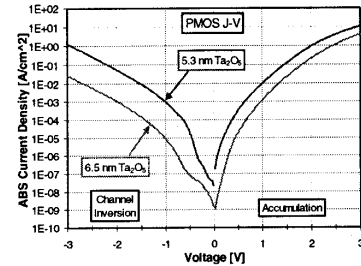


Figure 8. PMOS Gate current vs. Gate Voltage with n-well for Ta<sub>2</sub>O<sub>5</sub> cases of 5.3nm and 6.5nm thickness. Note that the gate currents are similar for NMOS and PMOS for Ta<sub>2</sub>O<sub>5</sub>. For 6.5nm Ta<sub>2</sub>O<sub>5</sub>, both NMOS and PMOS have lower gate currents for negative voltages than for positive voltages.

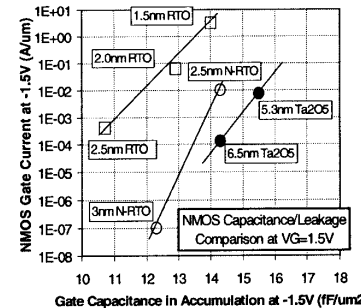


Figure 9. The use of Ta<sub>2</sub>O<sub>5</sub> gate insulator results in higher values of capacitance for the same value of gate leakage in accumulation compared to either pure SiO<sub>2</sub> or nitrided silicon dioxide gate insulators

## 29.1.3

IEDM 98-779

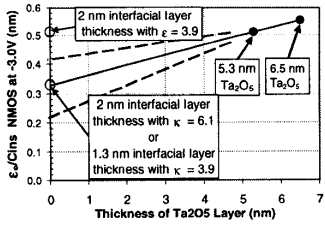


Figure 10.  $\epsilon_0/C_{g-acc}$  vs thickness of  $Ta_2O_5$  where  $C_{g-acc}$  is gate capacitance of the NMOS in accumulation. Dashed lines show confidence limits in extrapolation from  $Ta_2O_5$  data. The dielectric constant of  $Ta_2O_5$  obtained from the slope is 29. The intercept on the vertical axis is  $t/k$  where  $t$  is the interfacial layer thickness,  $k$  is its dielectric constant. Using  $t=2nm$  from TEM (Fig.3),  $k=6.1$ .

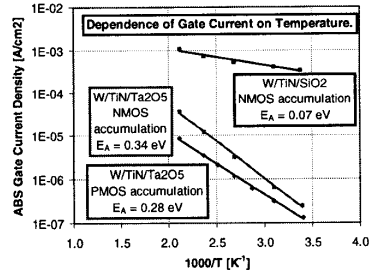


Figure 11. Gate leakage as a function of temperature comparing NMOS and PMOS in accumulation for  $Ta_2O_5$  to  $SiO_2$  gate insulator. The large activation energies for  $Ta_2O_5$  suggest a mixed conduction mechanism including Poole-Frenkel emission.

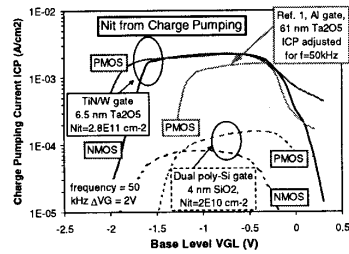


Figure 12. Fast interface state density for  $Ta_2O_5$  and pure  $SiO_2$  compared to prior work Ref.-2.

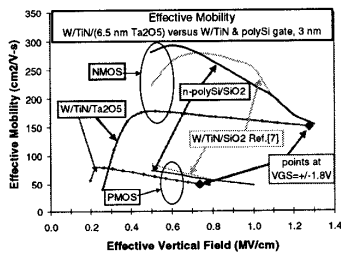


Figure 13. Hole mobility is similar to that reported in Ref.-8, but electron mobility is smaller.

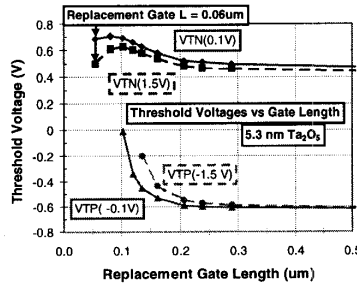


Figure 14. NMOS and PMOS  $V_T$ 's vs. replacement gate length. NMOS works well to 0.06 $\mu m$ . Shorter gate lengths fail because poly-Si linewidth varies ( $3\sigma = 0.015\mu m$ ) along the width of the transistor.

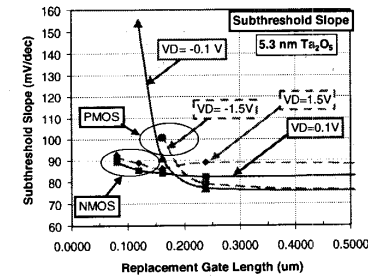


Figure 15. Subthreshold slope vs. replacement gate length for NMOS & PMOS.

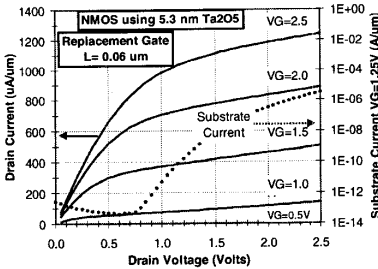


Figure 16. NMOS drive characteristics and substrate current for typical strong transistor.

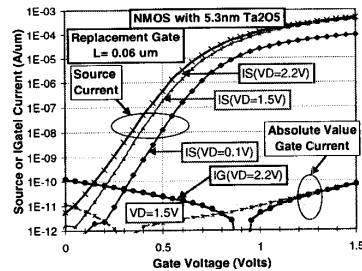


Figure 17. Typical NMOS strong transistor subthreshold current and gate current. The characteristics for PMOS transistors are similarly good, but at longer gate lengths.

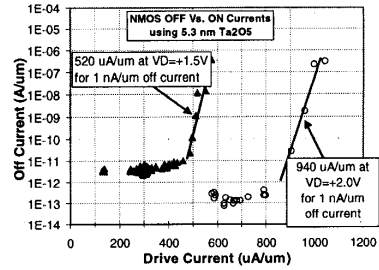


Figure 18. NMOS off-current vs. Drive Current for 1.5V and 2.0V.

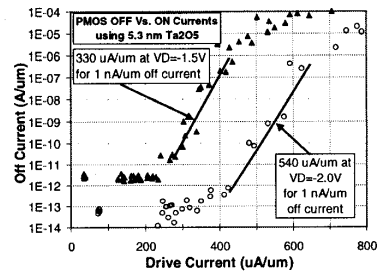


Figure 19. PMOS off-current vs. drive current for  $V_D=1.5V$  and  $2.0V$ .

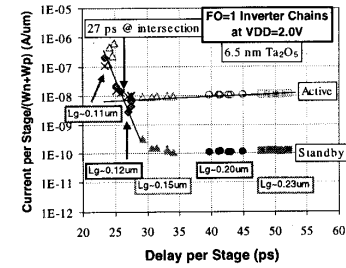


Figure 20. Inverter chain standby and active currents vs. delay time. Intersection at 27 ps is a "maximum FOM" quality factor [Ref. 9] when standby = active current. At long gate lengths (long delays), the standby current is limited by gate current.

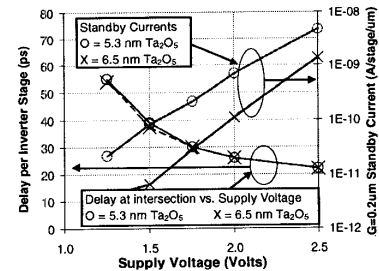


Figure 21. Inverter chain FOM(max) and standby current vs. supply voltage. Use of 6.5nm  $Ta_2O_5$  results in lower standby current for same speed.

## 29.1.4