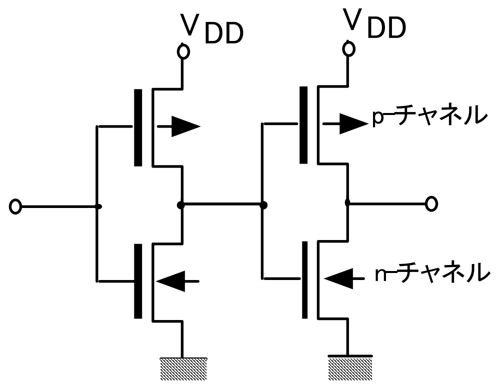
第 11 回　MOSFET　インバータと高速動作

インバータでの見積もり

　トランジスタが動作を始めたが、できれば早く動かしたい。早く動けばコンピューターの application も早く動く。パソコンのスピードの指標の一つにクロック周波数があるが、今は GHz を越える周波数があたりまえになってきた。この高速化はトランジスタが高速化することでなりたっている。CMOS の最も重要な応用はデジタル回路なので、デジタル回路の基礎となるインバーターで充電時間の大まかな見積もりをしよう。

　下に 2 段の CMOS のインバータ回路の大まかな回路図をしめす。



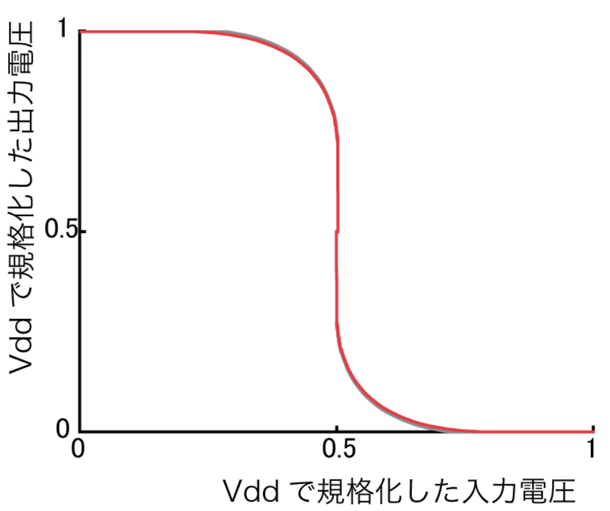
このインバータ回路で、まずは一段のみで、始め入力が Lo になっている状態を考えよう。簡単にするために Hi の電圧は電源電圧である V_{DD}、Lo は接地電位である(=0V)と考えよう。
まず入力電圧 V_{IN} が Lo=0V のとき、下側の n-チャネルのトランジスタが、V_{IN} が V_T より低いので遮断されている。上側の p-チャネルのトランジスタでは、ソースからみて- V_{DD} の電圧がゲートにかかる(p チャネルに注意)状態なので電圧差が有る限り電流が流れることが出来る。そこで p-チャネルトランジスタのソースドレイン間に電圧差が無くなるまで電流が流れ、出力電圧 V_{OUT} は Hi である V_{DD} になる。
入力電圧を Hi にすれば、p-チャネルのトランジスタのソースから見てゲートが同じ電位なのでは遮断されている。一方 n-チャネルトランジスタは電流が流れられる。そこで V_{OUT} は 0。これでインバータとなり、定常状態では電流は流れない。
　次に V_{IN} が変化した時を考えよう。まずは定常状態で V_{IN} が 0 でも V_{DD} でも無い時を考えよう。簡略化のために、pMOS と nMOS の移動度が等しく、同じ最大電流を得るためにチャネル幅 W は同じで良いとする。(実際は移動度は 2 倍以上違い、W を変えて $\mu\frac{W}{L}$ が同じになる様にする。)
$V_T > V_{IN} > 0$ では、n-チャネルトランジスタは遮断したままで変わらない。V_{OUT} は V_{DD} になる
$V_{DD}/2 > V_{IN} > V_T$ では、n-チャネルのゲート電圧がしきい値を超えるので流れることができる。ドレイン・ソース間には始め充分電圧がかかっているので、飽和領域で動作する。$C_{OX} = \frac{\varepsilon_{OX}}{d}$ として、
$I_n = \mu C_{OX}\frac{W}{2L}(V_{IN} - V_T)^2$。一方ドレインソース間にあまり電圧が印加されていない p-チャネルは線形領域であり、
$I_p = \mu C_{OX}\frac{W}{2L}(2(V_{DD} - V_T - V_{IN})(V_{DD} - V_{OUT}) - (V_{DD} - V_{OUT})^2)$ の電流が流れる。次段の入力はゲートで絶縁されているので、この電流が定常状態で流れられるのは下側の n-チャネルトランジスタだけであり、両者の式を解けば V_{OUT} がでる。電流は流れ続ける。
$V_{DD} - V_T > V_{IN} > V_{DD}/2$ では、逆に p-チャネルは充分電圧がかかっていて、飽和領域で動作する。$I_p = \mu C_{OX}\frac{W}{2L}(V_{DD} - V_{IN} - V_T)^2$。n-チャネルは、線形領域であり、$I_n = \mu C_{OX}\frac{W}{2L}(2(V_{IN} - V_T)V_{OUT} - V_{OUT}^2)$。両者の式を解けば V_{OUT} がでる、
$V_{DD} > V_{IN} > V_{DD} - V_T$ では p-チャネルが遮断して V_{OUT} は 0 となる。
対称性から考え計算しやすい $V_{DD} - V_T > V_{IN} > V_{DD}/2$ で V_{OUT} を計算しよう。

$$(V_{DD} - V_{IN} - V_T)^2 = 2(V_{IN} - V_T)V_{OUT} - V_{OUT}^2$$ 、

$$V_{OUT}^2 - 2(V_{IN} - V_T)V_{OUT} + (V_{DD} - V_{IN} - V_T)^2 = 0$$

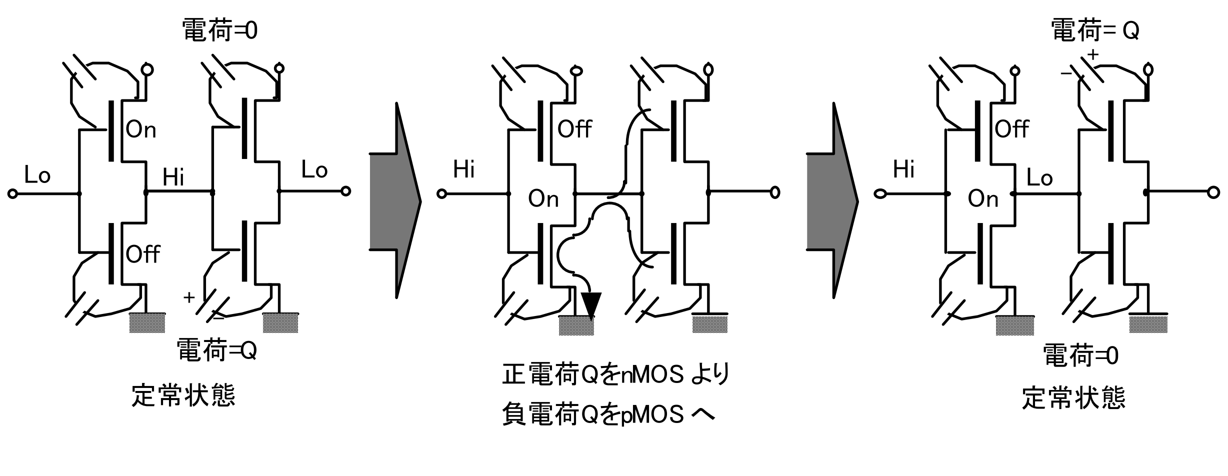
$$V_{OUT} = V_{IN} - V_T \pm \sqrt{(V_{IN} - V_T)^2 - (V_{DD} - V_{IN} - V_T)^2}$$

$V_{DD} - V_T = V_{IN}$ で $Vout = V_{DD} - 2V_T \pm \sqrt{(V_{DD} - 2V_T)^2}$ なので、マイナスを取るべき。$V_{IN} = V_{DD}/2$ で $V_{OUT} = V_{DD}/2 - V_T$ なので、$V_{DD}/2 + V_T > V_{OUT} > V_{DD}/2 - V_T$ は定常状態ではどんな V_{IN} でも作れない不安定領域であることが判る。$V_T = 0.2\,V_{DD}$ で計算すると下の様なカーブになる。このカーブを伝達特性またはトランスファーカーブと呼ぶ。このカーブを実際には後述する完全に飽和しない・サブスレッショルド特性などによりもっと緩いカーブが出る。



（CMOS の優れた点は、定常状態=保持状態に電流が流れず、低消費電力になることである。これ以外ののインバータでは、駆動用のトランジスタがオンになっている状態を保持すると、電流が流れ続ける。）

次に、2 段のインバータで入力が Lo、出力 Hi の段階から、入力を Hi に変えてやり、出力が Lo に変わるまでにどのくらい時間がかかるかみよう。



入力が切り変わってから、トランジスタのオン、オフ状態はすぐ変わる。一方、出力の電位は出力段のゲートがコンデンサとして働くので、このコンデンサの電荷がなくならないと出力は Lo にならない。即ち n-MOS のトランジスタで、V_{DD} の電圧まで充電された次段の n-MOS のコンデンサ 1 個を放電し、0 の電圧の次段の p-MOS のコンデンサ 1 個を V_{DD} まで充電することに相当する。すなわちコンデンサとしては 2 個分になる。

この容量を充放電する電流はオンとなった一個の nMOS トランジスタであり、このトランジスタはゲート電圧がV_{DD}であり、当初ドレイン電圧もV_{DD}なので、$I_{sat}=\mu\frac{\varepsilon_{ox}}{d}\frac{W}{L}\frac{(V_{DD}-V_T)^2}{2}$の電流が流れる。

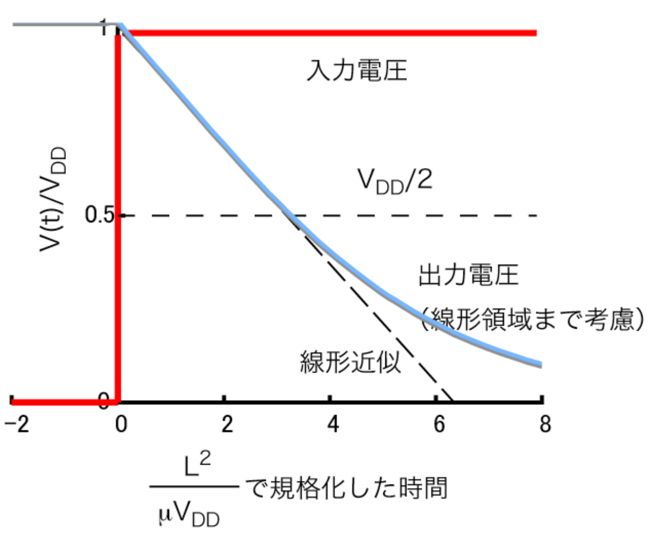
このトランジスタのドレインのバイアス条件は充放電状態に寄って代わり、ドレイン電圧がオーバードライブ電圧より小さくなった場合は電流が小さくなり指数関数的な降下になって、出力電圧が完全に切り替わるには相当の時間がかかるが、通常出力電圧が$V_{DD}/2$まで落ちる時間を遅延時間として定義し、その間ではドレイン電流は大きく変わらないとして計算する。

次段の充電/放電する容量は pMOS、nMOS それぞれの容量を考えなければならない。すると充放電する容量は$C_T=2C_{ox}LW$となる。

すると充電時間は$\tau=\frac{C_TV_{DD}}{2I_{sat}}=\frac{2L^2V_{DD}}{\mu(V_{DD}-V_T)^2}$となる。$V_T=\alpha V_{DD}$とすると、$\mu\frac{\varepsilon_{ox}}{d}\frac{W}{L}\frac{(V_{DD}-V_T)^2}{2}\tau=\frac{C_TV_{DD}}{2}$で$\tau=\frac{2L^2}{\mu V_{DD}(1-\alpha)^2}$となる。$V_T\approx 0.2V_{DD}$で計算すると、$\tau\approx\frac{3.1L^2}{\mu V_{DD}}$となる。チャネル長を短くすることが非常に重要となる。また移動度を高く保つことと電源電圧をあげることも有利になる。この電源電圧の依存性がオーバークロックの方法の一つである。(実際には、上げ過ぎると壊れる。13 回で説明しよう。)

しきい値は低い方がオーバードライブ電圧が大きくなり駆動力の上昇で高速動作が可能となるが、低くなりすぎると後述するサブスレッショルド特性によるオフ電流が大きくなることから、一般的には$V_T\approx 0.2V_{DD}$程度が用いられていた。これも最近はサブスレッショルド対策で大きくなりつつある。

念のため$V_T\approx 0.2V_{DD}$の場合に線形領域の遷移まで考慮した遷移を示す。$\tau\approx\frac{3.2L^2}{\mu V_{DD}}$で 3%の誤差しかない。

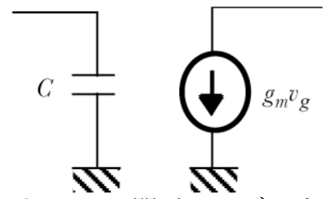


従って、チャネル長を短くすることが非常に重要となる。また、移動度が高い方がよい。電圧が高い方が早く動く。

インテルが 2001 年にテラヘルツトランジスタという呼称を発表した。これは$\frac{C_{MOS}V_{DD}}{I_{sat}}$が 1ps になったことを言っていた。実際のクロック動作で 1THz を作れる見込みは現時点のシリコンではない。いま駆動されるインバータは 1 つだった。しかし、これではインバータの直列接続だけで発振器を作れても論理動作はできない。通常 3 つ程度の回路を駆動する(fan-out=3 等と呼ぶ)。従って、コンデンサの容量が増え、さらに速度は遅くなる。さらに、遠いところの回路を駆動するとその間の配線の容量も無視できなくなる。また目安が充電時間なので、配線の抵抗も配慮すべきである。そのため、最近では単に負荷容量としている例もあり、この場合 CMOS の容量に比例しない。ただそれだと速度が粗く見積もれないので、ここでは昔ながらの MOS 容量のみで考える。

遮断周波数

さて、もう一つの速度の性能指標として、バイポーラトランジスタでも扱った遮断周波数(cutoff frequency: f_T)の話をしよう。遮断周波数とは、出力端を短絡したときの出力電流対入力電流の比率が 1 になる周波数で定義される。小信号等価回路で考えたときの四端子パラメータの一つである h_{21} が 1 になる周波数と定義しても良い。小信号時の電流利得で定義されるので、アナログ応用のための指標である。本来ゲートに電流は流れないが、交流ならば電流は流れるので、電流を定義できる。まず、等価回路で MOSFET を表そう。



等価回路は、通常ゲート-ソース間がコンデンサで、ドレインソース間に電流源という回路がもっとも簡単でよく用いられる。他に容量としてドレイン-ソース間を考えるときもあるが、簡単にするため、無視しよう。

また、電流源のゲート電圧依存性を相互コンダクタンス g_m で表す。g_m は V_G の変化に対して、I_D がどのように変化するかであり、飽和領域では、$I_D=\frac{\mu W\varepsilon_{ox}}{Ld}\frac{(V_G-V_T)^2}{2}$なので、$g_m=\frac{dI_D}{dV_G}=\frac{\mu W\varepsilon_{ox}}{Ld}(V_G-V_T)$となる。

つぎに容量を求める必要がある。ソース端では電荷密度は$Q_i=-C_{ox}(V_G-V_T)$だが、ドレイン側に行くに従って電荷量が減るので全体では 2/3 ほど小さくなり、コンデンサの容量は、$C=\frac{2}{3}C_{ox}LW$となる。

入力電流 $i_G = jwCv_g$ 出力高周波電流 $i_D = g_m v_g$ なので、電流利得の絶対値は$\left|\frac{i_D}{i_G}\right|=\frac{g_m}{wC}$となる。直流時はゲートが絶縁されていることから無限大であり、そのあと周波数と反比例して下がる。利得が 1 となる遮断周波数は $f_T=\frac{w_T}{2\pi}=\frac{g_m}{2\pi C}$ となる。前の C と g_m の式を代入すると

$$f_T=\frac{\mu WC_{OX}}{2\pi L}(V_G-V_T)\frac{3}{2C_{OX}LW}=\frac{3\mu(V_G-V_T)}{4\pi L^2}$$ となる。

従って、やはり、チャネル長を短くすることが非常に重要となる。また、移動度が高い方がよい。電圧が高い方が早く動く。

当たり前といえば当たり前だが、係数を除けば、速度は電圧、移動度に比例し、チャネル長さの二乗に反比例する関係は、デジタルでも、アナログでも同じ

さて、ここで容量の計算時に掛けられた 2/3 がどのようにでたか、示そう。飽和領域でのゲートのコンデンサの電子の電荷量 Q を計算する必要がある。$(V_G-V_T)V_c(z)-\frac{V_c(z)^2}{2}=\frac{I_{Dsat}d}{\mu W\varepsilon_{ox}}z$ であり、

$I_{Dsat}=\frac{\mu W\varepsilon_{ox}}{Ld}\frac{(V_G-V_T)^2}{2}$ となる。

$$(V_G-V_T-V_c(z))^2=(V_G-V_T)^2-\frac{2I_{Dsat}d}{\mu W\varepsilon_{ox}}z$$

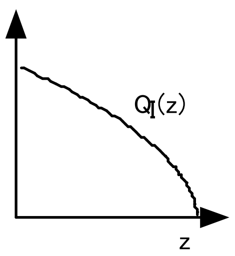
$$=(V_G-V_T)^2-\frac{2d}{\mu W\varepsilon_{ox}}\frac{\mu W\varepsilon_{ox}}{Ld}\frac{(V_G-V_T)^2}{2}z$$

$$=(V_G-V_T)^2\left(1-\frac{z}{L}\right)$$

ここから、 が、でて、

$$Q_I(z)=-\frac{\varepsilon_{ox}}{d}(V_G-V_T-V_c(z))$$

$$=-\frac{\varepsilon_{ox}(V_G-V_T)}{d}\sqrt{1-\frac{z}{L}}$$ がでる



$$Q=w\int_0^L Q_I(z)dz==-\frac{\varepsilon_{ox}W(V_G-V_T)}{d}\int_0^L\sqrt{1-\frac{z}{L}}dz=-\frac{2\varepsilon_{ox}LW(V_G-V_T)}{3d}$$

、($\int_0^L\sqrt{1-\frac{z}{L}}dz=\left[-\frac{2}{3\sqrt{L}}(L-z)^{3/2}\right]_0^L=\frac{2L}{3}$を使った。)

コンデンサの容量は、$C=\frac{2\varepsilon_{ox}LW}{3d}$であることが判る。