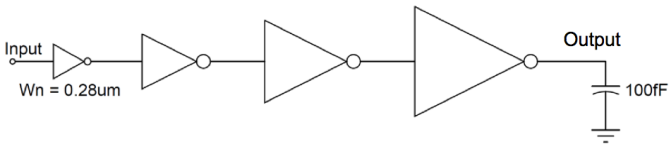
CMOS Inverter Dynamics: Designing for Propagation Delay and Power

Bhaskar Mookerji — 6.012 Fall 2010 — WF11*
(Dated: November 25, 2010)

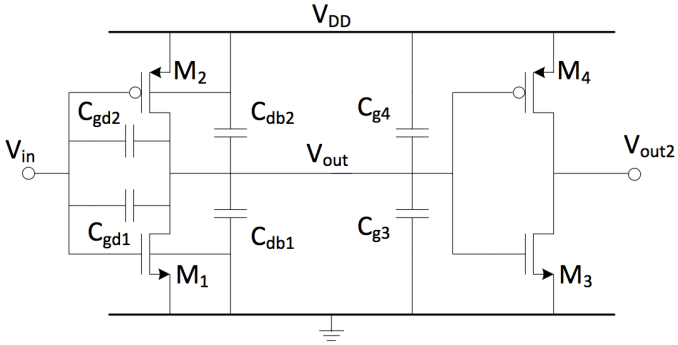
The propagation delay of a CMOS inverter circuit is determined by the charge-discharge time of its load; accounting for parasitic capacitances, MOSFET sizing and supply voltage are extremely important to designing high-performance CMOS circuits. We derive a general expression for the sizing of MOSFETS in an N-stage CMOS inverter circuit and discuss the relative influence of transistor sizing versus power-supply scaling. Using both numerical modeling (SPICE) and analytic arguments, we discuss power supply minimization using the energy-delay product (EDP) figure-of-merit.

1. INTRODUCTION

It is 2010, President Barack H. Obama faces defeat at the Midterm Elections unless a four-stage CMOS buffer made of inverters is designed to the GOP Minority Leader's specifications. The artful discussion covers the effects of circuit sizing and power supply voltage on speed, power, and energy dissipation in a relatively simple CMOS circuit. Figure 1 shows the four-inverter buffer and load, as well as the associated load capacitance for its first stage. Table I specifies the devices parameters of the FETS in the buffer. We first discuss the effects of circuit sizing and power supply voltage on power dissipation (Section 2), and conclude with a discussion of the circuit energy-delay product (EDP) using both dynamic and static power dissipation (Section 3).



(a)Design problem: A four-inverter buffer drives a 100fF load.



(b)Parasitic capacitances influencing the transient behavior of an inverter pair.

FIG. 1: Specified circuit parameters.

TABLE I: Device parameter values.

Parameter	Value	Parameter	Value
μ	$200\text{cm}^2/\text{V/s}$	$n = \alpha$	1.5
V_T	0.26V	$L_{\min}$	0.18μm
λ	0V^{-1}	$\rho = W_p/W_n$	2.0
V_{DD}	1.8V	C_{ox}	$8.6\text{fF}/\mu\text{m}^2$
I_0	42nA	C_{db}	$0.4\text{fF}/\mu$m
δ_{VT}	0V	C_{ov}	$0.4\text{fF}/\mu$m

2. POWER

2.1. Minimum delay specification

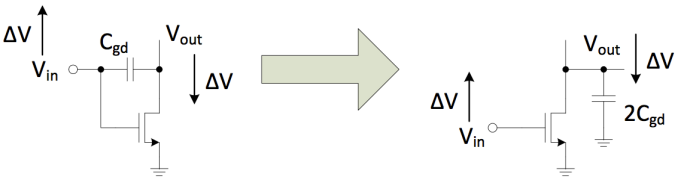


FIG. 2: Miller Effect. A voltage-driven MOSFET's floating gate-drain capacitance C_{gd} can be redefined with respect to ground.

In order to derive the conditions for minimum propagation delay t_p in the four-inverter buffer, we first derive the full input-output delay as a function of the circuit parameters specified in Table I. Noting that $\rho = W_p/W_n = 2$ is fixed and $V_{Tn} = -V_{Tp}$, the high-to-low (HL) and low-to-high (LH) propagation times are symmetrical, so

$$t_{pHL} = \frac{C_L \Delta V}{I_{Dn}} = \frac{C_L V_{DD}/2}{\frac{W_n}{L_n}\mu_n C_{ox} F_n^{SI}\left(V_{DD}, V_{DSAT_n}\right)} = t_{pLH}, \tag{1}$$

where

$$F_n^{SI}\left(V_{GS}, V_{DS}\right) = \left(V_{GS} - V_{Tn} - \alpha\frac{V_{DS}}{2}\right) V_{DS}, \tag{2}$$

C_L is the total load capacitance, the saturation voltage is $V_{DSAT_n} = V_{GS} - V_{Tn}$, and we have assumed that the

*Electronic address: mookerji@mit.edu

output-high voltage $V_{OH} = V_{DD}$. The average propagation time for a single MOSFET is

$$t_p = \frac{1}{2}\left(t_{pHL} + t_{pLH}\right) = t_{pHL}. \tag{3}$$

The load capacitance C_L in Equation 1 for a single inverter stage is given by

$$C_L = C_p + C_g, \tag{4}$$

where C_p is the total internal parasitic capacitance of the driving inverter and C_g is the external load capacitance. Equation 1 can be re-expressed as

$$t_{pHL} = \frac{\left(C_p + C_g\right) V_{DD}/2}{\frac{W_n}{L_n}\mu_n C_{ox} F_n^{SI}\left(V_{DD}, V_{DSAT_n}\right)} = t_{p0}\left(1 + \frac{C_p}{C_g}\right), \tag{5}$$

where

$$t_{p0} = \frac{C_p V_{DD}/2}{\frac{W_n}{L_n}\mu_n C_{ox} F_n^{SI}\left(V_{DD}, V_{DSAT_n}\right)} \tag{6}$$

is the propagation delay loaded only by the minimally-sized inverter's own intrinsic capacitances. Assuming uniform parasitic capacitances, as justified by Table I, the total delay of a chain of N inverters ($N = 4$) is

$$t_p = \sum_{i=1}^{N} t_{p0}\left(1 + \frac{C_{g,i+1}}{C_{g,i}}\right), \tag{7}$$

where $C_{g,N+1} = C_L$ is the fixed load capacitance at the end of the chain and $C_{g,1}$ is the input parasitic capacitance of the first inverter given by

$$\begin{aligned} C_{g,1} &= 2C_{gd2} + 2C_{gd1} + C_{db2}W_p + C_{db1}W_n \\ &= 2W_{p1}C_{ov} + 2W_{n1}C_{ov} + C_{db2}\rho W_{n1} + C_{db1}W_{n1} \\ &= W_{n1}\left(\rho + 1\right)\left(2C_{ov} + C_{db}\right) \end{aligned} \tag{8}$$

after applying the Miller effect equivalence shown in Figure 2. The minimum delay is achieved when adjacent MOSFETS are sized by a fixed ratio, or when (for $i = 2 \ldots N$)

$$\Delta = \frac{C_{g,i}}{C_{g,i-1}} = \frac{C_{g,i+1}}{C_{g,i}}. \tag{9}$$

The intermediate capacitances are unknown, but it can be shown by direct substitution of this expression that $C_L = C_{g,1}\Delta^N$, or that

$$\Delta = \left(\frac{C_L}{C_{g,1}}\right)^{1/N}, \tag{10}$$

and that

$$t_p = N t_{p0}\left(1 + \Delta\right). \tag{11}$$

Finally, from the specified device parameters, we know from $C_{g,1} = 1.008$pF, $C_L = 100$pF, and $N = 4$, that $\Delta = 3.15598$. Therefore, the inverter sizes are approximately

$$\begin{aligned} W_{n1} &= 0.28\mu\text{m} & W_{p1} &= 0.56\mu\text{m} \\ W_{n2} &= 0.884\mu\text{m} & W_{p2} &= 1.767\mu\text{m} \\ W_{n3} &= 2.789\mu\text{m} & W_{p3} &= 5.578\mu\text{m} \\ W_{n4} &= 8.801\mu\text{m} & W_{p4} &= 17.603\mu\text{m}. \end{aligned} \tag{12}$$

The minimum propagation time through the circuit is approximately $t_p = 95$psec.

2.2. Fixed delay specification

We are told that the acceptable propagation delay is $t_c = 200$ps, as such, the MOSFETS can be scaled larger. The simplest solution is to assume, once again, that adjacent nFETs scale (denoted by $\tilde{\Delta}$), but that the overall propagation time is no longer the minimum. Applying Equation 11 to the fixed delay, we have that

$$\tilde{\Delta} = \frac{t_c}{N t_{p0}} - 1 \approx 7.74304, \tag{13}$$

with the resulting FET widths.

$$\begin{aligned} W_{n1} &= 0.28\mu\text{m} & W_{p1} &= 0.56\mu\text{m} \\ W_{n2} &= 2.168\mu\text{m} & W_{p2} &= 4.336\mu\text{m} \\ W_{n3} &= 16.787\mu\text{m} & W_{p3} &= 33.574\mu\text{m} \\ W_{n4} &= 129.985\mu\text{m} & W_{p4} &= 259.969\mu\text{m}. \end{aligned} \tag{14}$$

The dynamic circuit power is

$$\begin{aligned} P_{\text{dyn}} &= C_L^{\text{tot}} V_{DD}^2 f_c \\ &= \left[2C_{g,1}\left(1 + \tilde{\Delta} + \tilde{\Delta}^2 + \tilde{\Delta}^3\right) + C_L\right] V_{DD}^2 f_c \\ &\approx 19.025\text{mW}. \end{aligned} \tag{15}$$

where $f_c = 1/t_c$ is the clock frequency of the driven input, compared to $P_{\text{dyn}} = 3.10765$mW. It is not surprising that the circuit power is higher for this case, as the channel width W_n is directly proportional to discharge current, which is in-turn proportional to the circuit power.

Note. I should point out now, that I am aware of the apparent contradiction that a longer propagation time results in wider nFET channels. In fact, it should be the opposite, as an increased W_n/L_n ratio will allow *more* charge carriers to flow in saturation, resulting in a shorter delay. These values seem in stark contradiction, although consistent with the first theory (perhaps a contradiction implying that scaling $\tilde{\Delta}$ won't work. Accordingly, smaller channels would imply a smaller dynamic circuit power.

2.3. Minimizing circuit power given fixed delay specification

In the following we analyze the dynamic power dissipation P_{dyn} subject to minimal inverter sizing and the

constraint of a fixed delay. Assuming this minimal sizing, the dynamic circuit power is

$$\begin{aligned} P_{\text{dyn}} &= C_L^{\text{tot}} V_{DD}^2 f_c \\ &= \left[2C_{g,1}\left(1+\Delta+\Delta^2+\Delta^3\right)+C_L\right] V_{DD}^2 f_c, \end{aligned} \quad (16)$$

where $f_c = 1/t_c$ is the clock frequency of the driven input. As a function of the power supply voltage, the sum propagation delay is given by

$$\begin{aligned} t_p &= \frac{N\left(1+\Delta\right) C_{g,1} V_{DD}/2}{\frac{W_n}{L_n}\mu_n C_{ox} F_n^{SI}\left(V_{DD}, V_{DSAT_n}\right)} \\ &= N\left(1+\left(\frac{C_L}{W_{n1}\left(\rho+1\right)\left(2C_{ov}+C_{db}\right)}\right)^{1/N}\right) \\ &\quad \cdot \frac{\left(W_{n1}\left(\rho+1\right)\left(2C_{ov}+C_{db}\right)\right) V_{DD}/2}{\frac{W_n}{L_n}\mu_n C_{ox} F_n^{SI}\left(V_{DD}, V_{DSAT_n}\right)}. \end{aligned} \quad (17)$$

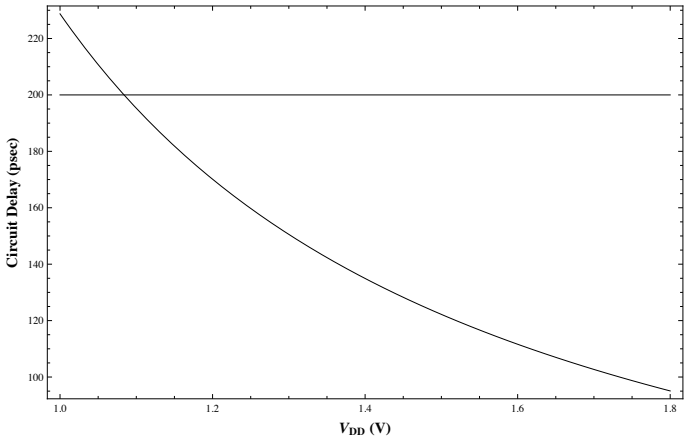
We opt to numerically solve for the power supply voltage V_{DD} that achieves $t_p = 200$psec, as shown in Figure 2.3. This constraint is achieved for

$$V_{DD_{\min}} = 1.08393\text{V} \qquad P_{\text{dyn}_{\min}} = 1.12691\text{mW}, \quad (18)$$

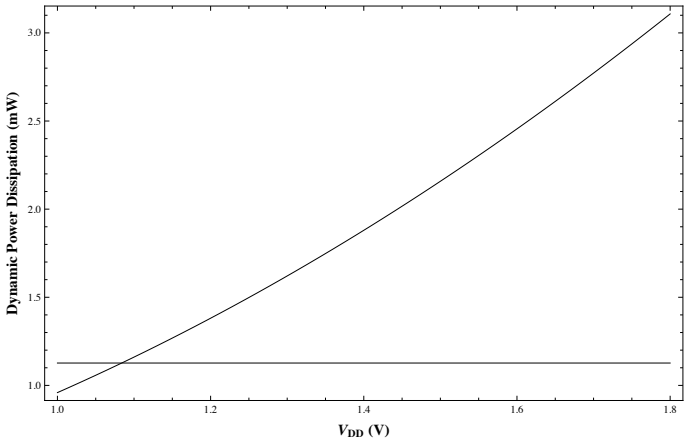
representing a significant power improvement over the previous two power considerations. Equation 17 represents some key dependencies of t_p on the relative importance of channel sizing and power supply voltage. It is particularly sensitive to power supply voltage ($t_p \propto V_{DD}^{-1}$), which decreases the effective ON resistance of a given MOSFET, thereby decreasing the propagation time across the inverter. While increasing the width-length ratio of the MOSFET will decrease the propagation time, it is worth noting that $t_p \propto W_{n1}^{-1/N}$, and therefore that as increasing channel width continuously will result in an inverter effectively loading itself, thereby increasing the overall circuit delay. Decreasing power supply voltage is not only more effective in considerations of propagation delay, is is also important for power dissipation optimization.

3. ENERGY-DELAY PRODUCT

In the following, we consider the power supply voltage in the contest of the energy-delay product (EDP), in particular, by accounting for the weak-inversion leakage current during the MOSFET's static OFF state.



(a)Propagation time. The clock period constraint is marked by the horizontal line intersecting at approximately 1.08V and 200psec.



(b)Dynamic circuit power. The achievable minimum power is marked by a horizontal line

FIG. 3: Power supply voltage effect on propagation time t_p and dynamic power P_{dyn}.

3.1. Energy-delay product (EDP) for power specification

The energy-delay product of the circuit is given by

$$\text{EDP} = \left[\underbrace{C_L^{\text{tot}} V_{DD}^2}_{\text{Dynamic}} + \underbrace{\left(\sum_{i=1}^{N=4} \bar{I}_{\text{off}}^{(i)}\right) V_{DD}\tau_{\text{period}}}_{\text{Static}}\right] \tau_{\text{period}}, \quad (19)$$

where τ_{period} is the cycle-time of the circuit as defined by the circuit specification, and the average leakage current of the ith inverter is given by

$$\bar{I}_{\text{off}}^{(i)} = \frac{1}{2}\left(I_{\text{OFF}_n}^{(i)} + I_{\text{OFF}_p}^{(i)}\right). \quad (20)$$

The nFET weak-inversion leakage current is given by

$$I^{(i)}_{\mathrm{OFF}_n} = \frac{W_{ni}}{L}\mu_n C_{ox}\frac{\gamma\phi_t^2}{2\sqrt{2\phi_{Fp}}}e^{\frac{V_{GS}-V_{Tn}}{n\phi_t}}\left[1-e^{-\frac{V_{DS}}{\phi_t}}\right]$$
$$= \frac{W_{ni}}{L}I_{0n}\exp\left(-V_{Tn}/n\phi_t\right). \quad (21)$$

The pFET leakage current $I^{(i)}_{\mathrm{OFF}_n}$ is similarly defined. Applying the power supply voltage from the power-supply scaling problem, we find that

$$V_{DD_{\min}} = 1.08393\mathrm{V} \qquad \mathrm{EDP}_{\min} = 0.0266443\mathrm{fJ}\cdot\mathrm{sec}. \quad (22)$$

3.2. Minimizing energy-delay product (EDP)

Figure 3.2 shows the total EDP of our four-inverter buffer, in addition to the EDP broken into its individual static and dynamic components. Using the minimally sized MOSFETS from the first part of this design, we find that the EDP is minimized when,

$$V_{DD_{\min}} = 1.11083\mathrm{V} \qquad \mathrm{EDP}_{\min} = 0.0266363\mathrm{fJ}\cdot\mathrm{sec}. \quad (23)$$

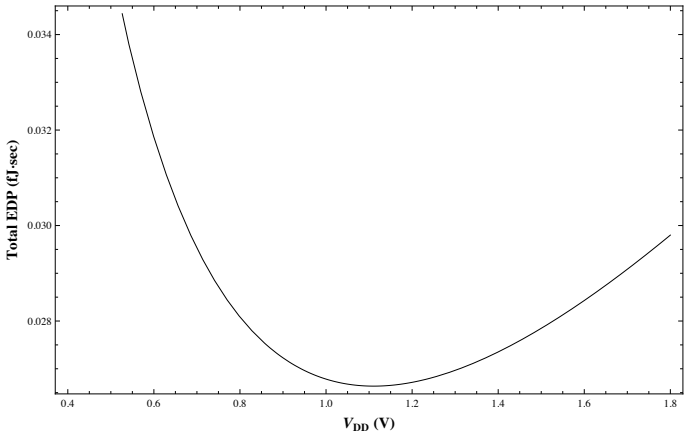
Figure 3.2 is not particularly surprising: we know that dynamic discharging and charging a capacitive load C_L is strongly dependent on the strength of the power supply voltage V_{DD}, so reducing it important. On the other hand, as the power supply voltage approaches the threshold voltage of the MOSFETS, subthreshold leakage current becomes increasingly important, and a tradeoff must be made to minimize the EDP product of the circuit.

4. SPICE SIMULATION

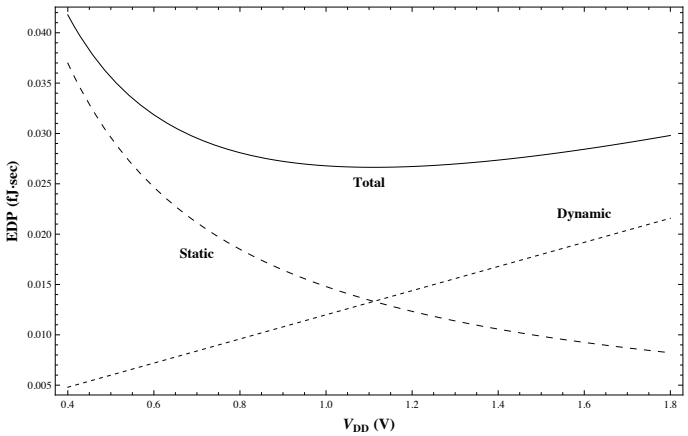
Figure 6 shows the schematic capture from our LTSpice simulation, and Figure 5 shows the results of the simulation, indicating, instead of a minimum for the EDP, a threshold voltage around 0.8V, at which point the EDP from dynamic power switching becomes increasingly important. This voltage is reasonably close to our postulated minimum around 1.1V. The magnitude of the EDP is clearly comparable to the total postulated EDP, and we can therefore conclude that subthreshold leakage current contributed less (But not nothing! Note that the EDP is flat before that threshold) to the EDP than would be expected in-theory. I posit that the MOSFET model provided in the SPICE library is ultimately less likely to generate thermal charge carriers relative to weak inversion model discussed this semester.

4.1. Netlist

The netlist code corresponding to the schematic capture in Figure 6 is listed below for your inspection.



(a)Total EDP.



(b)Total EDP and its static and dynamic components.

FIG. 4: Power supply voltage effect on EDP.

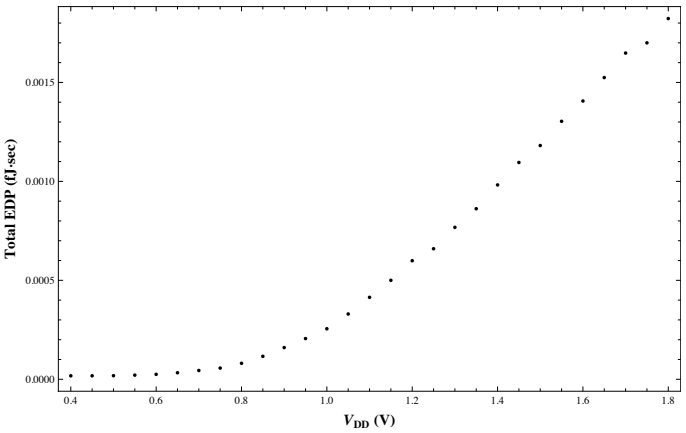


FIG. 5: Simulated, total EDP using LTSpice.

```
* C:\Users\mookerji\Desktop\6.012\design_project.asc
M1 N003 N002 0 NC_01 NMOS2 l=0.18um w=0.28um
M2 N001 N002 N003 NC_02 PMOS2 l=0.18um w=0.56um
M3 N004 N003 0 NC_03 NMOS2 l=0.18um w=0.883676um
M4 N001 N003 N004 NC_04 PMOS2 l=0.18um w=1.767352um
M5 N005 N004 0 NC_05 NMOS2 l=0.18um w=2.78887um
```

```
M6 N001 N004 N005 NC_06 PMOS2 l=0.18um w=5.7774um
M7 N006 N005 0 NC_07 NMOS2 l=0.18um w=8.80162um
M8 N001 N005 N006 NC_08 PMOS2 l=0.18um w=17.60324um
C1 N006 0 100fF
V1 N001 0 {vdd}
V2 N002 0 PULSE(0 {vdd} 0 0 0 {Tper/2} {Tper})
.model NMOS NMOS
.model PMOS PMOS
.lib C:\Program Files\LTC\LTspiceIV\lib\cmp\standard.mos
.lib 6012_180nm.lib
.options method=gear
.step param vdd 0.4 1.8 0.05
.param Tper=62.5u/vdd
*.meas tran Ptot integ -V(vdd)*I(vdd)*Tper
+ TRIG V(vdd) VAL=vdd/2 TD=5u FALL=1
+ TARG V(vdd) VAL=vdd/2 TD=5u FALL=2
.tran 0 0.5m 0
*.param vdd 1.5
.meas tran EDP integ
+( (Id(M1)+Id(M3)+Id(M5)+Id(M7))*Vdd*Tper*Tper)
.meas tran EDP1 integ (-I(V1)*Vdd*Tper*Tper)
.backanno
.end
```

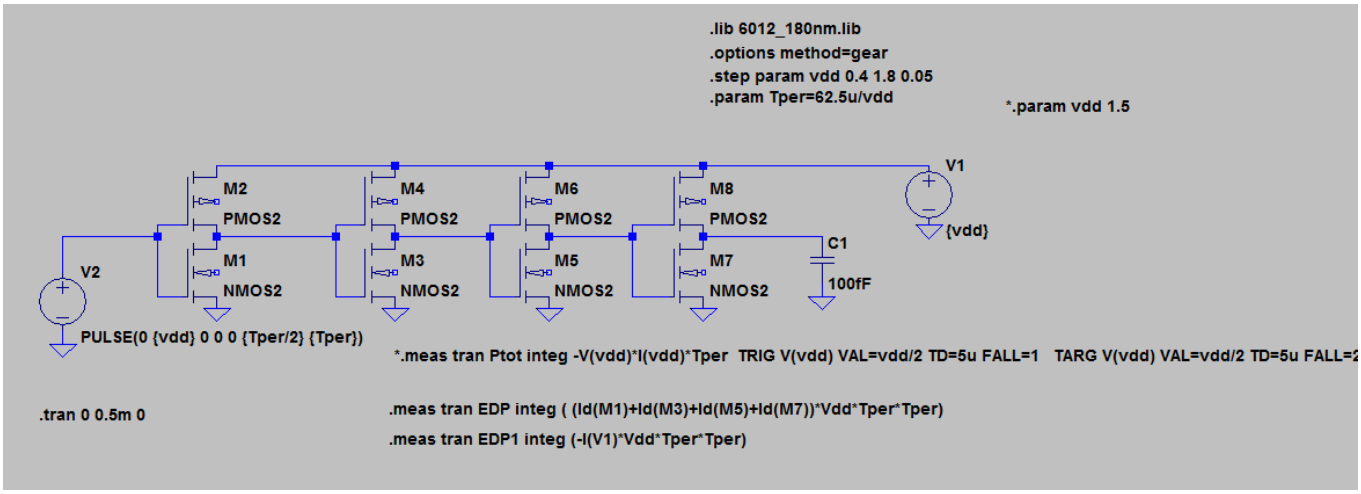


FIG. 6: Schematic capture of four-inverter buffer driving 100pF load in LTSpice.