

DICD | DIGITAL INTEGRATED CIRCUIT DESIGN

Prof. A. Boufaudi

di Pietro Giannuccaro

A.A. 2021/22



INTRODUCTION

DIGITAL REFERS TO A WAY OF CODING A SIGNAL OPOSED TO ANALOG WAY.
THE ADVANTAGES ARE:

- ROBUSTNESS TO NOISE:

THERE ARE ONLY TWO VALUES, '0' AND '1'
SO THERMAL NOISE, FLICKER NOISE, SHOT NOISE DO NOT ALTER THE
VALUE OF THE DIGITAL SIGNAL.
WE HAVE ONLY TO DEAL WITH QUANTIZATION NOISE, THAT CAN
BE ALWAYS PRECISELY PREDICTED INFLUENCING THE NUMBER OF BITS.

- EASY TO STORE DATA:

WE NEED TO STORE ONLY TWO VALUES OF VOLTAGE (0 AND V_{DD})
THROUGH FLIP-FLOPS OR SEMICONDUCTOR MEMORIES

- EASY TO PROCESS DATA:

THROUGH LOGIC GATES, ADDERS, SHIFTERS

INTEGRATED MEANS THAT CIRCUITS ARE IMPLEMENTED ON A SINGLE CHIP, ON THE SAME
SILICON SUBSTRATE, WITH ADVANTAGES IN TERMS OF AREA, DELAY AND POWER CONSUMPTION
WITH RESPECT TO A DISCRETE COMPONENT CIRCUIT.

DIGITALIZATION OF ICs

DIGITAL ICs ARE BECOMING MORE AND MORE COMPLEX, FAST
AND POWER HUNGRY THANKS TO:

ALSO NON-IDEALITIES
OF ANALOG DOMAIN
(LINEAR DISTORTION,
OFFSET, etc...)
ARE SOLVED

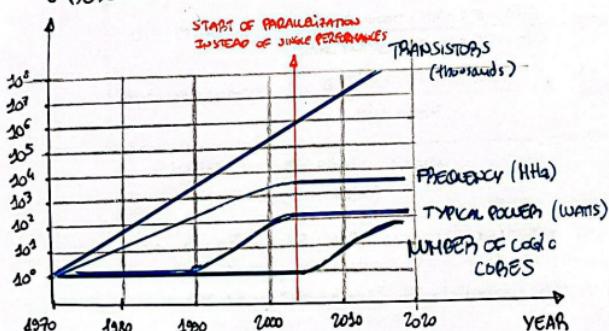
- PROCESS TECHNOLOGY SCALING

- EFFICIENT DESIGN TOOLS

PREDICTION OF DIFFERENT PARAMETERS
(MAXIMUM LENGTH, OXIDE THICKNESS, POWER SUPPLY)

- PREDICTION OF AREA PER TRANSISTOR (AND COST)
- INCREASE OF SPEED
- REDUCTION OF POWER CONSUMPTION (AT GIVEN SPEED.)

START OF PARALLELISATION
INSTEAD OF SINGLE PERFORMANCE



IN THE CMOS TECHNOLOGY
THE 'GROUND' DOES NOT
EXIST, BUT IT'S A COMBINATION
OF PASSIVE ELEMENTS
WHICH LIMIT THE PERFORMANCES
OF MODERN ARCHITECTURES

HOOPER'S LAW:

EVERY TWO YEARS THE
CHANNEL LENGTH DECREASES
BY A FACTOR $\sqrt{2}$

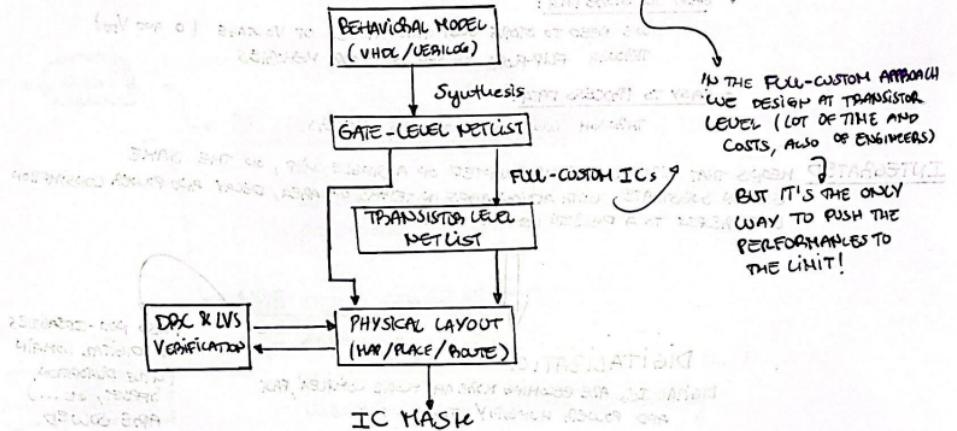
DESIGN TOOLS

SEMI-

- CUSTOM APPROACH: ASICs for which all the logic cells are PRE-DESIGNED (SEMI-CUSTOM ASICs) and AVAILABLE IN A LIBRARY, which makes the design easier and faster.

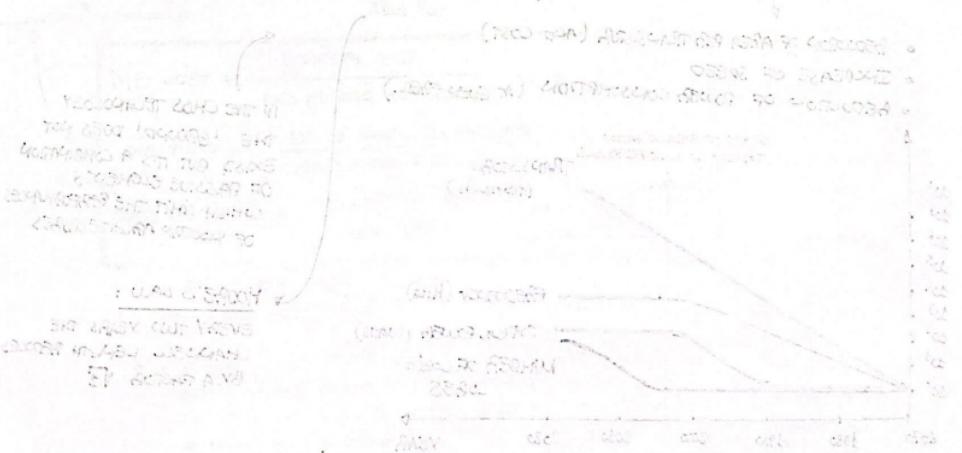
• STANDARD CELL APPROACH: DESIGN YOUR LOGIC CELLS FROM SCRATCH

• FULL-CUSTOM APPROACH: THE MOST COMMON SEMI-CUSTOM APPROACH IS THE STANDARD-CELL BASED APPROACH



NOTE THAT IN DIGITAL DESIGN WE ALWAYS USE THE MINIMUM CHANNEL LENGTH, AND ADJUST THE W (because we don't need to cope with analog constraints like noise and offset). THE MOSFETS ACT JUST AS SWITCHES. WE DON'T CARE ABOUT OTHER CONSTRAINTS

CONSTRAINTS THAT ARE TO CONSIDER
(WE'RE GOING TO EXPLAIN THEM IN THE NEXT SLIDES)



DIGITAL CIRCUITS PERFORMANCES EVALUATION

- FIGURES OF MERIT -

THE PARAMETERS TO MEASURE THE PERFORMANCES OF A DIGITAL CIRCUIT ARE:

- 1) COST (and AREA)
- 2) RELIABILITY
- 3) SPEED (PROPAGATION DELAY OR MAXIMUM OPERATING FREQUENCY)
- 4) POWER CONSUMPTION

CAST

RE COSTS (RECURRING ENGINEERING):

- proportional to volume
- silicon processing, packaging, test
- proportional to chip die area

NRE COSTS (NON RECURRING ENGINEERING):

- design time and effort
- mask generation (can cost up to \$ 2 MIL!)
- manufacturing machines and building

$$\text{DIE COST} = f(\text{DIE AREA})$$

$$IC = RE \text{ COSTS} + \frac{NRE \text{ COSTS}}{\text{PRODUCTION VOLUME}}$$

$$\begin{aligned} \text{DIE YIELD} &= 1 - \frac{\text{DEFECTS PER AREA} \times \text{DIE AREA}}{\text{WAFER DIAMETER}^2} \\ &\approx 3 \end{aligned}$$

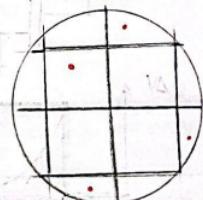
$$RE \text{ COSTS} = \text{DIE COST} + \text{PACKAGE COST} + \text{TESTING COST}$$

about 80%

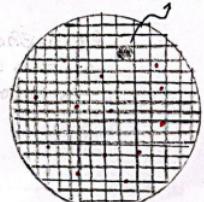
$$\text{DIE COST} = \frac{\text{WAFER COST}}{\text{DIES PER WAFER} \times \text{DIE YIELD}}$$

$$\text{DIE YIELD} = \frac{\text{NO. OF GOOD CHIPS PER WAFER}}{\text{TOTAL NUMBER OF DIES PER WAFER}} \cdot 100 [\%]$$

$$\text{DIES PER WAFER} = \frac{\pi \times (\text{WAFER DIAMETER}/2)^2}{\text{DIE AREA}} = \frac{\pi \times \text{WAFER DIAMETER}^2}{4 \times \text{DIE AREA}}$$



SINGLE DIE



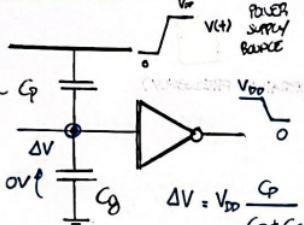
if in a chip there is a silicon defect it's invalid
AND IT CANNOT BE USED (RED DOTS). SO, THE LARGER THE
DIE AREA THE LARGER THE PROBABILITY THAT A CHIP HAS
A DEFECT IN ITS AREA.

Reliability

IT'S THE ROBUSTNESS TO NOISE, POWER SUPPLY BOURGES, AND OTHER NON IDEALITIES OR DISTURBANCES

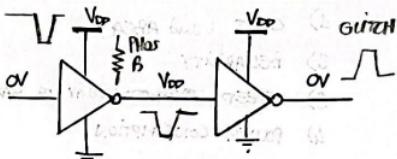
CAPACITIVE COUPLING:

PARASITIC CAPACITANCE

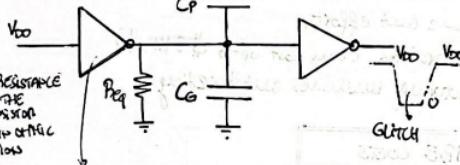


IF $C_p \gg C_g$ WE HAVE AN EPROM (LOGIC VALUES SHOULD BE V_{DD} AT THE OUTPUT)

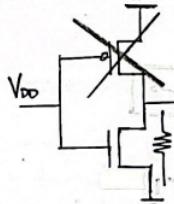
POWER AND GROUND NOISE



THE SUPPLY-VOLTAGE SOURCE IS NOT AN ISSUE FOR THE FIRST INVERTER BUT CAUSES GLITCHES FOR THE FOLLOWING GATES.

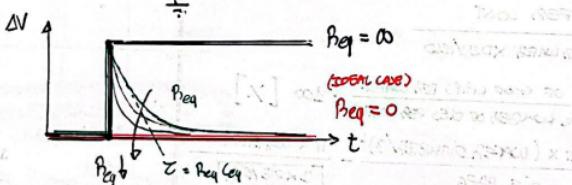


a way to practically reduce C_p (and capacitive coupling) is to push away the power supply wire, so the gap is larger



The larger the width of the inverter transistors the lower R_{load} (paying in area larger)

THE FLOATING NODE is more prone to cause logic errors than a node forced by a low-resistance driver.



this is an issue in DYNAMIC GATES (BASED ON THE STORAGE OF CHARGE ON A CAPACITOR THAT THEN IS LEFT FLOATING) THEY ARE VERY PRONE TO INTERFERENCES AND POLE SHIFTING

BY LOWERING R_{load} THE SPIKE IS SHORTER

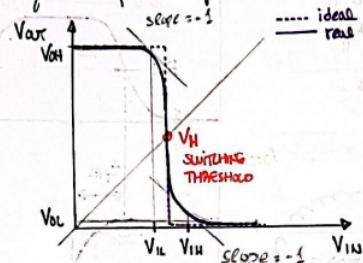
SO T IS LOWER AND THE INVERTER CASCODE HAS TIME TO MAKE THE EPROMED TOGGLE

INDEED, STATIC GATES ARE CHARACTERIZED BY A LOW-IMPEDANCE DRIVING OF THE OUTPUT VOLTAGE, WHICH IS ALWAYS CONNECTED THROUGH A LOW-IMPEDANCE PATH TO EITHER V_{DD} OR GROUND.

THE LOWER THE OUTPUT RESISTANCE OF THE DRIVER, THE BETTER; AT THE PRICE OF AREA (cost) AND POWER CONSUMPTION.

VOLTAGE TRANSFER CHARACTERISTIC (VTC)

Let's consider an inverter, which flip the input logic signal. It can be described by plotting the input voltage vs the output voltage in the VTC:



V_{IL} : MAXIMUM VOLTAGE RECOGNIZED AS '1'
 V_{IH} : MINIMUM VOLTAGE RECOGNIZED AS '0'

VALID INPUT VALUES
in between no valid logic values
(undefined region)

→ THERE ARE IMPORTANT PARAMETERS:

V_{OH} : HIGH OUTPUT VOLTAGE (corresponding to '1')

V_{OL} : LOW OUTPUT VOLTAGE (corresponding to '0')

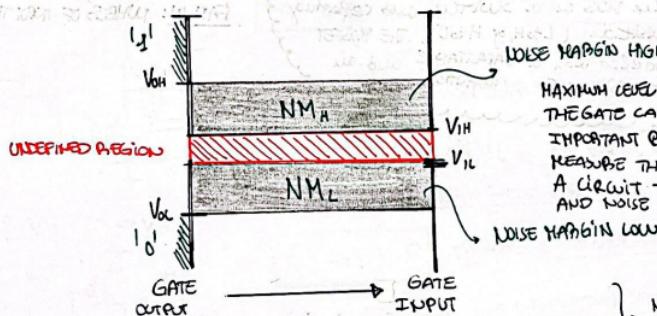
NOMINAL
OUTPUT
VALUES

CAN BE FOUND RECURSIVELY

$$V_{OH} = f(V_{OL}) \quad \left\{ \begin{array}{l} \text{the difference is the} \\ \text{SIGNAL SWING} \end{array} \right.$$

$$V_{OL} = f(V_{OH}) \quad \left\{ \begin{array}{l} \text{THRESHOLD INVERTING} \\ \text{VOLTAGE} \end{array} \right.$$

V_H : THRESHOLD INVERTING VOLTAGE
(at first approximation)



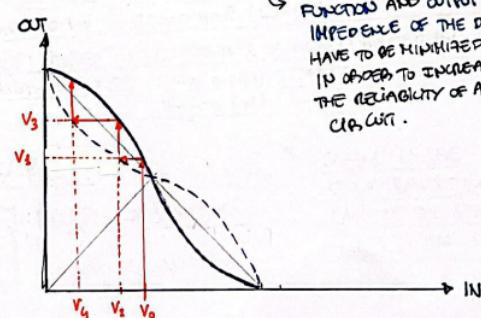
HIGH LEVEL OF NOISE THAT THE GATE CAN TOLERATE
IMPORTANT BECAUSE THEY MEASURE THE IMMUNITY OF A CIRCUIT TO INTERFERENCE AND NOISE

NOISE MARGIN LOW

NOTE THAT NOISE TRANSFER FUNCTION AND OUTPUT IMPEDANCE OF THE DRIVER HAVE TO BE MINIMIZED IN ORDER TO INCREASE THE RELIABILITY OF A CIRCUIT.

REGENERATIVE PROPERTY

WITH CASCDED INVERTERS we can reconstruct an ambiguous signal only if $|g| > 1$ where g is the gain at V_H



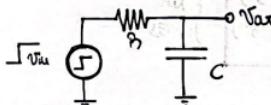
SPEED (PROPAGATION DELAY or MAX. OPERATING FREQUENCY)

- for PROPAGATION DELAY we distinguish between the two transitions t_{PHL} and t_{PLH} . To have a unique value we can take the average or the maximum value.

Others key values are falling-time and rising-time (t_F and t_R).

RC NETWORK MODEL

Since we treat the TRANSISTOR as a combination of a RESISTANCE, CAPACITANCE and SWITCH it's important to consider a simple RC networks:



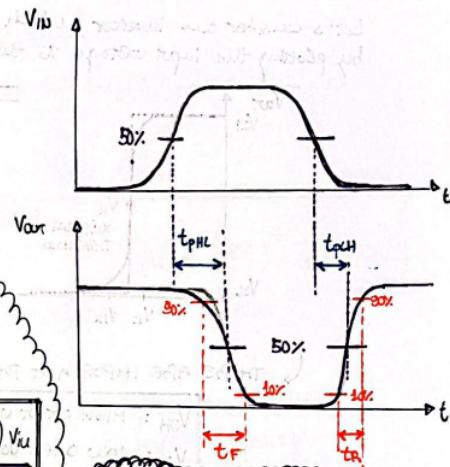
$$V_{out}(t) = (1 - e^{-t/\tau}) V_{in}$$

$$\tau = R \cdot C = 0.63 \cdot RC$$

In ANALOG DOMAIN, we model a MOS by linearizing its I-V characteristic around the BIAS POINT. It acts as a current generator.

In DIGITAL DOMAIN we deal with LARGE SIGNALS, so we don't care about the bias point.

We model the MOS as a SWITCH and referring to the transition (LtoH or HtoL). The MOSFET can be modeled with a CAPACITANCE and an EQUIVALENT RESISTANCE (PARASITIC)



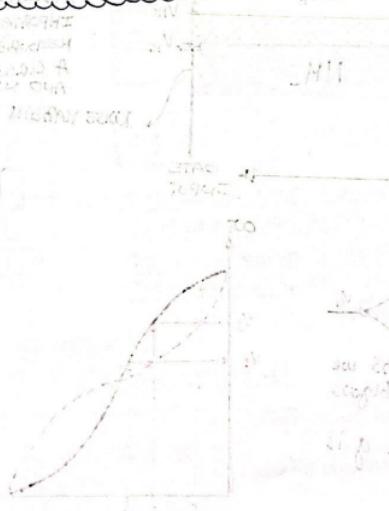
FAN-OUT and FAN-IN

FAN OUT: NUMBER OF PADS CONNECTED TO THE OUTPUT

$$f = \frac{C_{LOAD}}{C_{GATE}}$$

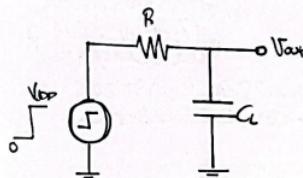
FAN IN: NUMBER OF INPUT TERMINALS

MINIMUM FAN-IN AND FAN-OUT
DEPENDING ON THE NUMBER OF
TRANSISTORS OR NUMBER OF
INPUTS OR NUMBER OF
OUTPUTS OR NUMBER OF
POWER SUPPLY VOLTAGES



WE DON'T WANT THIS, BUT SOMETIMES
WE USE IT DUE TO LOGIC DESIGN
OR LOGIC STYLING.
IF f = 2 & f = 1, f = 2
MUST BE USED

POWER CONSUMPTION



• INSTANTANEOUS POWER:

$$P(t) = v(t)i(t) = V_{\text{SUPPLY}}i(t)$$

• PEAK POWER:

$$P_{\text{PEAK}} = V_{\text{SUPPLY}} i_{\text{PEAK}}$$

• AVERAGE POWER:

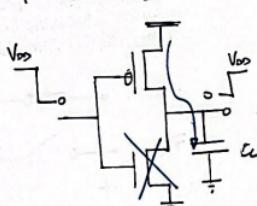
$$P_{\text{Ave}} = \frac{1}{T} \int_t^{t+T} P(t) dt = \frac{V_{\text{SUPPLY}}}{T} \int_t^{t+T} i_{\text{current}}(t) dt$$

SETS THE DURATION OF THE BATTERY

We can distinguish between:

- STATIC POWER CONSUMPTION: due to current that flows from positive rail to negative rail (due to a leakage)
- DYNAMIC POWER CONSUMPTION: associated to commutations, and it's the main contribution to the overall power.
IT'S DUE TO THE CURRENT FLOWING INTO THE OUTPUT CAPACITANCE OF THE GATES FROM THE SUPPLY.

$$\text{if } V_{\text{IN}} = V_{\text{DD}} \rightarrow 0$$



$$E_0 \rightarrow 1 = \int_0^{\infty} P(t) dt = V_{\text{DD}} \int_0^{\infty} i_{\text{current}}(t) dt = V_{\text{DD}} \int_0^{\infty} C_L \frac{dV_{\text{out}}}{dt} dt = \\ = V_{\text{DD}} \int_0^{V_{\text{DD}}} C_L dV_{\text{out}} = \frac{1}{2} C_L V_{\text{DD}}^2$$

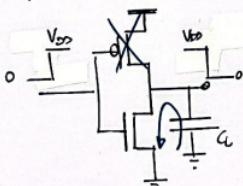
ENERGY FOR THE TRANSITION WITH PHOSPHOR half

CURRENT FLOW FROM SUPPLY TO CAPACITANCE

$$E_C = \int_0^{\infty} V_{\text{out}} i(t) dt = \int_0^{\infty} C_L V_{\text{out}} \frac{dV_{\text{out}}}{dt} dt = \frac{1}{2} C_L V_{\text{DD}}^2$$

ENERGY STORED IN THE CAPACITANCE

$$\text{if } V_{\text{IN}} = 0 \rightarrow V_{\text{DD}}$$



$$E_1 \rightarrow 0 = 0$$

DISCHARGE OF THE CAPACITANCE

THE OTHER HALF (OF THE ENERGY SPENT BY THE OPERATION)
IS DISSIPATED BY THE RESISTANCE (WHATEVER ITS VALUE) AS HEAT
THE HALF STORED IN THE CAPACITANCE IS ALSO DISSIPATED BY
THE RESISTANCE UP TO THE $0 \rightarrow V_{\text{DD}}$ (IN) TRANSITION
 $V_{\text{DD}} \rightarrow 0$ (OUT)

SO WHICH IS THE POWER CONSUMPTION?

IT'S PROPORTIONAL TO THE ROLL-UP FREQUENCY $f_{0 \rightarrow 1}$!

$$P = C_L V_{\text{DD}}^2 f_{0 \rightarrow 1}$$

POWER CONSUMPTION
(wasteful dynamic)

$E_{0 \rightarrow 1}$

$E_{0 \rightarrow 1}$ IS THE
ONLY TRANSITION
WHERE THE SUPPLY
(V_{DD}) IS INVOLVED!

the MOS TRANSISTOR

HOS vs BJT

MOS TRANSISTORS have completely overwhelmed the BJT since Federico Faggin introduced the SELF-ALIGNED GATE TECHNOLOGY in 1968

HOS

- 1) BETTER SWITCHES
- 2) GOOD COMPLEMENTARY DEVICES (PMOS AND NMOS)
- 3) NO STATIC POWER CONSUMPTION
- 4) LARGE INTEGRATION DENSITY
- 5) SIMPLER PROCESS

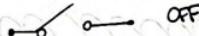
BJT

- 1) LARGE CURRENT DRIVE CAPABILITY
- 2) FAST DEVICES
- 3) LARGE GATE AND OUTPUT RESISTANCE
- 4) STATIC POWER CONSUMPTION

as switches: we are not interested in large transconductance and a large output resistance (HOS as current generator)
 we are interested in low impedance when it's ON (ideally short circuit) and high impedance when it's OFF (ideally open circuit)



ON



OFF

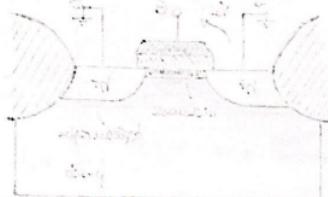
\rightarrow I_{eq}

there is NO POSSIBILITY TO IMPLEMENT COMPLEMENTARY DEVICES, AS PNP FEATURES BAD PERFORMANCE IN TERMS OF CUT-OFF FREQUENCY AND SO ON.
 on the other side NMOS and PMOS can be both good.

HOS transistors are good switches since they can be easily be turned ON and OFF. BJT instead are bad because biasing a bipolar device in DEEP ACCUMULATION REGION (TO HAVE A CLOSED SWITCH CAN USE THE DIODES ACROSS THE COLLECTOR-Emitter JUNCTION)

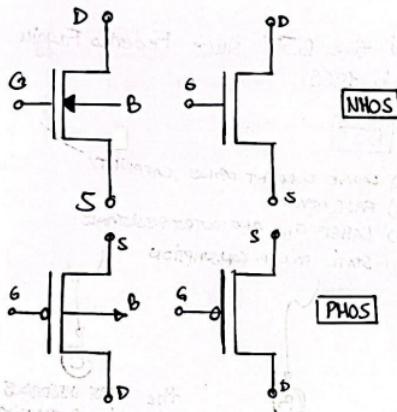
makes the BJT a slow device because of the charge that has to be removed from the BASE.

the INPUT IMPEDANCE OF A BIPOLAR TRANSISTOR IS FINITE, DUE TO THE BASE CURRENT REQUIRED BY A BJT. This means a static current consumption because of the current sink from the driving gate. This is in contrast to HOS design where a GATE PRESENTS ONLY A CAPACITIVE LOAD AT THE INPUT.



base
emitter
collector

HOS STRUCTURE



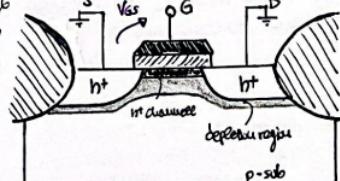
STATIC MODEL

CUT-OFF REGION

for $V_{GS} < V_t$

THERE IS NO FREE CHARGE BELOW THE THIN GATE OXIDE AND THIS NO CURRENT FLOWS. CORRESPONDS TO AN OPEN CIRCUIT $I_{DS} = 0$

LINEAR (OHMIC) REGION



$V_{GS} > V_t$ (mos on)

$V_D > V_t \Rightarrow V_{DS} < V_{Dr}$

THRESHOLD VOLTAGE

$$V_t = \Phi_{GC} - 2\Phi_F - \frac{Q_B}{C_{ox}} - \frac{Q_{ox}}{C_{ox}}$$

Φ_{GC} : work function difference between gate and channel

Φ_F : fermi voltage (-0.3V)

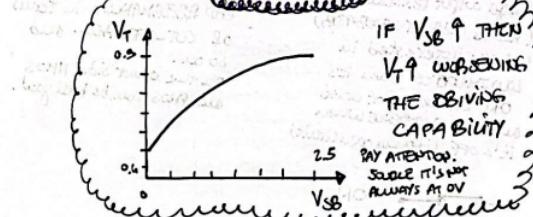
Q_B : depletion layer charge

Q_{ox} : electron charge trapped in the oxide

if a V_{SB} (back-gate) is applied:

$$V_t = V_{t0} + \eta [\sqrt{1-2\Phi_F + V_{SB}} - \sqrt{1-2\Phi_F}]$$

BODY EFFECT



$$I_{DS} = Q' W U \xrightarrow{\text{carrier velocity}} \int_0^L I_{DS} dx = \int_0^L Q'(x) W U(x) dx$$

$$\left\{ \begin{array}{l} Q'(x) = C_{ox} [V_{GS} - V_t - V_c(x)] \\ V_c(x) = \mu E(x) = \mu \frac{dV(x)}{dx} \end{array} \right. \rightarrow I_{DS} = W \mu C_{ox} \int_0^{V_{DS}} [V_{GS} - V_t - V_c(x)] dx$$

$$\hookrightarrow I_{DS} = (\mu C_{ox} \frac{W}{L}) (V_{GS} - V_t) V_{DS} - \frac{V_{DS}^2}{2} \quad \text{OHMIC CURRENT}$$

THE CHANNEL IS INVERTED ALSO AT THE DRAIN SIDE
GAIN FACTOR

PINCH-OFF SATURATION REGION

$$I_{DS} = \frac{1}{2} \mu C_{ox} \left(\frac{W}{L} \right) [V_{GS} - V_T]^2$$

can be obtained from:

A) drain current with $V_{DS} = V_{ov}$

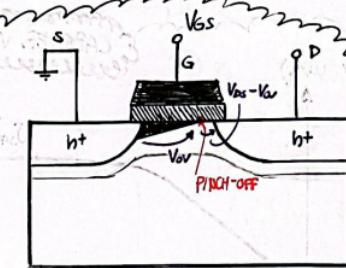
B) AVERAGE CHARGE + AVERAGE VELOCITY

$$\bar{Q} = \frac{C_{ox} (V_{GS} - V_T)}{2}$$

HEAD VALUE
WITH $C_{ox} V_{GS}$ AND $\mu L (D)$

$$\bar{v} = \mu \bar{E} = \mu \frac{(V_{GS} - V_T)}{L}$$

$$I_{DS} = \bar{Q} W \bar{v}$$

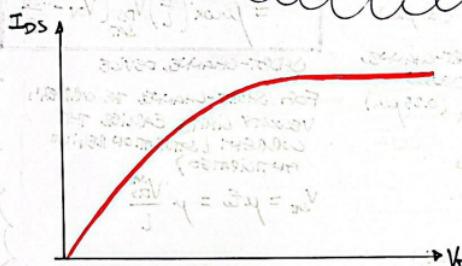


$$V_{GS} > V_T \text{ (MOS ON)}$$

$$V_{DS} > V_{ov} \text{ or } V_{DS} < V_T$$

- THE VOLTAGE ACROSS THE CHANNEL REMAINS THE SAME REGARDLESS OF V_{DS}

- BOTH OVERALL CHARGE IN THE CHANNEL AND LATERAL ELECTRIC FIELD ARE PROPORTIONAL TO V_{ov}



SINCE THE CURRENT IS PROPORTIONAL TO THE CHARGE AND TO THE VELOCITY OF THE CARRIERS WE OBTAIN A SQUARE-DEPENDENCE ON THE OVERDRIVE FOR SATURATION.

IN OTHER REGIONS INSTEAD, THE VOLTAGE ACROSS THE CHANNEL, THUS THE ELECTRIC FIELD AND THE CARRIER VELOCITY, DEPENDS ON THE DRAIN-SOURCE VOLTAGE, ONLY THE CHANNEL CHARGE DEPENDS ON V_{ov} , THAT'S WHY WE HAVE A LINEAR DEPENDENCE ON V_{ov}

CHANNEL LENGTH MODULATION

CHANNEL LENGTH MODULATION

ACTUALLY, THE I_{DS} (UP TO SATURATION) DEPENDS AT 2 ORDER FROM V_{DS} . THIS IS CALLED CHANNEL-MODULATION EFFECT OR EARLY EFFECT.

BY INCREASING V_{DS} WE OBTAIN A WIDTH DEPLETION REGION AT THE DRAIN JUNCTION REDUCING THE EFFECTIVE CHANNEL LENGTH $L' < L$. SO THE I_{DS} SHOULD BE LOWERED.

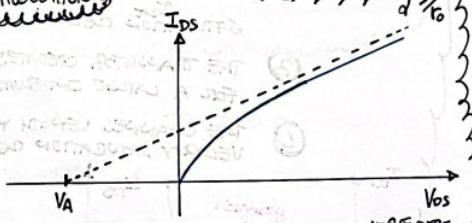
$$I_{DS} = I_{DS_0} (1 + 2V_{DS})$$

$$\lambda = 1/V_A$$

CHANNEL MODULATION EFFECT

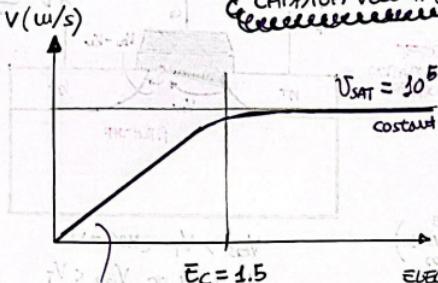
$$V_A d L$$

IF $L' \downarrow$ $V_{ds} \uparrow$ $L' \downarrow$ $V_{ds} \downarrow$ CHANNEL MODULATION EFFECT
DEPLETION REGION AREA

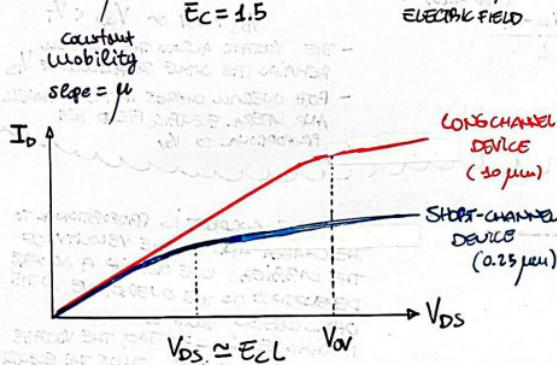


IF WE NEED HIGH OUTPUT IMPEDANCE IT'S BETTER TO USE LONG-CHANNEL TRANSISTORS

INSTEAD OF ANALOG ELECTRONICS WHERE WE NEED HIGH OUTPUT IMPEDANCES FOR GENERATORS; IN DIGITAL ELECTRONICS WE WANT SWITCHES SO WE DON'T CARE



THE EQUATIONS, ALREADY SHOWN ARE NOT PROPERLY CORRECT FOR SHORT CHANNELS, BECAUSE A VELOCITY SATURATION OF THE CARRIER'S HAPPEN.



$$I_{DS} \approx V_{SAT} C_o k' W \left(V_{DS} - V_T - \frac{V_{DS}}{2} \right)$$

$$= \mu C_o k' \left(\frac{W}{L} \right) V_{DS} \left(V_{DS} - \frac{V_{DSAT}}{2} \right)$$

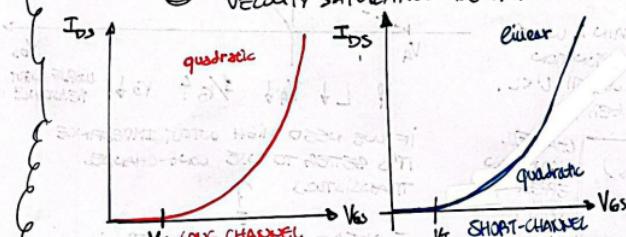
SHORT-CHANNEL DEVICE

FOR SHORT-CHANNEL THE CARRIER'S VELOCITY LIMITS EARLIER THE CURRENT (SATURATION REGION ANTICIPATED)

$$V_{SAT} = \mu E_c = \mu \frac{V_{DS}}{L}$$

SO, WHERE VELOCITY SATURATION TAKES PLACE :

- ① THE I_{DSAT} of V_{DS} NO MORE DEPENDENCE ON ELECTRIC FIELD WHICH IS PROPORTIONAL TO VOLTAGE ACROSS THE CHANNEL ΔV . THERE IS NO DEPENDENCE ON L TOO
- ② THE I_{DSAT} IS SMALLER THAN "SQUARE LAW MODEL IN THE VELOCITY SATURATION REGION".
- ③ THE TRANSISTOR OPERATES WITH A CURRENT APPROXIMATELY CONSTANT FOR A LARGE INTERVAL OF V_{DS} .
- ④ THE CHANNEL LENGTH REGULATION EFFECT IS ALSO PRESENT IN THE VELOCITY SATURATION REGION



⑤ LINEAR DEPENDENCE OF I_{DS} ON V_{DS} IN SATURATION REGION FOR SHORT CHANNEL DEVICE (INSTEAD OF QUADRATIC OF LONG CHANNEL)

- ⑥ THE SMALLER CURRENT IN THE SHORT-CHANNEL DEVICE DUE TO THE VELOCITY SATURATION
- ⑦ LARGER SATURATION REGION FOR SHORT-CHANNEL DEVICE DUE TO VELOCITY SATURATION

If $V_{DS} > V_{DSAT}$ there is still a small quadratic region because the voltage across the channel is not sufficient for saturating the carrier velocity so for $V_{DS} > V_{DSAT}$ there is the pinch-off but no carrier-velocity saturation

the UNIFIED Model of MOS

① CUT-OFF

$$V_{GS} - V_T \leq 0$$

② MOSFET ON

$$V_{GS} > V_T$$

③ LINEAR or OHMIC REGION

$$V_{GS} < V_{DS}$$

$$V_{DS} < V_{DS, \text{sat}}$$

(V_{DS} is the lowest)

④ PINCH-OFF SATURATION

$$V_{DS} > V_{DS, \text{sat}}$$

$$V_{DS, \text{sat}} > V_{DS}$$

$$V_{DS, \text{sat}} > V_{DS, \text{sat}}$$

(V_{DS} is the lowest)

④(b) VELOCITY SATURATION

$$V_{DS} > V_{DS, \text{sat}}$$

$$V_{DS, \text{sat}} > V_{DS}$$

(V_{DS} is the lowest)

So, in other words:

$$V_{GS} < V_T \rightarrow \text{MOSFET OFF}, I_{DS} \approx 0$$

$$V_{GS} > V_T \rightarrow \text{MOSFET ON}$$

NHOS

$I_{DS} = \mu C_{ox} \left(\frac{W}{L} \right) V_{GS} (V_{GS} - V_T - \frac{V_{HIN}}{2}) (1 + 2V_{DS})$

$$V_{HIN} = \min(V_{GS}, V_{DS}, V_{DS, \text{sat}})$$

PINCH-OFF OHMIC VELOCITY
SATURATION SATURATION SATURATION

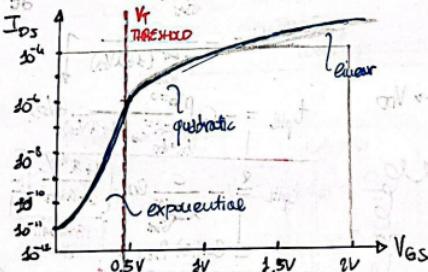
FOR PHOS, IF WE USE ABSOLUTE VALUES nothing REALLY CHANGES:

$$I_{DS} = \mu C_{ox} \left(\frac{W}{L} \right) \left[(|V_{GS}| - |V_T|) V_{HIN} - \frac{|V_{HIN}|^2}{2} \right] (1 + 2|V_{DS}|)$$

$$V_{HIN} = \min(|V_{GS}| - |V_T|, |V_{DS}|, |V_{DS, \text{sat}}|)$$

PHOS

for $V_{GS} \leq V_T$, there is still a small leakage current



THE CURRENT DECAYS EXPONENTIALLY SUBTHRESHOLD

$$I_{DS} = I_s \exp \left[\frac{V_{GS}}{nV_t} \right] \left[1 - \exp \left(-\frac{V_{DS}}{U_t} \right) \right]$$

where:

I_s : reference current

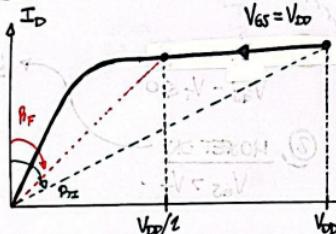
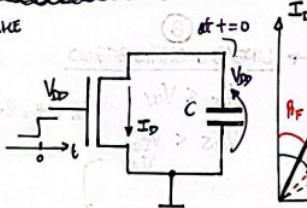
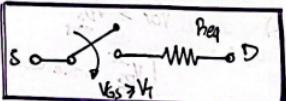
U_t : thermal voltage = 25.8 mV at $T=300K$

n : factor ≈ 3.5 ($1 + C_0/C_{ox}$)

THE SLOPE OF THE EXPONENTIAL IS VERY IMPORTANT AS SINCE IT STATES HOW MUCH WE SHOULD REDUCE V_{GS} TO REDUCE THE LEAKAGE CURRENT TO A POWER-CONSUMPTION ACCEPTABLE LEVEL.

$$S = n U_t \ln 10 = 60 \mu V/\text{dec}$$

A MOS TRANSISTOR BEHAVES LIKE A SWITCH IN DIGITAL CIRCUITS



THE MAIN PROBLEM IS THAT P_{eq} IS TIME-VARIANT, NON-LINEAR AND DEPENDS ON OPERATION POINT OF THE TRANSISTOR.

We will use a constant linear resistance

$$R_{eq} = \frac{V_{DD}}{I_{DS}}$$

OPERATION REGIONS

IF V_{DD} IS SUFFICIENTLY LARGE THE TRANSISTOR WORKS ALWAYS IN VELOCITY SATURATION REGION
(Both $-V_{DS}$ and $V_F > V_{DS, sat}$)

C has an initial voltage of V_{DD} . As the step is applied the current discharges the capacitor and we move on the characteristic towards the origin.

As we are in digital domain we are interested in the transient to reach V_H of an inverter, which is most likely $V_{DD}/2$.

TO FIND R_{eq} WE MAKE THE AVERAGE

$$R_{eq} = \frac{R_F + R_{FS}}{2} = \frac{1}{2} \left(\frac{V_{DD}}{I_{DS}} + \frac{V_{DD}/2}{I_{DS, sat}} \right)$$

HOP EQUIVALENT RESISTANCE FOR HALF V_{DD} TRANSIENT

$$R_{eq} = \frac{1}{2} \left[\frac{V_{DD}}{I_{DS, sat}(1+AV_{DD})} + \frac{V_{DD}/2}{I_{DS, sat}(1+AV_{DD}/2)} \right] = \frac{3}{4} \frac{V_{DD}}{I_{DS, sat}^2} \left(1 - \frac{5}{6} \frac{AV_{DD}}{V_{DD}} \right)$$

R_{eq} in Velocity saturation region

- to get R_{eq} we need $\left(\frac{W}{L}\right) \uparrow$ and since $L = \text{fixed}$, $W \uparrow$
- for $V_{DD} \gg V_T + V_{DS, sat}/2$, R_{eq} becomes independent of V_{DD}
- once $V_{DD} \approx V_{TE} = 0.745V$ we observe a huge increase of R_{eq}

WE CAN EVALUATE THE PROPAGATION TIME:

the MOS current is equal to capacitance current

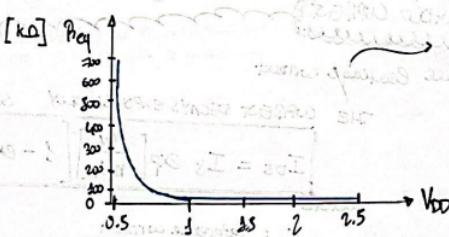
$$I_{DS} = I_{DS, sat}^2 \frac{(1+AV_{DD})}{V_{DS, sat}} = -C \frac{dV_{DS}}{dt}$$

$$dt = -\frac{C}{I_{DS, sat}^2 (1+AV_{DD})} dV_{DS}$$

$$t_{pHL} = -C \frac{\frac{V_{DD}}{2} - V_{DS, sat}}{I_{DS, sat}^2 (1+AV_{DD})} dV_{DS}$$

$$t_{pHL} = -\frac{C}{2I_{DS, sat}^2} \ln \left(\frac{1+AV_{DD}/2}{1+AV_{DD}} \right) \approx \ln(2) R_{eq} C$$

PROPAGATION TIME



HOSFET CAPACITANCES

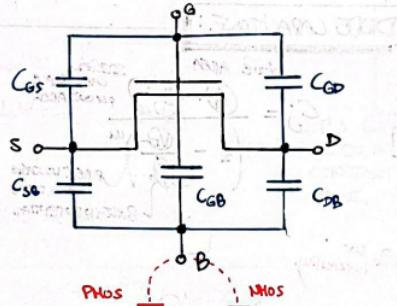
THREE TYPES OF CAPACITANCES

① GEOMETRICAL CAPACITANCES, OVERLAP CAPACITANCES

② CHANNEL CAPACITANCES, INTRINSIC CAPACITANCES

(depends on the operating region channel shape)
CUT-OFF, OHMIC, SATURATION

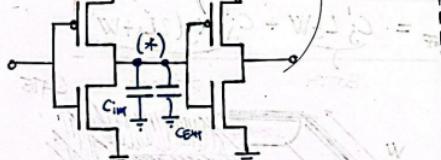
③ SINTERING CAPACITANCES (of drain diode junction, in inverse)



OK, BUT WHAT DO WE REALLY WANT TO KNOW?
WE ARE INTERESTED IN THE BEHAVIOR
OF LOGIC GATES ABOUT TIMING

$$T = f_{\text{in}}(2) \cdot \text{Perf}_C$$

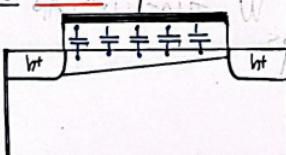
WHAT IS THIS?



We are interested in the capacitance at the node (*) which is made of two contributions:

- all INTRINSIC CAPACITANCE \rightarrow DRAIN CAPACITANCES
- all EXTERNAL CAPACITANCE \rightarrow GATE CAPACITANCES

B2: LINEAR



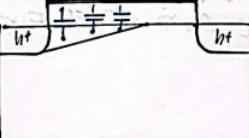
$$C_{gb} = 0$$

$$C_{gs} = \frac{C_{ox}}{2}$$

$$C_{gd} = \frac{C_{ox}}{2}$$

LINEAR

B3: SATURATION



$$C_{gb} = 0$$

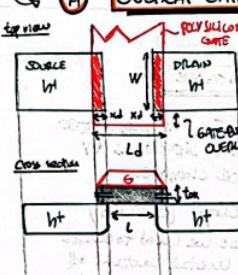
$$C_{gs} = \frac{2}{3} C_{ox}$$

$$C_{gd} = 0$$

SATURATION

The electrons does not come anymore from bulk depletion region but from drain and source. The capacitance is also with drain and source

A OVERLAP CAPACITANCES



$$C_{ov} = C_{ox} \frac{1}{W} \cdot X_d$$

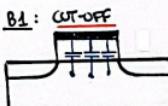
OVERLAP CAPACITANCE

$$C_{ov} = \frac{C_{ox}}{t_{ox}}$$

$C_{ox} = 0.33 \text{ pF/mm}^2$

$t_{ox} = 6 \text{ nm}$

B CHANNEL CAPACITANCES



$$C_{gb} = C_{ox} \parallel C_{gb}$$

$$C_{gs} = 0$$

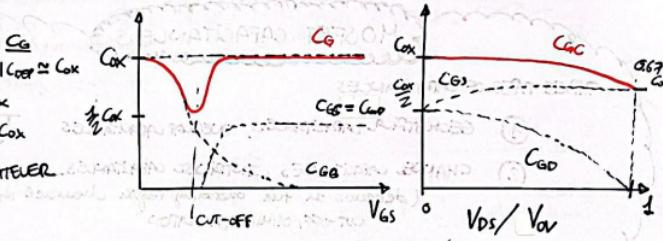
$$C_{gd} = 0$$

CUT-OFF

THEN FOR GATE CAPACITANCES

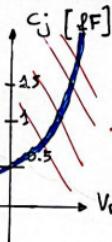
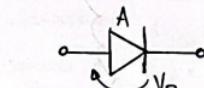
$$\begin{array}{l} \text{CUT-OFF} \rightarrow C_{ox} \parallel C_{ox} \approx C_{ox} \\ \text{OHMIC} \rightarrow C_{ox} \\ \text{SATURATION} \rightarrow \frac{1}{2} C_{ox} \end{array}$$

APPROXIMATE TO C_{ox} ! WHATEVER THE CONDITION.



(around 10% of what's going on) (DEGREE OF SATURATION)

DIODE CAPACITANCE



we are interested in the REVERSE BIAS REGION (fortunately)

DIODE AREA

$$C_d = \frac{A}{C_s}$$

$$(1 - \frac{V_d}{V_b})^m$$

specific
capacitance
per unit area

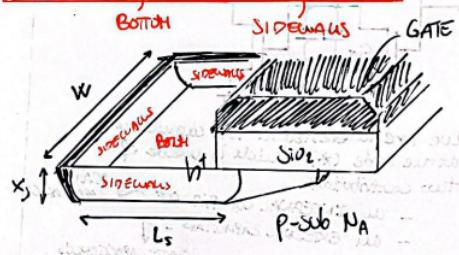
reverse
bias
across diode
built-in potential

DIFFUSION CAPACITANCE

SOURCE N WELL IS SURROUNDED ON BOTTOM AND SIDEWALLS BY P-SUBSTRATE OR CHANNEL

L_s very large ($\approx 3L_{\text{hub}}$) because we need to create the metal contact of source/drain which occupies area

$$C_{diff} = C_j' L_s W + C_j' X_j (2L_s + W)$$



NOTE: ALL CAPACITANCES DEPENDS ON WIDTH W , THAT WE CAN CHANGE AND OTHER FIXED PARAMETERS.

EQUIVALENT CAPACITANCE ALONG TRANSIENT

WE DEAL WITH LARGE VOLTAGE VARIATIONS!

$$V_{DD} \rightarrow V_{DD}/2$$

$$0 \rightarrow V_{DD}/2$$

THE CAPACITANCES CAN CHANGE!!
(DURING THE TRANSIENT)

$$C_{EQ} = \frac{\Delta Q}{\Delta V} = \frac{Q(V_2) - Q(V_1)}{V_2 - V_1} = \\ = \frac{1}{V_2 - V_1} \int_{V_1}^{V_2} C(V) dV \quad \text{very complex!}$$

HOS CAPACITANCE MODEL

GATE-CHANNEL: $C_{GC} \approx C_{ox} WL$

GATE-OVERLAP:

$$C_{GSW} = C_{ox}^f W$$



JUNCTION CAP:

$$C_{jj} = C_j^f L_s W + C_{jSW}^f (2L_s + W)$$

We are interested in GATE CAPACITANCE
and DRAIN CAPACITANCE:

$$C_G \approx C_{ox} + 2C_{ov} = \frac{E_{ox}}{t_{ox}} (L + 2x_s) W$$

$$C_{GC} \quad C_{GSW} + C_{GOW}$$

$$C_D \approx C_{ov} + C_j^f L_s W + C_{jSW}^f (2x_s + W)$$

Cox/WL

BOTTOM plate SIDEWALL plate

SIMPLIFIED APPROACH

$$C_{EQ} \approx \frac{C_G + C_D}{2}$$

LET'S CONSIDER THE WORST CASE (AS WE DO IN DIGITAL ELECTRONICS), SO, CONSTANT CAPACITANCES AT THEIR MAXIMUM VALUE.

PER UNIT WIDTH: GATE AND DRAIN CAPACITANCES PER UNIT WIDTH

$$C_G^f \approx \frac{E_{ox}}{t_{ox}} (L + 2x_s)$$

$$C_D^f \approx \frac{E_{ox}}{t_{ox}} x_s + C_j^f L_s + C_{jSW}^f$$

ARE TECHNOLOGY INDEPENDENT
(L, x_s, t_{ox} SCALE WITH TECHNOLOGY)
HOWEVER THE CAPACITANCES ARE
SMALLER BECAUSE WE CAN SCALE
DOWN THE WIDTH!!

INVERTER SIZING

Increasing the size makes the delay subside:

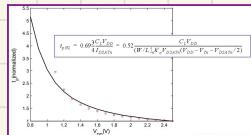
$$t_p = t_{po} (1 + \frac{f}{f_p})$$

$$f = C_{ox} / (s C_0^{(1)})$$

$$s = C_0^{(1)} - f_p \Rightarrow s > f_p$$



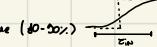
↳ some guidelines for performances:



- * KEEP CAPACITANCES SMALL
- * INCREASE TRANSISTOR SIZES
- * INCREASE VDD (f_p and t_{po})

↳ let's assess the **IMPACT OF RISE-TIME ON DELAY** (since we have considered ideal steps so far):

o if t_{po} is the rise time (10-90%)



o we call t_{step} the delay with ideal input steps of i -th inverter of the chain:

$$t_p^{(i)} = t_{step}^{(i)} + \frac{1}{f_p} (i-1) \quad i \leq 0.25$$

PUSH-IN DELAY WITH RISE-TIME

$$t_p \approx t_{step} + t_{po} + t_{ps}$$

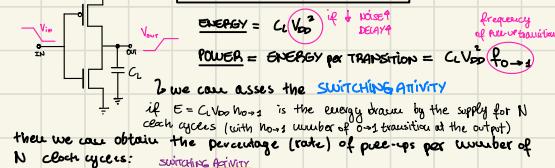


POWER DISSIPATION

There are 3 types of power consumption in a CMOS INVERTER:

- (I) **DYNAMIC POWER CONSUMPTION**: charge and discharge of capacitances
- (II) **CROSS-CONDUCTION CURRENT**: short-circuit path between supply rails during switching.
- (III) **LEAKAGE**: leaching diodes and transistors

DYNAMIC POWER CONSUMPTION



SWITCHING ACTIVITY

if $E = C_L V_{DD}^2$ is the energy stored by the supply for N clock cycles (with N as number of 0-to-1 transitions at the output)

then we can obtain the percentage (rate) of power diss per number of N clock cycles:

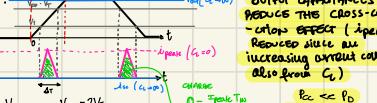
$$d_{sw} = \lim_{N \rightarrow \infty} \left(\frac{E_{0 \rightarrow 1}}{N} \right) = P(0)P(1) \quad \text{eqivalent square = 0.5} \cdot 0.5 = 1/4$$

$\rightarrow 1/4 \text{ cases}$

$$P = C_L V_{DD}^2 d_{sw} \cdot f_{CK}$$

CROSS-CONDUCTION POWER CONSUMPTION

due to a short circuit current when both transistors are ON:



$$\Delta T = T_{on} = V_{DD} - V_{TH} \approx 0.68 T_{on}$$

$$I_{peak} = V_{DD} \cdot \frac{W}{L} \cdot V_{DD} \cdot \left(\frac{V_{DD}}{2} - \frac{V_{TH}}{2} \right)^2 \approx 33 \mu\text{A} \cdot \frac{(W/L)}{2}$$

$$E = V_{DD} Q = V_{DD} \frac{\Delta T}{2} \approx 23 \mu\text{J} \cdot \frac{(W/L)}{2} \rightarrow P_{cc} = E_{cc} \frac{f_o}{2}$$

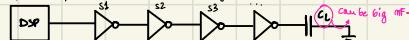
TECHNOLOGY SCALING

We want to obtain:

- more function per chip (transistors) for same money
- build same product cheaper
- reduce transistors price
- we want to be faster, smaller, with less power consumption

INVERTER CHAIN

the MOSFET question is: HOW TO OBTAIN THE BEST DELAY FOR A TRANSMISSION? and so, HOW MANY INVERTERS? HOW TO SIZE THEM? HOW TO DEAL WITH TRANSMISSION LINE DELAY? (for now, we will postpone the discussion about layout of transmission lines)



the first thing we may think is to INCREASE s to make $t_p \rightarrow t_{po}$. HOWEVER, this does not happen since the DSP delay now INCREASE because it has to drive a (much) bigger load.

$$t_{p,j} = 0.69 \operatorname{Re}f \cdot C_{0,j} \left(1 + \frac{C_{0,j} s_j}{\operatorname{Re}f C_{0,j}} \right) = t_{po} \left(1 + \frac{C_{0,j} s_j}{\operatorname{Re}f C_{0,j}} \right) \quad C_{0,j,\text{MS}} = C_L$$

$$t_p = \sum_{j=1}^N t_{p,j} = t_{po} \sum_{j=1}^N \left(1 + \frac{C_{0,j} s_j}{\operatorname{Re}f C_{0,j}} \right)$$

OVERALL DELAY

$$\text{CALCULUS: } \partial t_p / \partial C_{0,j} = 0 \rightarrow \frac{1}{\operatorname{Re}f C_{0,j}^2} \cdot C_{0,j} s_j = 0 \rightarrow$$

OPTIMUM VALUES of s_j among s_j

$$f_{opt}^N = C_L / C_{0,j} = F \quad f_{opt} = \sqrt{F} \quad t_p = N t_{po} \left(1 + \frac{f_{opt}}{\operatorname{Re}f} \right)$$

OPTIMAL FAIR-OUT

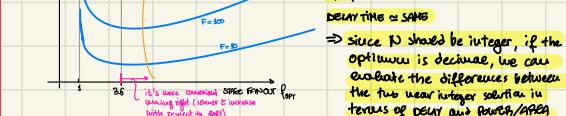
$$\text{THEN, HOW TO SIZE STAGES? } f_{opt} = \frac{C_{0,j}^{(1)}}{\operatorname{Re}f C_{0,j}^{(1)}} = \frac{C_{0,j}^{(1)} s_j}{\operatorname{Re}f C_{0,j}^{(1)} s_j} \Rightarrow s_j = s_0 f_{opt} \quad \text{OPTIMAL SIZING}$$

Now let's asses the **OPTIMUM NUMBER OF STAGES**: once we find f_{opt} , for a given C_L and C_0 ...

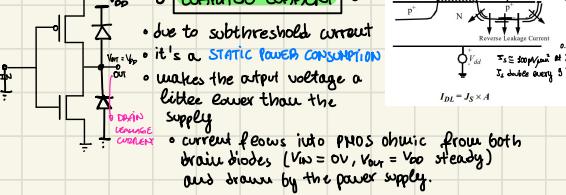
$$t_p = N t_{po} \left(1 + \frac{f_{opt}}{\operatorname{Re}f} \right) \Rightarrow \frac{\partial t_p}{\partial N} = f_{opt} = \sqrt{F} \cdot \frac{f_{opt}}{N} = 0 \Rightarrow N_{opt} = \exp \left(1 + \frac{f_{opt}}{\operatorname{Re}f} \right)$$

OPTIMUM NUMBER OF STAGES

we will consider directly the results $f_{opt} = 3.6$ $\operatorname{Re}f = 1 \Rightarrow N_{opt} = \frac{\ln(F)}{\ln(f_{opt})}$

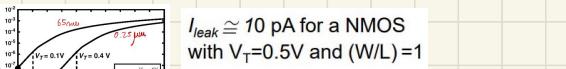


LEAKAGE CURRENT



$$P_{DC} = I_{leak} \cdot V_{DD} = I_{leak} \cdot \mu \left(\frac{W}{L} \right) e^{\left(\frac{V_{DD}-V_t}{nV_t} \right)}$$

STATIC POWER CONSUMPTION



FULL SCALING (CONSTANT ELECTRICAL FIELD)

ideal model → dimensions and voltage scale together by the same factor S

FIXED VOLTAGE SCALING

most common model until recently → only dimension scales, voltages remain constant

GENERAL SCALING

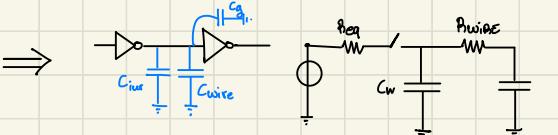
most realistic for today situations → voltages and dimensions scale with different factors

Parameter	Relationship	Full Scaling
W, L, t_{ox}		$1/S$
V_{DD}, V_T		$1/S$
N_{sub}		S
area	WL	$1/S^2$
C_0	$\epsilon_{ox} / (t_{ox} \cdot WL)$	$1/S$
I_{DSAT}	$C_{0N}(W/L) \cdot V_{DSAT} \cdot V_{DD}$	$1/S$
R_{eq}	V_{DD} / I_{DSAT}	1
delay	$R_{eq} \cdot C_0$	$1/S$
frequency	$1/t$	S
energy	$C(V_{DD})^2$	$1/S^3$
power	$C(V_{DD})^2 f$	$1/S^2$

Wires in ICs

The IC interconnections are in multiple layers with in-between parasitics which worsen performances

- CAPACITIVE → - AFFECT PROPAGATION DELAY
- RESISTIVE → - REDUCE RELIABILITY
- INDUCTIVE → - INCREASE POWER CONSUMPTION



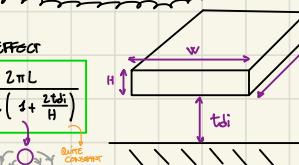
HOW TO SIMPLIFY THE PROBLEM?

- 1) **inductance** can be neglected if the wire resistance is large or if the rise/fall of the input signal is large. $R_w \gg L$ (kHz-few GHz)
- 2) when the **WIRE** is **SHORT** and the **EQUIVALENT RESISTANCE** R_{eq} of the drivers is **LARGE**, the **WIRE RESISTANCE** can be neglected $R_{wire} \ll R_{eq}$
- 3) when the separation between nearby wires is large or when the wires run in parallel for a short distance, the **INTERWIPE CAPACITANCE** can be neglected.

CAPACITANCE MODEL

PARALLEL PLATE + FRINGING EFFECT

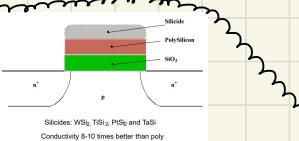
$$C_{tot} = E_{di} \frac{WL}{tdi} + E_{fr} \frac{2\pi L}{ln(1 + \frac{2tdi}{H})}$$



WIRE RESISTANCE

$$R_w = \rho \frac{L}{H \cdot W}$$

SHEET RESISTANCE

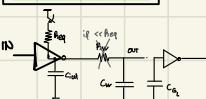


Material	ρ ($\Omega \cdot \text{m}$)
Silver (Ag)	1.6×10^{-8}
Copper (Cu)	1.7×10^{-8}
Gold (Au)	2.2×10^{-8}
Aluminum (Al)	2.7×10^{-8}
Tungsten (W)	5.5×10^{-8}

Material	Sheet Resistance (Ω/\square)
n- or p-well diffusion	1000 - 1500
n^+, p^+ diffusion	50 - 150
n^+, p^+ diffusion with silicide	3 - 5
n^+, p^+ polysilicon	150 - 200
n^+, p^+ polysilicon with silicide	4 - 5
Aluminum	0.05 - 0.1

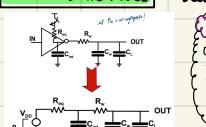
INTERCONNECT MODELING

the LUMPED C MODEL:



$$Z_p = \ln(z) R_{eq} (C_{out} + C_w)$$

the LUMPED R/C MODEL:



Delay can be assessed through...

(MORE THEOREM)

ONLY IF...

- ONE INPUT

- NO FEEDBACK

- ALL CAPACITANCES REFERRED TO GND OR FIXED MESHES

① for each capacitance we look at the resistances in the paths from the input to the capacitance itself.

② we look for resistances in the path INPUT-OUTPUT

③ the delay for a capacitance is given by the resistances in the two paths.

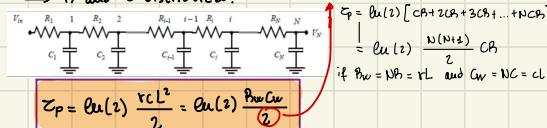
$Z_p = C_{out}(z) Z_{pi}$

$Z_p = \sum_i C_{out} R_{ik}$

overestimate!

we can apply it to a wire:

→ R and C distributed.



$$Z_p = \ln(z) \frac{R_{eq} L^2}{2} = \ln(z) \frac{R_{eq} C_w}{2}$$

INTERWIPE CAPACITANCE

the INTERWIPE CAPACITANCE contribution is becoming more and more dominant in modern technology

↳ the OVERALL PARASITIC CAPACITANCE affecting the wire can be evaluated by means of a parasitic capacitor in the layout environment.

Typical for a given technology a table is given reporting PARALLEL TYPE and THE FRINGING EFFECT contributions for a wire in a layer with respect to another wire in another layer.

Layer	Poly	Alu	Alu	Alu	Alu	Alu	Alu
Capacitance	40	95	85	85	85	115	115

single table for evaluating the capacitance towards a nearby wire interconnected with the same layer, supposed that the two wires are at the maximum distance allowed by the technology process.

INDUCTANCE

Assuming $W < t_{di}$ and a negligible thickness H of the wire

$$\text{INDUCTANCE} = L = \frac{\mu_0}{2\pi} \ln \left(\frac{8t_{di}}{W} + \frac{W}{4t_{di}} \right) \approx 0.4 \text{ pH}$$

if $R = 0.075 \Omega/\mu\text{m}$, it's negligible for freq. smaller than

$$2\pi f L = R \Rightarrow f = \frac{r}{2\pi L} \approx 30 \text{ Hz}$$

inductance can be obtained from $V = \frac{L}{C} = \sqrt{\frac{C}{2\pi f}}$

to take into account the effect of a distributed model, we can use the T-MODEL or the pi-MODEL



$$Z = \left(C_{out} + \frac{C_w}{2} \right) R_{eq} + \left(\frac{C_w}{2} \right) (R_{eq} + R_{in})$$

$$Z = C_{out} R_{eq} + C_w \left(R_{eq} + \frac{R_{in}}{2} \right)$$

we obtain:

$$Z_p = \ln(z) \left[C_{out}(2R_{eq} + r_L) + R_{eq} C_L + \frac{C_r L^2}{2} \right]$$

BUFFERS are equivalent

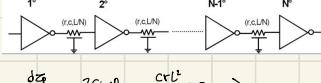
$L = 50 \rightarrow L^2 = 500$

$L = 5.5 \rightarrow L^2 = 30.25$

$L = 5 \rightarrow L^2 = 25$

INSERTING BUFFERS

the delay of a wire can be optimized by inserting buffers:



$$\frac{dZ_p}{dN} = 2C_{\text{buf}} \omega_{\text{req}} - \frac{CrL^2}{2N^2} = 0 \Rightarrow$$

NOTE:
OPT of
buffers

$$N = L \sqrt{\frac{Cr}{4C_{\text{buf}} \omega_{\text{req}}}}$$

for the optimal size:

$$Z_p = \text{Ex}(2)N \left[sC_{\text{buf}} \left(\frac{\omega_{\text{req}}}{s} \right) + \frac{Cl}{N} \left(\frac{\omega_{\text{req}}}{s} + \frac{rL}{2s} \right) + sC_{\text{buf}} \left(\frac{\omega_{\text{req}}}{s} + \frac{rL}{N} \right) \right]$$

$$\frac{dZ_p}{ds} = -\frac{Cl}{N} \left(\frac{\omega_{\text{req}}}{s^2} \right) + C_{\text{buf}} \left(\frac{rL}{N^2} \right) = 0$$

Z_{opt} is

$$Z_{\text{opt}} = N_{\text{opt}} \left[2Z_{\text{po}} + 2Z_{\text{pu}} + 2Z_{\text{pl}} + 2Z_{\text{po}} \right] = 8N_{\text{opt}} Z_{\text{po}}$$

$$= 2.7 \sqrt{(C_{\text{buf}} C_{\text{req}})(\omega_{\text{req}}^2 C_{\text{buf}}^2)}$$

it's worth inserting a buffer if:

$$L > \sqrt{\frac{16 C_{\text{buf}} \omega_{\text{req}}}{Cr}} = \sqrt{\frac{23 Z_{\text{po}}}{Cr}}$$

LENGTH CONDITION

SUBSTITUTE N_{opt} IN EQ.

$$S = \sqrt{\frac{\omega_{\text{req}}^2 C}{r C_{\text{buf}}^2}}$$

SIZE

INDUCTANCE MATTERS

The inductance of a line has to be taken into account when:

- the rise/fall time (t_r) of the signal is much lower than the time-of-flight (t_{prop}) along the wire;
- $t_r \ll t_{\text{prop}} \Rightarrow L \gg \frac{1}{r \sqrt{C}}$
- I and c being the inductance and capacitance per unit length, respectively.

see the video value for this topic

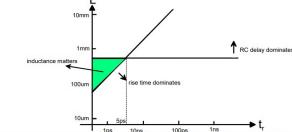
- the RC delay is much lower than the time-of-flight (i.e., the signal propagates as an electromagnetic wave)

$$\frac{RC}{n^2} \ll t_{\text{prop}} \Rightarrow L \ll r \sqrt{\frac{C}{Z}}$$

Where the inductance matters

$$r=0.075 \Omega/\mu\text{m}, c=110 \text{ aF}/\mu\text{m}, l=0.39 \text{ pH}/\mu\text{m}, Z_0=60 \Omega, V=0.21 \text{ ns/cm}$$

$$\frac{t_r}{l} \ll L \ll \frac{1}{r \sqrt{C}} = 800 \mu\text{m}$$



SCALING

Scaling all physical dimensions of same factor S :

Parameter	Relation	Local Wire	Constant Length	Global Wire
W, H, t		$1/S$	$1/S$	$1/S$
L		$1/S$	1	$1/S_C$
C	LW/t	$1/S$	1	$1/S_C$
R	L/WH	S	S^2	S^2/S_C
CR	L^2/Ht	1	S^2	S^2/S_C^2

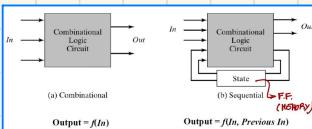
constant resistance scaling:

Unchanging the height of the wires only gives a small advantage since it brings into foreground the fringing and inter-wire capacitance problems

Parameter	Relation	Local Wire	Constant Length	Global Wire
W, t		$1/S$	$1/S$	$1/S$
H		1	1	1
L		$1/S$	1	$1/S_C$
C	$\epsilon_r L W/t$	$\frac{1}{S} (\epsilon_r S)$	ϵ_r	$\epsilon_r S_C$
R	L/WH	(1) Corrected	S	S/S_C
CR	L^2/Ht	$\epsilon_r S$	$\epsilon_r S$	$\epsilon_r S_C^2$

FC-CHMOS LOGIC GATES

- It's a DIGITAL CIRCUIT with complementary PULL-UP and PULL-DOWN NETWORKS which realize a BOOLEAN FUNCTION.
- It's a STATIC LOGIC (output always connected to V_{DD} or GND) [in contrast with DYNAMIC LOGIC based on the storage of a charge on a capacitance]
- It's also a COMBINATIONAL LOGIC:



BOOLEAN LOGIC

DE MORGAN THEOREM: PUN is the dual of the PDN:

$$\overline{A+B} = \overline{A}\overline{B}$$

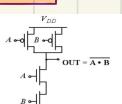
$$\text{and } \overline{AB} = \overline{A} + \overline{B}$$

DE MORGAN THEOREM

NAND and NOR:

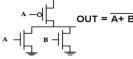
A	B	Out
0	0	1
0	1	0
1	0	0
1	1	0

Truth Table of a 2 input NAND gate



A	B	Out
0	0	1
0	1	0
1	0	0
1	1	0

Truth Table of a 2 input NOR gate

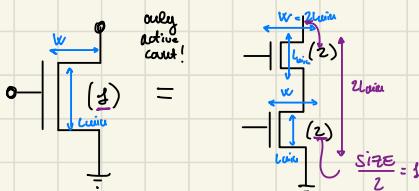


BUBBLE PUSHING:

$$\begin{aligned} \text{NAND} &= \text{NOR} + \text{INVERTED TERMINALS} \\ \text{NOR} &= \text{NAND} + \text{INVERTED TERMINALS} \end{aligned}$$

APPLICATION OF DE MORGAN THEOREM

EQUIVALENT CMOS GATE



MOSFET EQUIVALENT AC MODEL

RESISTANCE: B_i of TRANSISTOR $M_i^{(s)}$ is R_{ds}/s

CAPACITANCE:

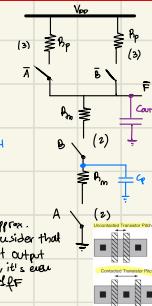
OBTAINED by $C_g^i = C_d^i = 2 \frac{PF}{\mu m^2}$ OF WIDTH

for example:

$$C_{out} = 2 \frac{PF}{\mu m^2} (3+3+2) \frac{\mu m}{\mu m} = 4 \mu F$$

$$C_P = \dots = 2 \frac{PF}{\mu m^2}$$

If we consider that there is not output pull current, it's even lower... $\approx 1 \mu F$



FC-CHMOS LOGIC BASICS

Characterized by a PULL-UP NETWORK connected to V_{DD} and a PULL-DOWN NETWORK connected to GND.

They work in a complementary way. The PUN works when the output should be '1'. The PDN works when the output is '0'.

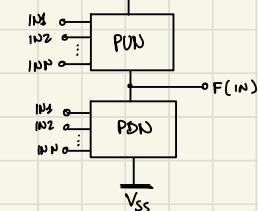
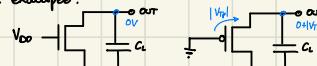
(In TRANSIENT they can be both active)

• WHY NMOS for PDN and PMOS for PUN?

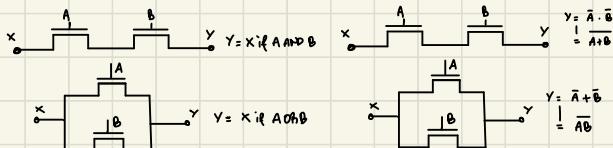
- NMOS does a better pull-down

- PMOS does a better pull-up

for example:



• WE CAN REALIZE BASIC Boolean FUNCTIONS BY PLACING NMOS/PMOS IN SERIES/PARALLEL:



FC-CHMOS LOGIC DESIGN AND PROPERTIES

To DESIGN the FC-CHMOS logic which synthesizes a Boolean function we:

1) Design PDN

2) Identify subnets

3) Design PUN

o HIGH PERFORM

$$V_{dd} = \text{OV}$$

$$V_{ss} = \text{V}_0$$

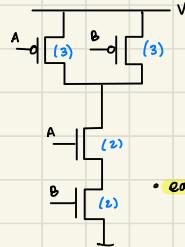
V_h approx. in the middle
achievable static power consumption
(no direct path between V_{DD} and GND)

we obtain...



STATIC CHMOS PROPERTIES

Let's work an example:



- equivalent $(W/L)_m = 1$, $(W/L)_p = 6$ for transistor involving both PMOS. Then larger V_h . $[A=B=0 \rightarrow 1]$

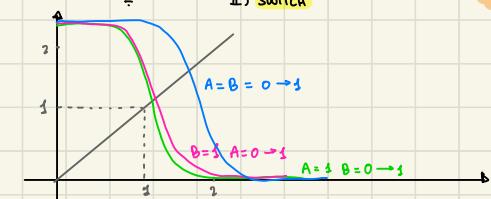
- if only one PMOS active then $B=3$, V_h in middle
- if $A=0 \rightarrow 1$, $B=2 \rightarrow 1$, V_h is little larger than that of $B=0 \rightarrow 1$, $A=1$, because the source of M_{out} has a source at $V>GND$.

TWO GATES HAVE THE SAME SIZE IF THEY EXPERIENCE THE SAME EQUIVALENT RESISTANCE IN INPUT-CASE PATHS.

• each transistor is modeled as a

- I) RESISTOR (eq. large signal resistance, scaled by ratio of device)
- II) SWITCH

$$R_D \propto \frac{L}{W} = \frac{1}{(w/l)} \propto \frac{1}{S}$$

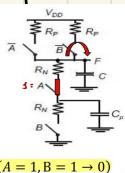


PROPAGATION DELAY

τ is assed as the mean between the τ_{PLH} (PUN) and the τ_{PHL} (PDN):

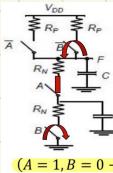
It can happen the one of the networks is slower due to PARASITIC CAPACITANCES

- Example: with Elmore's theorem



$$\tau_{PLH} = \text{Elm}(2) R_P^{(1)} [C + C_P]$$

(A = 1, B = 1 → 0)

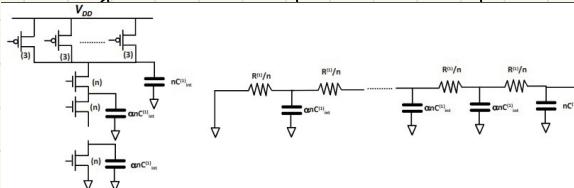


(A = 1, B = 0 → 1)

$$\tau_{PHL} = \text{Elm}(2) \left[C R_m^{(2)} + \frac{R_m^{(2)}}{2} C_P \right]$$

$$\begin{aligned} \tau_{PLH} &= \text{Elm}(2) R_P^{(1)} C \\ \tau_{PHL} &= \text{Elm}(2) (2 R_m) C \end{aligned}$$

We can generalize an equivalent model for the PDN:



HOWEVER, USUALLY THE INTERNAL NODE CAPACITANCE IS NEGLECTED FOR 1ST ORDER ANALYSIS.

• Other notes:

① INCREASING GATE DIMENSION BRINGS TO AN IMPROVEMENT OF THE PROPAGATION DELAY, UNTIL THE EXTERNAL CAPACITANCE IS DOMINANT.

② IN CASE OF LARGE FAN-IN AND A LOT OF TRANSISTORS IN SERIES, IT'S POSSIBLE TO ADOPT A PROGRESSIVE SIZING (WITH HUG LAYER THE OUTPUT BEING THE SMALLEST, SINCE IT HAS TO DISCHARGE LESS CAPACITANCES)

③ THE LAST SIGNAL COMING TO THE GATE IS CALLED CRITICAL PATH; IT HAS TO BE CONNECTED AS CLOSE AS POSSIBLE TO THE OUTPUT, TO ALLOW CAPACITANCES TO BE DISCHARGED BEFORE THE ARRIVAL OF LAST SIGNAL

④ LARGE FAN-IN CAN BE AVOIDED IN FAVOUR OF MORE SIMPLEX GATES, SO TO SPLIT IT IN CASCADE OF SMALLER GATES

TRANSISTOR SIZING IN COMBINATIONAL NETWORKS (PT.1)

Let's see how we can formalize the gates sizing for the OPTIMIZATION OF THE DELAY, by extending the INVERTER CHAIN THEORY.

Let's consider the PROPAGATION DELAY, neglecting the parasitic capacitances of internal nodes:

$$\tau_p = \text{Elm}(2) R_{eq} (C_{int} + C_{ext})$$

EQUIVALENT RESISTANCE OF THE FAN-IN/FAN-OUT PATH

LOAD CAPACITANCE

We size R_{eq} to be equal to the resistance of a minimum size inverter.

$$R_{eq} = \frac{\text{P}_{eq}(s)}{5} \quad \text{size of combinational gate}$$

also:

$$C_{int} = s P C_{int}^{(1)}$$

INTRINSIC CAPACITANCE

so, we obtain:

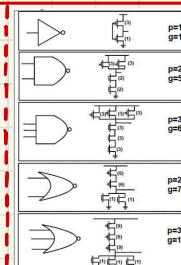
$$\begin{aligned} \tau_p &= \text{Elm}(2) P_{eq} (C_{int} + C_{ext}) = \text{Elm}(2) \frac{\text{P}_{eq}(s)}{5} (P_1 C_{int}^{(1)} + C_{ext}) = \text{Elm}(2) P_{eq} C_{int}^{(1)} \left(P + \frac{C_{ext}}{5 C_{int}^{(1)}} \right) = \\ &= \tau_{po} \left[P + \frac{C_{ext}}{5 P_1 C_{int}^{(1)}} \right] = \tau_{po} \left[P + \frac{C_{ext}}{5 P_1 C_{int}^{(1)} C_0} \right] = \tau_{po} \left[P + \frac{1}{P_1} \cdot \frac{C_{ext}}{C_0} \right] = \tau_{po} \left[P + \frac{g}{P_1} \right] \end{aligned}$$

*NOTE: P and g depends on the topology, complexity of gate, NOT the size. τ_{po} is the INTRINSIC DELAY (τ_{po} and P fixed by TECHNOLOGY).

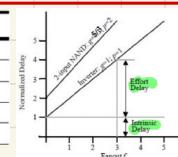
τ_{po} is the EXTRINSIC DELAY. if s_1, P_{eq1}

$\tau_{ext. delay + wire} \approx \tau_{po}$

only f (FAN-OUT) depends on the size (ratio C_0, C_{ext})



Gate type	P	g
Inverter	1	
n-input NAND	n	
n-input NOR		
2-way multiplexer	2n	
XOR, NXOR	n^{2+1}	
2-to-1 MUX		



Some examples:

to REDUCE THE DELAY WE NEED TO REDUCE THE GATE EFFORT
 $b = \frac{g}{P_1}$ by
 - INCREASING THE GATE SIZE
 - CHOOSE A TOPOLOGY WITH A LOWER LOGICAL EFFORT

to EVALUATE THE OVERALL DELAY OF A CHAIN WE CAN SIMPLY SUM THE DELAY OF THE GATES (NEGLECTING THE EFFECT OF DELAYS IN CASCADE):

$$\begin{aligned} \tau_p &= \sum_{j=1}^n \tau_{p,j} = \tau_{po} \sum_{j=1}^n (P_j + \frac{f_j g_j}{5}) \Rightarrow F = \prod_{j=1}^n f_j = \frac{C_L}{C_{g,1}} \quad \text{PATH LOGICAL EFFORT} \\ G &= \prod_{j=1}^n g_j \quad \Rightarrow H = F \cdot G \quad \text{PATH EFFECTIVE EFFORT} \end{aligned}$$

since ALL THE OPTIMIZED PATHS MUST HAVE THE SAME GATE EFFORT, WE OBTAIN:

$$H = h_{opt} \Rightarrow h_{opt} = \sqrt[n]{H} = \sqrt[n]{FG} \quad \text{PATH EFFECTIVE EFFORT}$$

$$\tau_{p,min} = \tau_{po} \left[\left(\sum_j P_j \right) + \frac{N h_{opt}}{5} \right] \quad \text{OPTIMUM DELAY}$$

so, we can obtain the size of the elements in the chain:

$$S_j = \frac{s_1 g_j}{\sum_{i=1}^{j-1} f_i} \quad \text{OPTIMUM SIZING}$$

FURTHER OPTIMIZATION OF THE DELAY (PT. 2)

for INVERTER CHAIN we had verified that, once N number of inverters is decided, the minimum delay can be pursued sizing the inverters so that the fan-out is equal for every stage. However, a further optimization can be done if the number of stages can be changed.

→ THE OPTIMUM FAN-OUT IS 3.6 (for $P=1$), with $N = \text{elu}(F)/\text{elu}(3.6)$

THE SAME PROCEDURE CAN BE DONE FOR A GENERIC CHAIN OF GATES. ONCE THE GATES ARE FIXED THE DELAY IS OPTIMIZED WHEN:

$$b = \text{hop} = \sqrt[n]{FG} = \sqrt[n]{\left(\frac{C_L}{g_{j,n}}\right) \prod_{i=1}^n g_i}$$

CAN WE FURTHER OPTIMIZE INSERTING AN EVEN NUMBER OF INVERTERS?

LET'S CONSIDER 16 GENERIC GATES IN AN OPTIMIZED CHAIN:

$$\begin{aligned} \tau_{\text{path}} &= \tau_0 \sum_{j=1}^k \left(p_j + \frac{k_j}{\eta} \right) \quad \rightarrow \text{LET'S NOW INSERT AN EVEN NUMBER} \\ &\quad \text{OF INVERTERS} \quad \text{WE INSERT INVERTERS} \\ &\quad (\text{OVERALL NUMBER OF STAGES } N = k+2) \quad \text{DO NOT CHANGE SINCE} \\ &\Rightarrow \tau_p = \tau_0 \sum_{j=1}^{k+1} \left(p_j + \frac{k_j}{\eta} \right) = \tau_0 \sum_{j=1}^k \left(p_j \right) + \tau_0 (N-k-1) + \tau_0 \frac{N-1}{\eta} = \tau_0 \left[\sum_{j=1}^k \left(p_j \right) \right] - \tau_0 k + \tau_0 N + \frac{N-1}{\eta} \\ &\Rightarrow \frac{\partial \tau_p}{\partial N} = 0 \rightarrow \left\{ \begin{array}{l} \eta = \text{hop} = \text{elu}(b_{\text{opt}}) = 0 \\ N = 16 \end{array} \right. \quad \rightarrow \text{OPTIMUM NUMBER OF STAGES/STEPS} \\ &\quad \text{OF INVERTERS} \quad \text{hop} = 3.6 \quad \rightarrow \text{Nopt} = \frac{\text{elu}(H)}{\text{elu}(3.6)} \\ &\quad \text{LESS TRANSISTORS WITH BIG SIZES!} \end{aligned}$$

Typically it's better to insert inverters pair at the end of the chain because of the increasing steps (less transistors with big sizes!).

POWER REDUCTION

- (A) LOGIC RECONSTRUCTION: CHAIN STRUCTURE usually features a lower switching activity with respect to TREE STRUCTURE
however
GLITCHES CAN OCCURS IN THE CHAIN STRUCTURE BUT NOT IN THE TREE STRUCTURE
- (B) INPUT RE-ORDERING: It's beneficial to SWITCHING ACTIVITY to POST-PONE THE SIGNALS WITH HIGH TRANSITION RATE
- (C) GLITCH REDUCTION: It's useful to re-design in order to have all ARRIVAL TIMES IDENTICAL.
- (D) CIRCUIT PARAMETERS: REDUCE GATE SIZE, VOLTAGE

$$P = C_L V_{DD}^2 d_{SW} f_{CH}$$

POWER CONSUMPTION FOR A GATE

BRANCHING

We can define a BRANCHING FACTOR for a gate as the ratio between the overall extrinsic capacitance connected to the output and the extrinsic capacitance along the path.

$$b = \frac{\text{cou-path} + \text{Coff-path}}{\text{cou-path}}$$

BRANCHING FACTOR

$$B = \prod_{j=1}^n b_j$$

BRANCHING EFFORT

$$H = F \cdot G \cdot B$$

PATH EFFECTIVE EFFORT

$$\tau_p = \tau_0 \left(p + \frac{f_{gb}}{\eta} \right)$$

Same formulae of NO BRANCHING can be adopted if we consider the fan-out along the path:

$$S_j = \frac{1}{g_j} \prod_{i=s}^{j-1} f_i \quad f_i = \frac{\text{hop}}{g_i b_i}$$

POWER CONSUMPTION OF FC-CHOS LOGIC GATES

The SWITCHING ACTIVITY η_{sw} is fundamental and depends on the logic function.

It's the probability p_0 that output is '0' in one cycle multiplied for probability p_1 that the output is '1' in the next clock cycle:

$$P_0 P_1 = P_0 (1 - P_0)$$

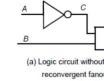
(valid if input signals are uniformly distributed and independent)

HOWEVER, HOW TO DEAL WITH NON-UNIFORM DISTRIBUTIONS? HOW PATH MODIFY THE DISTRIBUTION OF THE INPUTS?

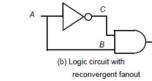
P_A, P_B PROBABILITY OF 'A'/B' TO BE '1'

$$\Rightarrow d_{SW} = P_0 P_1 = [3 - (1 - P_A)(1 - P_B)](1 - P_A)(1 - P_B)$$

WE NEED ALSO TO CONSIDER SHARED INPUTS (BECOMING FAN-OUT):



(a) Logic circuit without reconvergent fanout



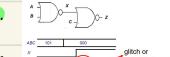
(b) Logic circuit with reconvergent fanout

Case (a): $P(Z=1) = P(B=1) \cdot P(C=1) = 0.25$

Case (b): $P(Z=1) = P(B=1) \cdot P(C=1|B=1) = 0$

ALSO GLITCHES CAN CAUSE PROBLEMS.

SPURIOUS TRANSITIONS BECAUSE OF FINITE TRANSITION DELAY DIFFERENCES BETWEEN PATHS



PATIOED LOGIC

Both STATIC and DYNAMIC properties depend on the RATIO BETWEEN THE SIZE OF PULL-UP AND PULL-DOWN NETWORKS (differently from FC-CMOS architecture)

WHY DO WE NEED A DIFFERENT APPROACH?

FC-CMOS LOGIC:

ROBUST AND SIMPLE APPROACH FOR IMPLEMENTING A GATE

HOWEVER, AS THE FAN-IN INCREASES,

FC-CMOS LOGIC IS PARIOLESS
 $V_{OL} = GND$, $V_{OH} = V_{DD}$ DO NOT DEPEND ON RATIO BETWEEN PMOS AND NMOS SIZE

① for a N FAN-IN we need $2N$ GATE.

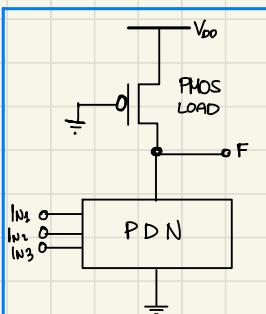
AREA CONSUMING

(also PMOS are wider than NMOS by a factor 3 for a switching threshold in the middle)

② LARGE GATE/OUTPUT CAPACITANCE (due to each input connected to PMOS and NMOS)

LARGE LOGICAL EFFORT

PSEUDO-NMOS LOGIC



THE PSEUDO-NMOS LOGIC IMPLEMENTS THE SAME PDN OF A FC-CMOS GATE but ONLY ON PMOS WITH GROUNDED GATE, ALWAYS ON.

Allow to REDUCE THE NUMBER OF TRANSISTORS and the AREA REQUIRED.
Also reduced the INPUT GATE CAPACITANCE

③ REDUCED NOISE MARGINS ④ LARGER DC POWER CONSUMPTION
⑤ ASYMMETRIC OUTPUT TRANSITIONS

- PMOS ALWAYS ON. IF PAN INPUTS ARE ZERO, PMOS CHARGES OUTPUT CAPACITANCE and output is at V_{DD} .
- IF PAN is active, the PULL-DOWN is STRONGER THAN THE PULL-UP, and THE OUTPUT GOES TO GND, but $V_{OL} > 0V$ (still there is a pull-up).
- the NUMBER OF TRANSISTORS is N_{Tg} (instead of $2N$). Area $\downarrow\downarrow$
- also CAPACITANCES (both INPUT/OUTPUT) are reduced. DELAY $\downarrow\downarrow$

- $V_{OH} = V_{DD}$ (PDN OFF). $V_{OL} > 0V$, obtained by equating the two currents of PHOS and PDN

$$\text{for a pseudo NMOS inverter: } V_{OL} \cong \frac{I_{PDN} \left(\frac{W}{L} \right)_P V_{DD,OFF}}{I_{PHOS} \left(\frac{W}{L} \right)_D}$$

- The THRESHOLD VOLTAGE can be similarly obtained by equating the two currents:

$$I_{PHOS} \left(\frac{W}{L} \right)_D V_{DD,OFF} [V_H - V_{TH} - \frac{V_{DD,OFF}}{2}] = I_{PDN} \left(\frac{W}{L} \right)_P V_{DD,OFF} [V_{DD} - V_{TP} - \frac{V_{DD,OFF}}{2}] \Rightarrow V_H = V_{TH} + \frac{V_{DD,OFF}}{2} + \frac{I_{PDN} \left(\frac{W}{L} \right)_P V_{DD,OFF}}{I_{PHOS} \left(\frac{W}{L} \right)_D} \left[(V_{DD} - V_{TP} - \frac{V_{DD,OFF}}{2}) \right]$$

$$P_{DD} = V_{DD} I_{DD} = V_{DD} k_P \left(\frac{W}{L} \right)_P \left[(V_{DD} - V_{TP}) V_{DD,OFF} - \frac{V_{DD}^2}{2} \right] \left[1 + 2 \left(\frac{V_{DD} - V_{TP}}{V_{DD,OFF}} \right) \right]$$

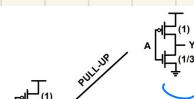
(REAL UNIT OF THIS ARCHITECTURE)
USED FOR STATIC AREA CONSTRAINED APPLICATIONS

- Direct path between V_{DD} and GND when output is low. STATIC POWER CONSUMPTION $\uparrow\uparrow$
- Reducing $(W/L)_P$ then V_{OL} & DC-POWER but DELAY \uparrow (TRADE-OFF STATIC AND DYNAMIC PROPERTIES)
- We have to choose between DELAY $\downarrow\downarrow$ or $\uparrow\uparrow$ (which one is faster). If $(W/L)_P \uparrow$ then $\tau_{RHS} < \tau_{PHS}$. It's evaluated in the same way of FC-CMOS with the concept of equivalent resistance and output capacitance.
- o PULL-UP TRANSITION as in FC-CMOS. Pull-down transition hindered by PMOS, but we CAN NEGLECT IT AT 1ST ORDER.
- o for PULL-DOWN: overall current = PMOS current. Therefore:

$$I_{PHOS,DOWN} \cong \left(\frac{1}{R_{PM}} - \frac{1}{R_{PP}} \right)^{-1} = \frac{P_{DD,IN}}{1 - \frac{R_{PM}}{R_{PP}}}$$

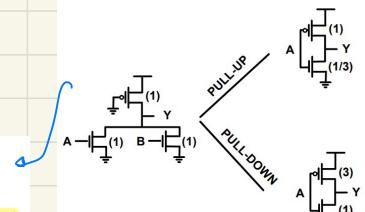
EVALUATION OF LOGICAL EFFORT AND INTRINSIC DELAY FACTOR MUST BE ASSESSED FOR EACH TRANSITION (COMPARING THE GATE HAVING THE SAME CURRENT CAPABILITY FOR THAT PARTICULAR TRANSITION).

We obtain BETTER LOGICAL EFFORT, NO IMPROVEMENT FOR INTRINSIC DELAY



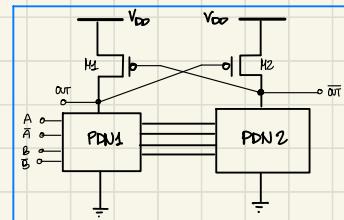
$$\begin{aligned} &\left\{ \begin{array}{l} p_u = \frac{1+1}{1+1/3} = 1.5 \\ g_u = \frac{1}{1+1/3} = 0.75 \end{array} \right. \quad \left\{ \begin{array}{l} p_d = \frac{1+1}{1+3} = 0.5 \\ g_d = \frac{1}{1+3} = 0.25 \end{array} \right. \rightarrow \left\{ \begin{array}{l} p = \frac{p_u + p_d}{2} = 1 \\ g = \frac{g_u + g_d}{2} = 0.5 \end{array} \right. \\ &\left\{ \begin{array}{l} p_u = \frac{1+1+1}{1+1/3} = 2.25 \\ g_u = \frac{1}{1+1/3} = 0.75 \end{array} \right. \quad \left\{ \begin{array}{l} p_d = \frac{1+1+1}{1+3} = 0.75 \\ g_d = \frac{1}{1+3} = 0.25 \end{array} \right. \rightarrow \left\{ \begin{array}{l} p = \frac{p_u + p_d}{2} = 1.5 \\ g = \frac{g_u + g_d}{2} = 0.5 \end{array} \right. \end{aligned}$$

For a FC-CMOS gate the intrinsic delay factor and the logical effort are 2 and $7/4=1.75$, respectively.



DCVSL: DIFFERENTIAL CASCODE VOLTAGE SWITCH LOGIC

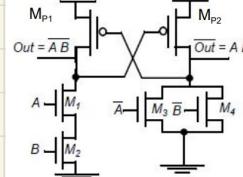
IT'S AN IMPROVEMENT OF PSEUDO NMOS LOGIC, THAT ELIMINATES DC POWER CONSUMPTION AND MAKES $V_{OL} = GND$.
THE IMPROVEMENT IS OBTAINED WITH A **POSITIVE FEEDBACK** AND A **DIFFERENTIAL STRUCTURE**



assures that the LOAD device is turned off when not needed

differential structure requires DIFFERENTIAL INPUTS. For any input also the complementary signal is needed.
COMPLEMENTARY OUTPUTS ARE PROVIDED.

- It's made of 2 COMPLEMENTARY PMOS NETWORKS, MUTUALLY EXCLUSIVE. This connection implements a positive feedback. Also inputs are complementary. Load is built with two PMOS with the gates connected to the complementary output.



- AT STEADY STATE THERE IS NO CONDUCTIVE PATH BETWEEN V_{DD} AND GND. AVOID DC POWER CONSUMPTION.
- V_{OL} IS AT GND
- The logic is PARTITIONED since if the PDN NETWORK is not stronger than the PMOS, the output cannot be DRIVEN LOW.
- PRODUCES BOTH OUTPUT AND INVERTED WITH NO TIME-SHIFT
HOWEVER

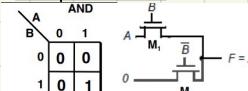
↳ DIFFERENTIAL ARCHITECTURE DOUBLE THE NUMBER OF WIRES
CROSS CONNECTION OF PMOS IS CRITICAL
CROSS-CONDUCTION CURRENT (PM ON, PMOS STILL ACTIVE)
DYNAMIC POWER CONSUMPTION INCREASED WITH RESPECT TO SINGLE ENDED
due to LARGER OUTPUT CAPACITANCE and LARGER SWITCHING ACTIVITY

IT'S AIM IS TO REDUCE THE NUMBER OF TRANSISTORS BY ALLOWING THE SIGNALS TO DRIVE BOTH GATE AND SOURCE DRAIN TERMINALS.

Used for AND, XOR, ...
then ADDERS, MULTIPLIERS

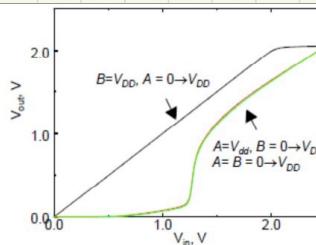
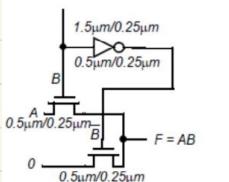
PASS-TRANSISTOR LOGIC

Example of an AND:



corresponds to building a MULTIPLEXER.
B is used to select which branch has to convey the input signal

REQUIRE ONLY 2 TRANSISTORS



$B = V_{DD}$ $A = 0 \rightarrow V_{DD}$: when B is high, the upper device is ON and copy the voltage V_A on the node F. When input voltage reaches $V_{DD} - V_m$, the output voltage starts to increase since the device is turned off.

$A = V_{DD}$ $B = 0 \rightarrow V_{DD}$: at the beginning of the transition, the lower device is ON while the upper device is OFF. The voltage at the node F remains at 0V until B is equal to V_m , after which the upper transistor starts to turn on. But the LOWER device is STRONGER since it has the GATE at V_{DD} . NOTE: output voltage remains low until the input reaches half V_{DD} .

Then the inverter switches and the lower device is turned off. (however in this case V_A is no longer transferred to the output, which is $V_B - V_m$)

There are some PROBLEMS:

① NMOS not so effective to pull-up a node Up to $V_{DD} - V_m$

$$V_t = V_{DD} + \eta P (\sqrt{2\mu F + V_{DD}} - \sqrt{2\mu F})$$

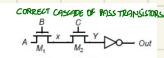
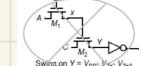
BODY EFFECT WORSENS THE SITUATION (high output voltage)

② PROPAGATION DELAY. The output node charges up quickly at the beginning of the transient, but as input voltage (source) increases, the V_{AS} is decreased, limiting the DRIVING CAPABILITY OF THE DEVICE

$$\tau_{PLH} \gg \tau_{PHL}$$

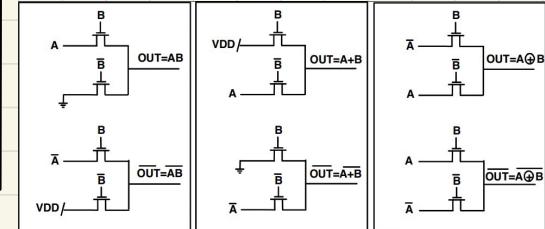
③ IF IT DRIVES A FC-CHOS INV. (for ex.) then the PHOS is cleverly turned off but there is always a DC POWER CONSUMPTION

④ PASS-TRANSISTOR GATES CANNOT BE CASCDED DUE TO THE PROBLEM OF THE HIGH OUTPUT VOLTAGE



DIFFERENTIAL (COMPLEMENTARY) PASS-TRANSISTOR LOGIC (DPL/CPL)

Complementary Pass-transistor Logic

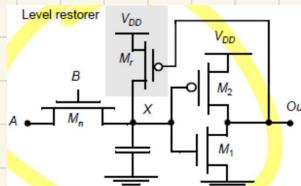


The SAME PASS-TRANSISTORS with the SAME GATE

SIGNALS are adopted for the TRUE and the COMPLEMENTARY FUNCTION. The DIFFERENCE is that INPUT SIGNALS are the COMPLEMENTARY ONES.

HOW TO SOLVE HIGH VOLTAGE PROBLEM?

I) USE LEVEL RESTORERS: PHOS TRANSISTOR IN FEEDBACK, GATE CONNECTED TO INVERTED OUTPUT OF PASS-TRANSISTOR



- if $A=0$, $B=1$, $X=0$, PHOS off, NO PROBLEMS

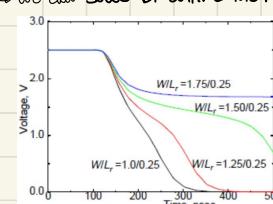
- if $A=1$, " $X=1$, PHOS ON pull-up even more!

↳ ELIMINATES ANY STATIC POWER DISSIPATION IN THE INVERTER.

⇒ STILL PASTED! HOW TO SIZE PHOS and PASS-TRANSISTORS?

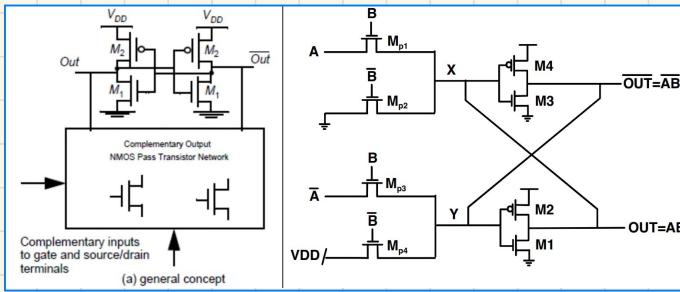
V_m and M_F must be such that X voltage goes below SWITCING THRESHOLD of the inverter (if PHOS is stronger than NMOS, out remains low and PHOS on!)

↳ we can solve by cutting the feedback loop on the inverter gate.



for $(W/L)_P < 3.5 (W/L)_m$ X goes below threshold.

for DIFFERENTIAL IMPLEMENTATIONS of PASS-TRANSISTOR LOGIC

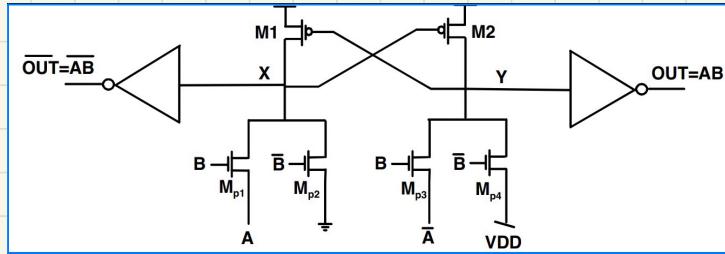


with TWO INVERTERS connected in CROSS-COUPLED FASHION.

- steep RATIOED logic, right sizing is MANDATORY

" SAPTL: SWING RESTORED PASS TRANSISTOR LOGIC "

HOW CAN WE SOLVE THE PROBLEM OF HIGH-VOLTAGE WITHOUT PRESORTING A RATIOED SOLUTION?



- two LEVEL RESTORERS (with GATES connected to opposite pass-transistor block and the two inverters) INVERTERS used as BUFFERS to allow CASCADING

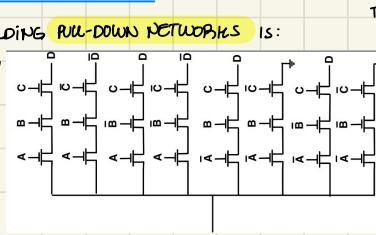
- works for every sizing.

Usually PMOS are sized with the same aspect ratio of the equivalent pull-down device.

" COMPLEMENTARY PASS-TRANSISTOR LOGIC "

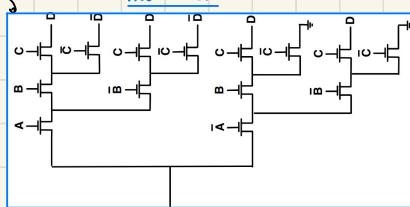
OTHER ALTERNATIVES FOR BUILDING FULL-DOWN NETWORKS IS:

	AB	CD
CD	00 01 11 10	00 01 11 10
00	0 0 1 1	0 0 1 1
01	0 0 0 0	0 0 0 0
11	1 1 1 1	1 1 1 1
10	0 0 0 0	0 0 0 0

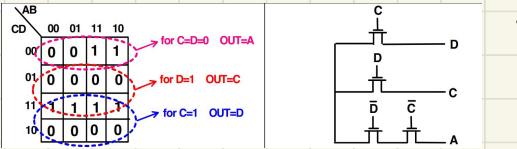


TREE STRUCTURE

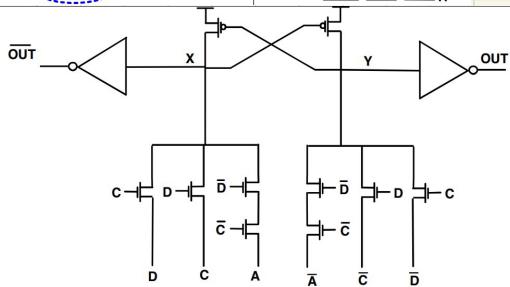
MUXPLEXER



OR BUILD IT FROM SCRATCH :

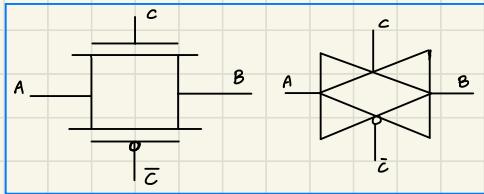


- The largest the circle, the lower the number of transistors in series that selects a signal



the TRANSMISSION GATE

to face VOLTAGE-DROP PROBLEM EMPLOY PHOS and NMOSTTRANSISTORS in PARALLEL to combine the best of both devices:



- IF $C = '1'$ then $A = B$. Otherwise OPEN CIRCUIT.

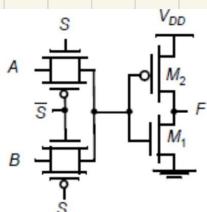
- if $A = 0 \rightarrow 1$ and $B = 0$ PHOS pull up to V_{DD}
NMOS pull up to $V_{DD} - V_{THN}$ (then it's OFF)

- if $A = 1 \rightarrow 0$ and $B = 1$ PHOS pull down to V_{THP} (then it's OFF)
NMOS pull down to GND

- it's essential to have LOW IMPEDANCE NODES!

- NMOS passes a strong '0'
but a weak '1'
- PHOS passes a strong '1'
but a weak '0'.

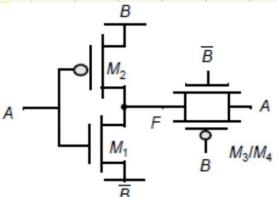
examples:



TRANSMISSION GATE

MUXPLEXER:

$$F = AS + BS$$



XOR

we can model a TG as a RESISTANCE.

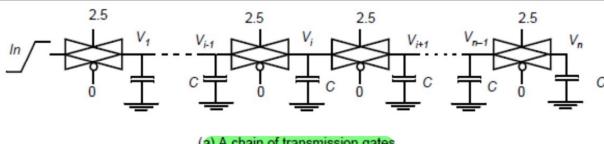
- Let's suppose $(W/L)_m = 1$; $(W/L)_p = 3$. We have to consider a LARGE SIGNAL RESISTANCE

$$R_{eq} \approx R_m \parallel R_p = \frac{V_{DD} - V_{out}}{I_m} \parallel \frac{V_{DD} - V_{out}}{V_p}$$

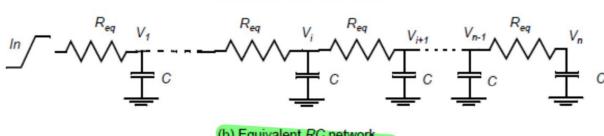
we can approximate (from our transistors):

$$R_{eq} = \frac{1}{2} R_{eq\text{ CMOS}} \frac{(W/L)_m^2}{(W/L)_p^3} \approx 5k\Omega$$

WHAT ABOUT THE DELAY OF A CASCADE OF N TRANSMISSION GATE?



(a) A chain of transmission gates

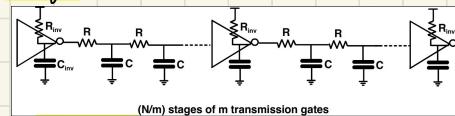


(b) Equivalent RC network

by considering a capacitance of 4ff for chosen sizing we end up with an equivalent RC model:

$$\tau_p \approx \ln(2) R_{eq} C \frac{N(N+1)}{2} \quad \text{d) N}$$

also here we can BREAK THE CHAIN with BUFFERS every m TGs.



(N/m) stages of m transmission gates

$$\tau_p \approx \ln(2) \frac{N}{m} \left[R_{inv} (2C_{inv} + mC) + \frac{RC}{2} \ln(m+1) + mRC_{inv} \right]$$

$$\partial \tau_p / \partial m = 0 \Rightarrow m_{opt} \approx \sqrt{\frac{4R_{inv}C_{inv}}{RC}}$$

but practically, to save power, every 3/4.

STATIC LOGIC: output NODE always connected to V_{DD} or GND through a low impedance path

DYNAMIC LOGIC: features HIGH IMPEDANCE at the output NODE in the HIGH-STATE

it's similar to PSEUDO-NMOS since it **AIMS TO REDUCE THE INPUT CAPACITANCE AND OVERALL AREA** by replacing the PMOS TRANSISTORS with a SINGLE PULL-UP DEVICE.

ALSO OUTPUT CAPACITANCE IS REDUCED, we obtained FASTER LOGIC GATES.



two phases:

- ① **PRE-CHARGE** at $CLK = 0$ only $0 \rightarrow 1$
- ② **EVALUATION** at $CLK = 1$ only $1 \rightarrow 0$

• When $CLK = 0$ the output node is charged to V_{DD} through the PMOS transistor. NMOS is OFF, pull-down path disabled.

PULL-UP OCCURS ALWAYS

No static current from supply to ground
output capacitance is parasitic (related to MOS connected, wiring, input of fan-out gates)

• for $CLK = 1$, pull-up is turned off and pull-down device is activated.

OUTPUT IS DISCHARGED ONLY IF LOGICAL INPUTS ACTIVATES THE PULL-DOWN NETWORK OR OTHERWISE IT REMAINS HIGH

SPEED and Power:

• **REDUCED AREA**

• **INCREASED SPEED** due to:

- reduced C_{out}
- absence of cross-conduction current

$$t_{PUL} = 0$$

t_{PUL} benefits of **REDUCED C_{out}** (however t_{PUL} slightly increased due to higher pre-charge V_{DD})

~ about 3/4 of t_{PUL} (ratio of C_{out} and C_{in})

• **DYNAMIC** is advantageous for **POWER** because

- smaller C_{in} (lower dynamic power)
- NO short circuit current
- smaller input capacitance. only NMOS transistors

HOWEVER

• **ADDITIONAL SOURCE OF POWER CONSUMPTION**

• **LARGER SWITCHING ACTIVITY** $d_{sw} = p(0)$

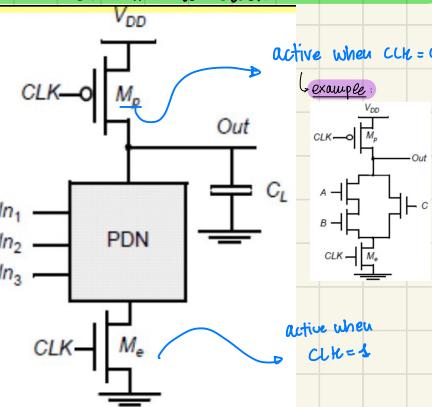
$$(V_{out} \downarrow t_{sw} = p(0)p(1) <)$$

DYNAMIC LOGIC

→ to the **MAIN DRAWBACKS OF THE RATIOED LOGIC**.

- 3) a **STATIC POWER CONSUMPTION** due to the fact that the pull-up device is always ON and contrast the pull-down network when the latter is ON.
- 2) a **SLOW PULL-UP TIME** since the PMOS device cannot be too much conductive
- 3) a low output voltage larger than OV that causes **REDUCED NH**.

SOLVE BY USING a CLOCHED PULL-UP DEVICE



$$\text{imperfections: } OUT = \overline{CLK} + CLK [AB'C]$$

• only **N+2 TRANSISTORS**

• it's **NON-RATIOED** (as in PSEUDO-NMOS)

PMOS size can be chosen to speed up the pre-charge phase (HOWEVER $Cap \uparrow t_{PUL} \uparrow$ and $Power diss. \uparrow t_{PUL} \uparrow$)

$$V_{dd} = GND, V_{oh} = V_{dd} \text{ ALWAYS}$$

• **NO STATIC POWER CONSUMPTION** since pull-up and pull-down are never on simultaneously

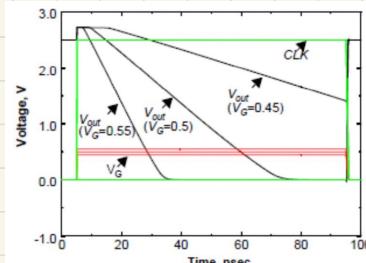
• **FASTER LOGIC** than static.

→ no conflict between pull-up and pull-down which instead happens both in FC-CHS if input does not change naturally and PATED-logic where pull-up is always ON and limiting the available load current.
(→ in PATED t_{PUL} is intrinsically slow because for different reasons pull-up device cannot be made large)

• **NOISE MARGINS** cannot be assessed since they are static metrics (VTC curve cannot be assessed statically)

IF INPUT VOLTAGE $< V_{th}$ OUTPUT REMAINS HIGH.

OUTPUT VOLTAGE still depends on DURATION OF EVALUATION PERIOD

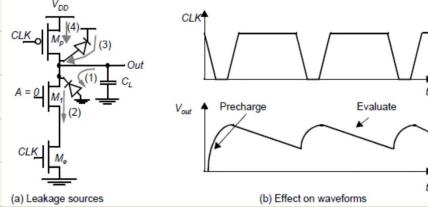


We can find V_{th} (V_{th} voltage considered 'low') equal to V_{thm}

LOW VALUE OF LOW-ROSE HAVING NML

let's see other ISSUES OF DYNAMIC CIRCUITS:

I) LEAKAGE CURRENT: because of SUB-TRESHOLD CURRENT OF NMOS and DRAIN JUNCTION



- final output voltage depends on the current flowing into the RESISTIVE DIVIDER, composed by pull-down and pull-up paths

↳ can bring a LOGIC ERROR

HOW TO SOLVE? use a BLEEDER keeps the node at V_{DD} during the evaluation if the OUTPUT HAS TO STAY HIGH.

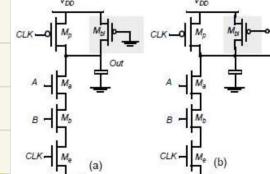
TWO POSSIBLE SOLUTIONS:

- ADD PHOS (like PSEUDO NMOS)

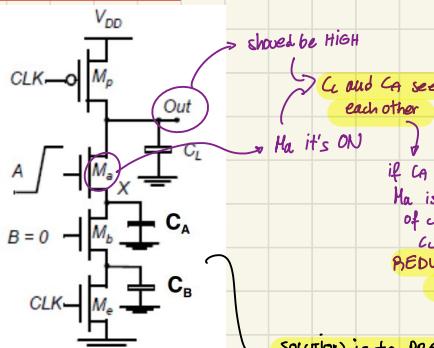
Strong enough for pull-up weak in pull-down

→ IT'S PARTIED

- PHOS in FEEDBACK
Also here PHOS weak enough to allow pull-down



II) CHARGE SHARING:



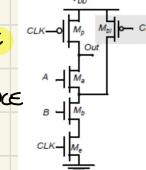
SOLUTION IS TO PRECHARGE INTERNAL NODE

$$\Delta V_{out} = V_{DD} - V_{out} = \frac{C_A}{C_A + C_B} V_{DD} \geq V_{min}$$

if $\frac{C_A}{C_A + C_B} > \frac{V_{min}}{V_{DD} - V_{min}}$

$$\Delta V_{out} = " " = \frac{C_B}{C_A + C_B} (V_{DD} - V_{min}) \leq V_{min}$$

if $\frac{C_B}{C_A + C_B} < \frac{V_{min}}{V_{DD} - V_{min}}$



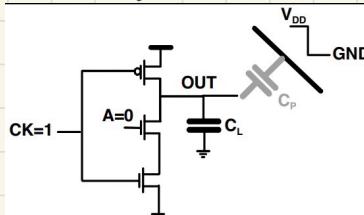
but
AREA ↑
POWER CONST
(more GATES connected to CLK)

III) CAPACITIVE COUPLING:

an HIGH-Z NODE is very sensitive to INTERFERENCE and CROSS-TALK.

For ex. if a wire is capacitively coupled to the output node of a dynamic gate transitioning from high to low, it can cause a loss of CHARGE in the floating output node (at high-Z)

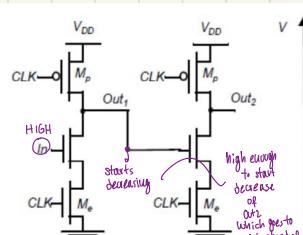
→ the larger the parasitic capacitance, the larger the voltage drop.



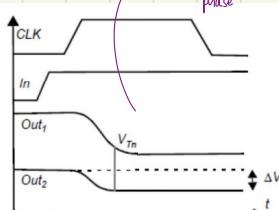
$$\Delta V_{out} = \frac{C_P}{C_P + C_L} V_{DD}$$

↳ CLOCK FEEDTHROUGH: aggressor is clock signal (cap of pre-charge device)

IV) CASCADING IS NOT POSSIBLE



Correct value CANNOT BE RECOVERED since pull-up cannot occur in evaluation phase



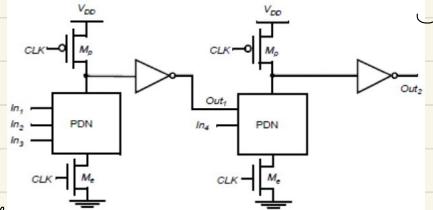
two possible SOLUTIONS:

- DOMINO LOGIC
- hp-CHOS DYNAMIC LOGIC

to ENABLE CASCADING

DOMINO LOGIC

CONSIST IN INSERTING A STATIC INVERTER after a DYNAMIC GATE



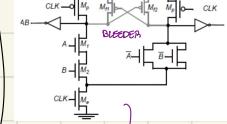
→ also INVERTER HAS THE SUBSEQUENT DYNAMIC GATE IS DRIVEN BY LOW-IMPEDANCE, THUS INCREASING RELIABILITY

DURING EVALUATION PHASE THE OUTPUT OF THE 1ST GATE EVENTUALLY ACTIVATES THE PULL-DOWN OF THE 2ND GATE WITH A DELAY, AND SO ON.

ONLY NON-INVERTING FUNCTION CAN BE IMPLEMENTED

We could eliminate the NMOS evaluation MOS to increase pull-down drives but PROBLEM WITH CASCADE (see transcript)

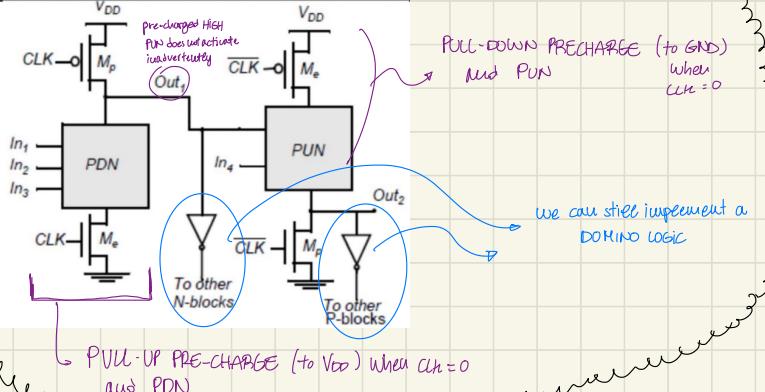
to solve DUAL RAIL DOMINO
very similar to DCVSL with the difference that in DCVSL pull-up is always active (here only in PRE CHARGE)



NON RATIOED

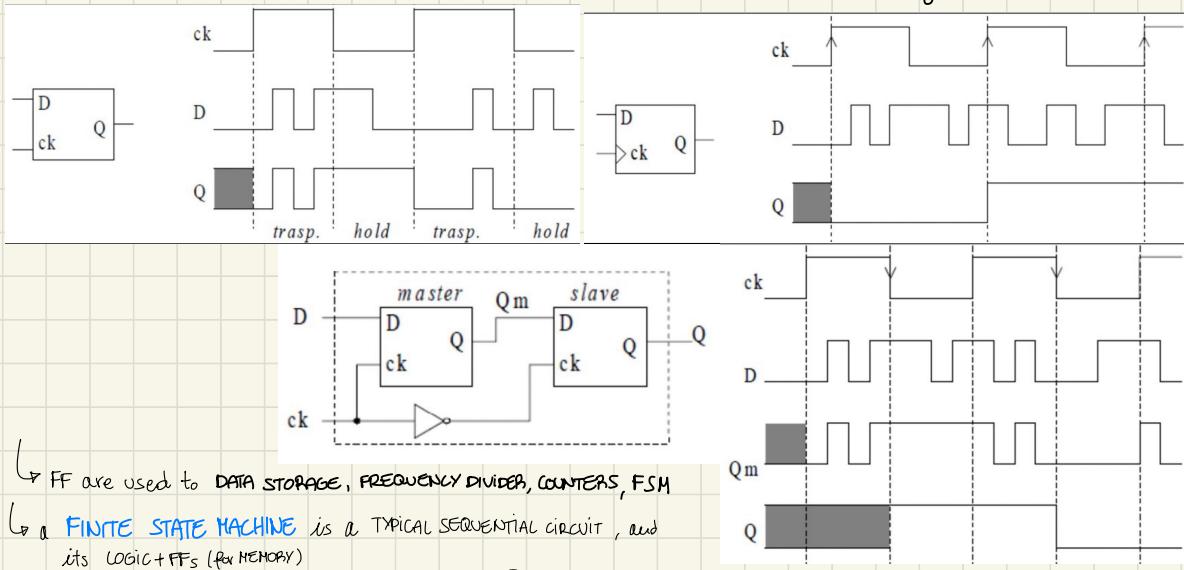
NP - CHOS DYNAMIC

IT'S A CASCADE OF AN N-TYPE DYNAMIC GATE AND A P-TYPE DYNAMIC GATE:



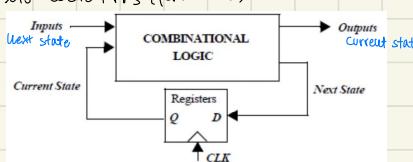
SEQUENTIAL CIRCUITS

- COMBINATIONAL LOGIC CIRCUITS → OUTPUT is a BOOLEAN LOGIC FUNCTION of the inputs.
 - SEQUENTIAL LOGIC CIRCUITS → OUTPUT depend not only on the current values but also PREVIOUS VALUE.
 - it remember the history of the input data. The circuit have a STATE.
- We need a MEMORY ELEMENT: a. LATCH; b. FLIP-FLOP
-
- LATCH:** level-triggered
Memory element sensitive to the LEVEL of the CLOCK
- FF:** edge-triggered
sensitive to the EDGE of the CLOCK
- THE STORED DATA is the one present at the input terminal in correspondence to the edge of the clock.



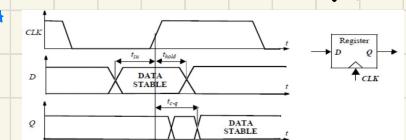
↳ FF are used to DATA STORAGE, FREQUENCY DIVIDER, COUNTERS, FSM

↳ a **FINITE STATE MACHINE** is a TYPICAL SEQUENTIAL CIRCUIT, and its LOGIC + FFs (for MEMORY)

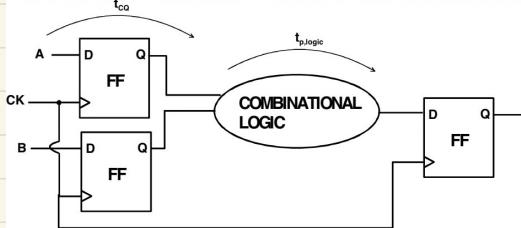


TIMING CONSTRAINTS:

- 1) **SET-UP TIME T_{su}** : time for which input data MUST be valid BEFORE the sensitive edge of the clock
- 2) **HOLD-TIME T_{hold}** : time for which input data must be STABLE AFTER the edge of the clock. (can be negative)
- 3) **PROPAGATION DELAY T_{pd}** : from $V_{DD}/2$ of clock edge to $V_{DD}/2$ of signal (to correct value)
- 4) **CONTINUATION DELAY**: time needed for the output to start changing.



Given a generic sequential circuit:



MEMORY DEVICES

→ **STATIC**: adopts active circuits
for ex. couple of inverters connected in positive loop
DATA IS STORED AS LONG AS SUPPLY VOLTAGE IS APPLIED

→ **DYNAMIC**: stores data on a capacitor left floating in the hold or memory phase
→ REDUCE POWER CONSUMPTION
→ NEED OF REFRESH

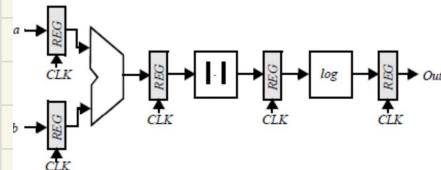
$$T_{CQ} + T_{P, \text{logic}} \leq T_{Cyc}$$

let the signal propagate

$$T_{CQ} + T_{P, \text{logic}} \geq T_{Hold}$$

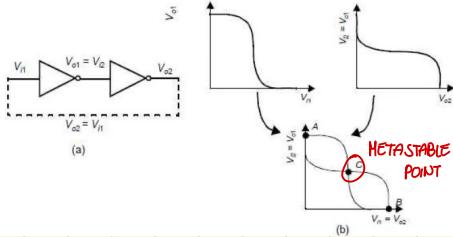
Logic cannot propagate to rapidly

Pipeline Techniques:



STATIC MEMORY DEVICES

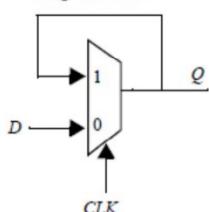
I based on a COUPLE OF INVERTERS CONNECTED IN A POSITIVE LOOP FEEDBACK FASHION. It has only two STABLE STATES (**BISTABLE**):



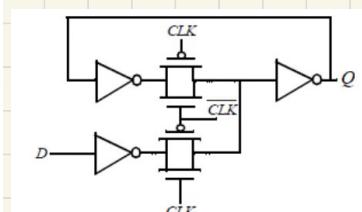
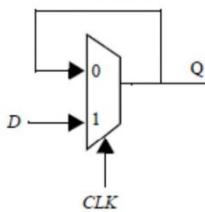
• we can WRITE by applying a PULSE with duration longer than delay loop ($2 \times \tau_{IN}$)

II MULTIPLEXER STATIC LATCH : Use of TRANSMISSION GATE MULTIPLEXER → LATCH can be opened to write data and then CLOSED to STORE IT.

Negative Latch

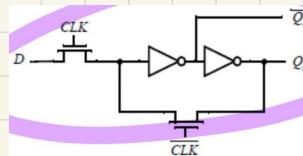


Positive Latch



→ since the clock has SWITCHING ACTIVITY = 1 the number of transistors connected to CLK is very important (impacts on POWER CONS.)

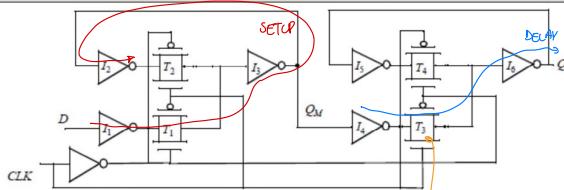
ALTERNATIVE ↗



→ DRAWBACK: it pass a higher high voltage $V_{DD} - V_{THN}$

longer propagation delay
reduced NM
STATIC POWER CONSUMPTION

MASTER-SLAVE FF: CASCADE OF POSITIVE AND NEGATIVE LATCH



$$T_{\text{Setup}} = 3 \tau_{\text{INV}} + \tau_{\text{TGS}}$$

$$T_{\text{Delay}} = \tau_{\text{INV}} + \tau_{\text{TGS}}$$

$$T_{\text{Hold}} = -\tau_{\text{INV}}$$

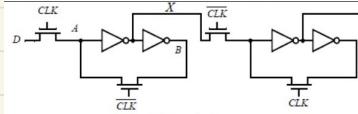
CLK ↑ TURNS TGS OFF

ANY CHANGE OF INPUT DOES NOT CHANGE FF STATE
BECAUSE OF τ_{INV} OF INV1 D CAN CHANGE ALSO AFTER RISING EDGES WITHOUT BEING SAMPLED

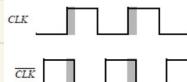
?
negative

OVERLAP OF THE CLOCK:

due to different paths of clock signal



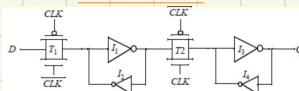
(a) Schematic diagram



(b) Overlapping clock pairs

DRAWBACK:
lot of transistors

RATIOED SOLUTION:



- BOTH MASTER AND SLAVE ARE TRANSPARENT

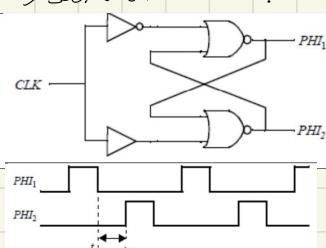
OUTPUT CHANGE IN WRONG EDGE IF INPUT CHANGE DURING OVERLAP (CRITICAL RACE)

- in overlap phase $Q_L = 1$ and $Q_H = 1$ the first latch is active during the sampling phase.

(NODE A) is function of STRENGTH OF INVERTER CONNECTED TO NODE B AND OF THE DRIVER OF NODE D)

(STATE CAN BE UNDEFINED)

⇒ SOL: USE TWO CLOCKS!



DYNAMIC LATCHES AND REGISTERS 3

STATIC MEMORY → COMPLEX. HIGH POWER CONS

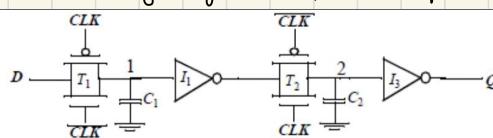
DYNAMIC

store data across a PARASITIC CAPACITANCE
presence of charge '1'. absence = '0'
Charge may be lost if we don't refresh.

$$T_{\text{Setup}} = t_{\text{TGS}}$$

$$T_{\text{Hold}} = 0$$

$$T_{\text{Delay}} = t_{\text{inv}} + t_{\text{TGS}} + t_{\text{inv}}$$



↳ each overlap is a problem: a signal that occurs just after falling edge of CLK can propagate to the output node.

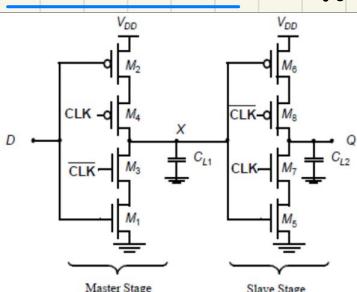
$$t_{\text{overlap}} \leq t_{\text{TGS}} + t_{\text{inv}} + t_{\text{TGS}}$$

for overlap 1-1

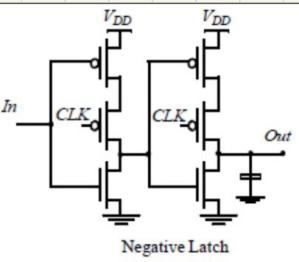
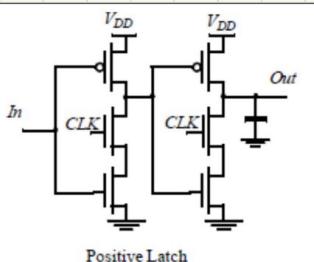
$$t_{\text{overlap}} \leq t_{\text{hold}}$$

C²MOS DYNAMIC FF: The two patches are tri-state inverters. Output can be '1', '0', or HIGH-Z depending on inputs and CLK.

o IT'S INSENSITIVE TO OVERLAP: PDN and PUN never on at same time

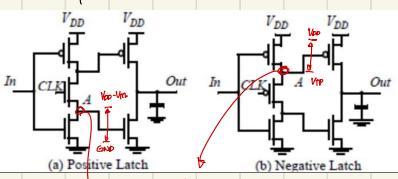


TRUE SINGLE PHASE CLOCK: adopt only one phase of CLK relying on fact that NMOS and PMOS are ON for different gate voltages



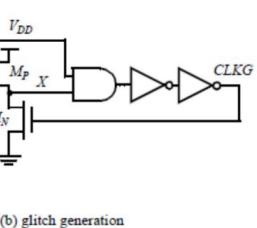
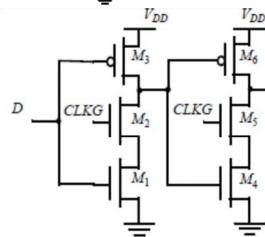
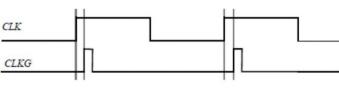
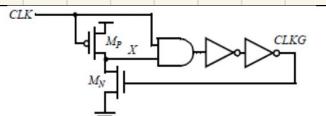
- if CLK is high circuit reduces to 2 cascaded inverters.
- if CLK is low both inverters are disabled
- only two PWN are ON, so central node of latch can be charged up by the first PMOS transistor, but 2nd PMOS is OFF
- more MOS used but clock is single phase

↳ can we further reduce MOSFETs



PULSED FLIP-FLOPs: adopt a latch with a pulse clock

Short clk samples incoming signal in a small window.



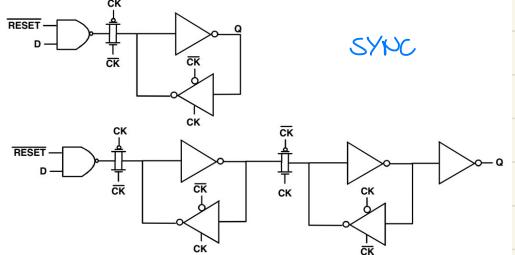
$$T_{\text{SETUP}} = 0$$

$T_{\text{HOLD}} = \text{pulse duration}$

$T_{\text{DELAY}} = \text{delay of inverters}$

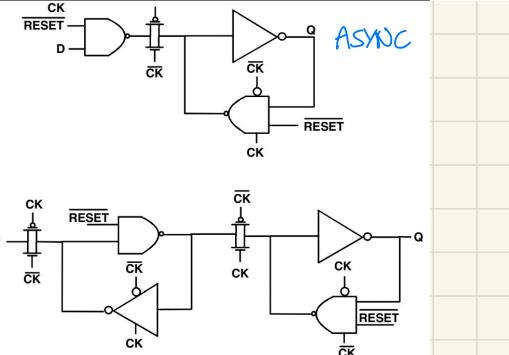
- reduced number of transistors connected to CLK and power consumption
- glitch generation circuit is shared
- substantial increase in verification process

RESETTABLE LATCHES and FF: RESET needed for initial state. Can be SYNC or ASYNC



SYNC

change immediately
force change at edge



ASYNC

D -> CK -> RESET -> CK -> Q