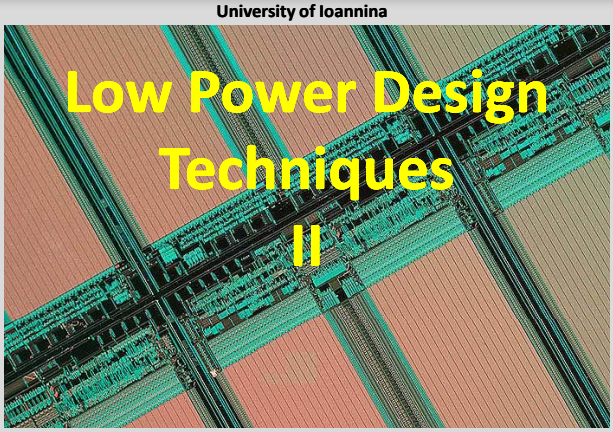
CMOS INTEGRATED CIRCUIT DESIGN TECHNIQUES

University of Ioannina



Dept. of Computer Science and Engineering

Y. Tsiatouhas

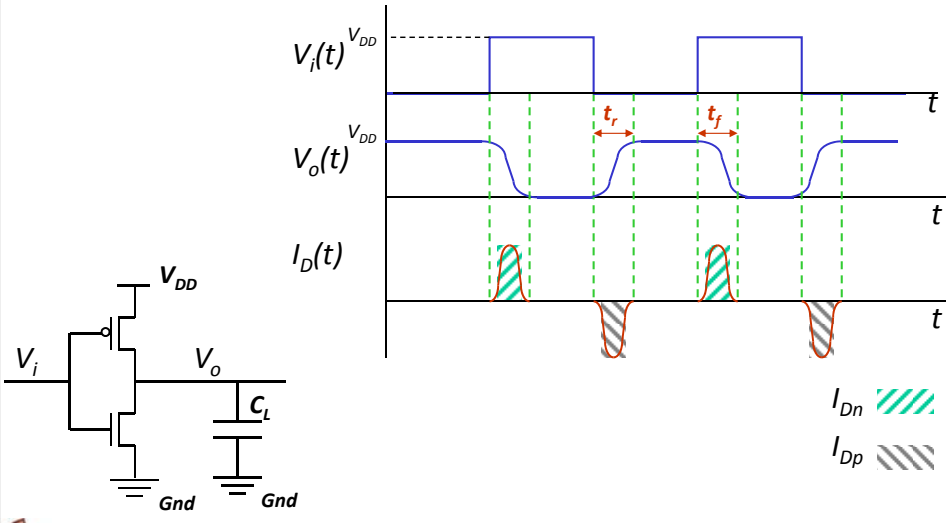
CMOS Integrated Circuit Design Techniques

VLSI Systems
and Computer Architecture Lab

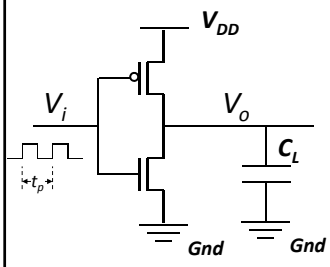
Overview

1. ***Dynamic power consumption reduction***
2. ***Architecture level power reduction***
3. ***Short circuit power consumption reduction***

Dynamic Power Consumption I



Dynamic Power Consumption II



The mean dynamic power consumption for an input pulse with frequency $f_p = 1/t_p$ is provided by:

$$P_d = \frac{1}{t_p}\int_0^{t_p/2} i_n(t)V_o(t)dt + \frac{1}{t_p}\int_{t_p/2}^{t_p} i_p(t)(V_{DD} - V_o(t))dt$$

Given that: $$i(t) = C\frac{dV(t)}{dt}$$

$$P_d = \frac{C_L}{t_p}\int_0^{V_{DD}} V_o dV_o + \frac{C_L}{t_p}\int_{V_{DD}}^{0} (V_{DD} - V_o)d(V_{DD} - V_o) = \frac{C_L \cdot V_{DD}^2}{t_p} = f_p \cdot C_L \cdot V_{DD}^2$$ ✓

Dynamic Power Consumption Reduction

Technology Scaling

1

- A new CMOS technology every 2-3 years.
- Technology scales down by a factor $\kappa=1/\alpha\approx0.7$
  - $W=\kappa W$, $L=\kappa L$
  - $V_{DD}=\kappa V_{DD}$, $V_{th}=\kappa V_{th}$
  - $t_{ox}=\kappa t_{ox}$
  - $C_{ox}=\varepsilon_{ox}/t_{ox}=C_{ox}/\kappa$
  - $I_D=\mu_n C_{ox} W/L(V_{DD}-V_{th})^2=(1/\kappa)(\kappa/\kappa)(\kappa)^2 I_D=\kappa I_D$
  - $C_{gate}=C_{ox}WL=(1/\kappa)(\kappa)(\kappa)C_{gate}=\kappa C_{gate}$
  - C_{wire} ($\propto L$) $=\kappa C_{wire}$
- Delay $=CV_{DD}/I_D=(\kappa\cdot\kappa/\kappa)$Delay $=\kappa$Delay
- Energy $=CV^2_{DD}=\kappa^3$Energy
- Power = Energy / Delay $=\kappa^2$Power
- Energy $\times$ Delay $=\kappa^4$ (Energy $\times$ Delay)

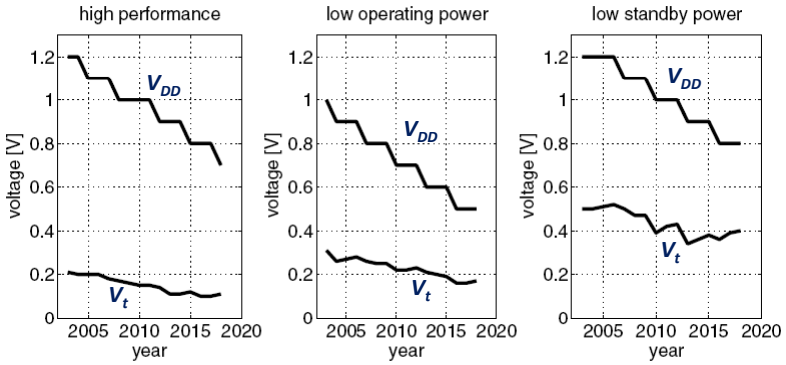
Power Supply V_{DD} Reduction

2 *Since:* $P_d = f_p \cdot C_L \cdot V_{DD}^2$ *the reduction of* V_{DD} *will result in significant power consumption reduction (square low).*

However, the power supply reduction will result in the increment of the circuit signal propagation delays. To confront with this drawback, the following techniques are proposed :

- *Threshold voltage* V_{th} *reduction (but static power is increased).*
- *Use of reduced threshold voltage* V_{th} *only for the gates at time (performance) critical signal paths in the circuit.*
- *Use of multi power supplies or an adaptive power supply* V_{DD}.
- *Exploitation of pipeline design techniques or other parallelism design schemes for the reduction of the signal propagation delay.*

$V_{DD} - V_t$ Scaling Scenarios

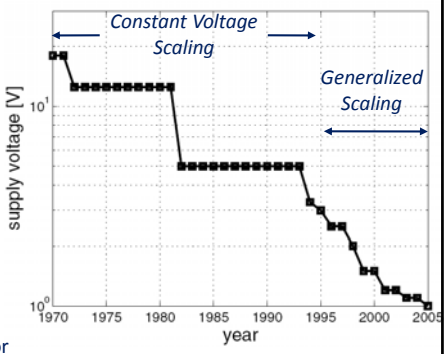


V_{DD} Reduction and Performance

Scaling Strategies

device/circuit parameter	constant voltage scaling	constant field scaling	generalized scaling
channel length	$1/\alpha$	$1/\alpha$	$1/\alpha$
channel width	$1/\alpha$	$1/\alpha$	$1/\alpha$
oxide thickness	$1/\alpha$	$1/\alpha$	$1/\alpha$
area per device	$1/\alpha^2$	$1/\alpha^2$	$1/\alpha^2$
terminal voltages	1	$1/\alpha$	ϵ/α
threshold voltage	1	$1/\alpha$	ϵ/α
internal electric fields	α	1	ϵ
doping	α^2	α	$\epsilon\alpha$
saturation current	α	$1/\alpha$	ϵ/α
gate capacitance	$1/\alpha$	$1/\alpha$	$1/\alpha$
gate delay	$1/\alpha^2$	$1/\alpha$	$1/\alpha$
power dissipation	α	$1/\alpha^2$	ϵ^2/α^2
power density	α^3	1	ϵ^2
power delay product	$1/\alpha$	$1/\alpha^3$	ϵ^2/α^3

α is the scaling factor while α/ε is the reduced scaling factor



Influence of V_{DD} on the current I_D and propagation delay t_d:

$$I_D = k \cdot (V_{DD} - V_{th})^s \qquad t_d = \frac{C \cdot V_{DD}}{k \cdot (V_{DD} - V_{th})^s}$$

where s is the velocity saturation index (1 < s < 2).

Power Supply V_{DD} Scaling (I)

For s = 2

Propagation Deley

$$t_d = \frac{C \cdot V_{DD}}{I_D}$$

$$I_D = k \cdot (V_{DD} - V_{th})^2$$

$$t_d = \frac{C \cdot V_{DD}}{k \cdot (V_{DD} - V_{th})^2}$$

Energy

$$E = C \cdot V_{DD}^2$$

Metric

Energy × Delay

$$E \cdot t_d = \frac{C^2 \cdot V_{DD}^3}{k \cdot (V_{DD} - V_{th})^2}$$

$$E \cdot t_d = \gamma \frac{V_{DD}^3}{(V_{DD} - V_{th})^2}$$

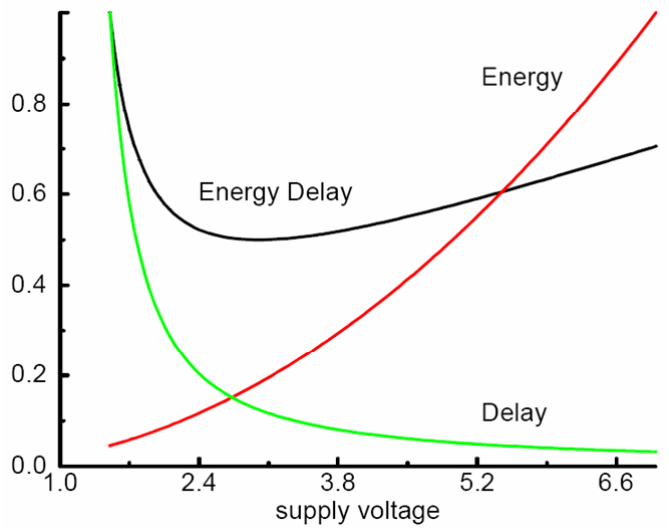
Metric Minimization

$$\frac{d(E \cdot t_d)}{dV_{DD}} = \gamma \frac{(V_{DD} - V_{th})^2 \cdot 3 \cdot V_{DD}^2 - V_{DD}^3 \cdot 2 \cdot (V_{DD} - V_{th})}{(V_{DD} - V_{th})^4} \Rightarrow$$

$$\frac{d(E \cdot t_d)}{dV_{DD}} = \gamma \frac{V_{DD}^2 (V_{DD} - 3V_{th})}{(V_{DD} - V_{th})^3}$$

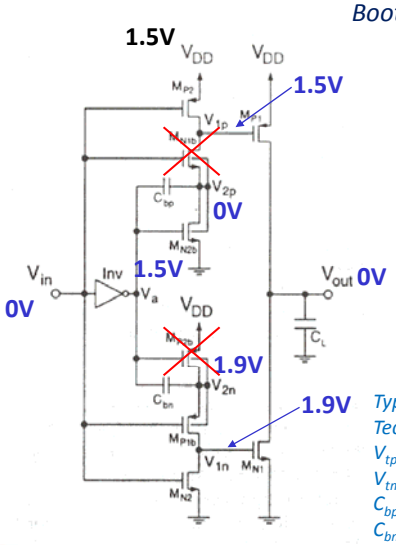
The derivative is vanished for $V_{DD}=3V_{th}$

Power Supply V_{DD} Scaling (II)

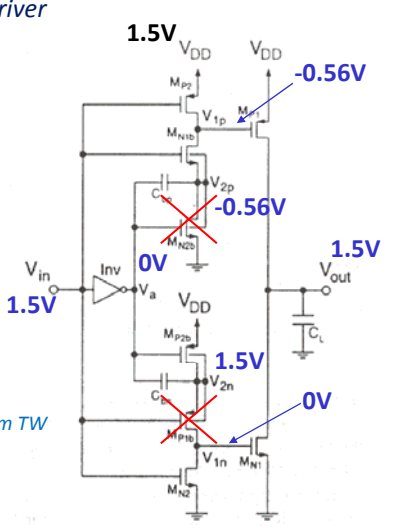


Low Voltage Oriented Design (I)

Bootstrapped Driver

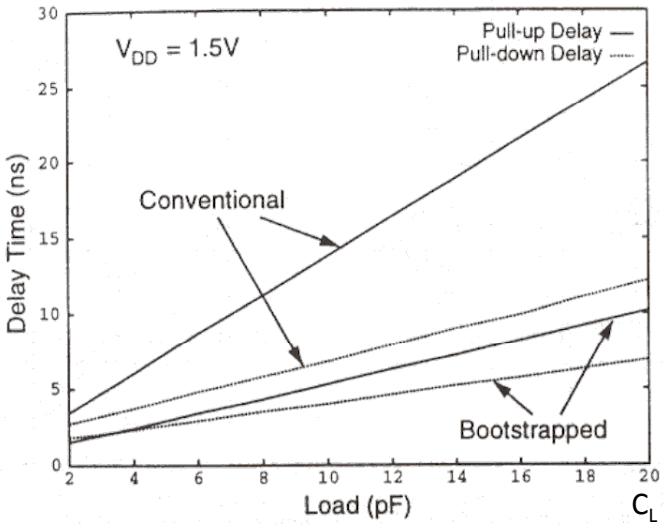


Typical V_{DD}: 5V
Technology: 0.8µm TW
V_{tp}= -0.9V
V_{tn}= 0.75V
C_{bp}= 0.35pF
C_{bn}= 0.25pF



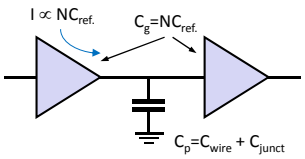
J. Lou, J. Kuo, IEEE J. Solid-State Circuits, 32(1) 1997

Low Voltage Oriented Design (II)



Transistor Size Scaling (I)

3



$$I = \kappa'(NC_{ref}) \cdot (V_{ref} - V_t)^2$$

$$t_d = (C_p + NC_{ref}) \frac{V_{ref}}{I} \Leftrightarrow$$

$$t_d = \kappa \frac{(C_p + NC_{ref}) \cdot V_{ref}}{(NC_{ref}) \cdot (V_{ref} - V_t)^2} \Leftrightarrow$$

$$t_d = \kappa(1 + \alpha/N) \frac{V_{ref}}{(V_{ref} - V_t)^2}$$

$$\alpha = C_p / C_{ref}$$

Let us consider a reference design with transistor sizes W/L so that the input (transistor gate related) capacitance is C_{ref} .
Initially, the power supply is V_{ref}.

Aiming to reduce the dynamic power consumption, we reduce the power supply voltage to V_N.

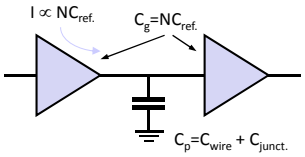
This will result in the increment of the propagation delay time of the circuit.

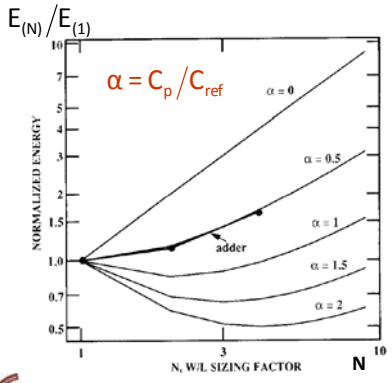
In order to recover the operating speed, we decide to increase (scale up) the transistor sizes in the design by a factor N
(our target is to reduce the propagation – $t_d \downarrow$).

Is this approach applicable and if yes which is the proper N factor?

Chandrakasan et.al., IEEE J. Solid-State Circuits, 27(4) 1992

Transistor Size Scaling (II)





For each N value, there is a proper power supply voltage V_N where both designs (the reference & the scaled one) exhibit the same speed performance.

Given that the propagation delay t_d increases with $1/V_{ref}$ (V_{th}<<)*, the V_N voltage for which the t_d of the scaled design is equal to this of the reference design (N=1), is provided by:

$$V_N = \frac{(1+\alpha/N)}{1+\alpha} V_{ref} \underset{(N>1)}{\leq} V_{ref}$$

$$t_d \approx \kappa(1+\alpha)\frac{1}{V_{ref}} = \kappa(1+\alpha/N)\frac{1^*}{V_N}$$

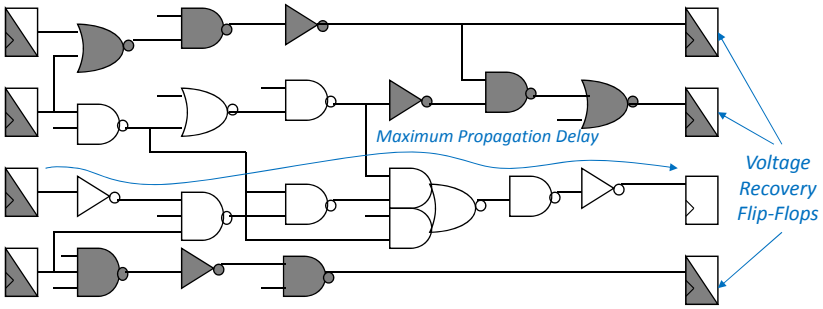
Consequently, for every N, the energy consumption at the first stage is given by:

$$E(N) = (C_p + NC_{ref})V_N^2 = \frac{NC_{ref}(1+\alpha/N)^3}{(1+\alpha)^2} V_{ref}^2$$

Multiple Power Supplies

4

Selective use of a lower than V_{DD} power supply voltage in order to reduce the dynamic power consumption.



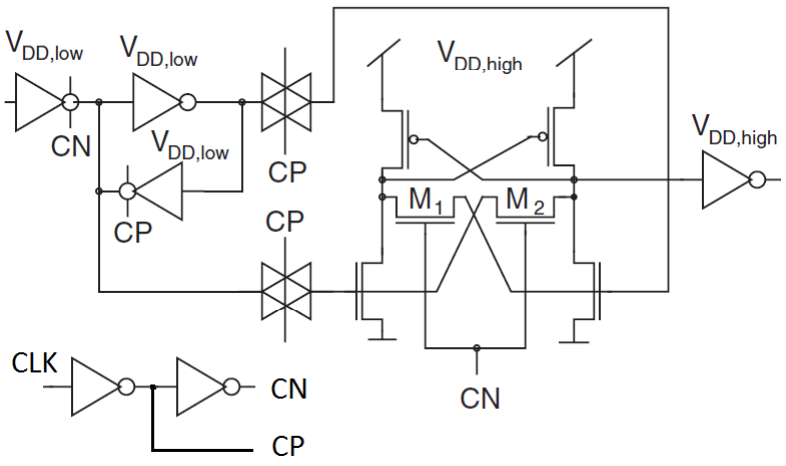
Use of a single threshold voltage V_{th} for the transistors in the design

The shaded logic gates are fed with a lower voltage V_L. The shaded Flip-Flops recover the signal voltage level $V_H \equiv V_{DD}$ at the output.

Note that when a logic block is fed with a lower than V_{DD} voltage, then every subsequent block up to the recovery Flip-Flops must be fed with the same voltage level.

Voltage Level Recovery Flip-Flop

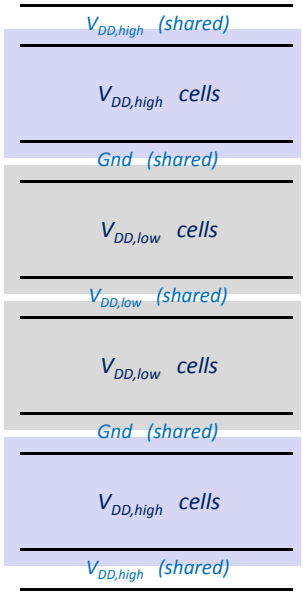
Slave-Latch Level-Conversion Flip-Flop

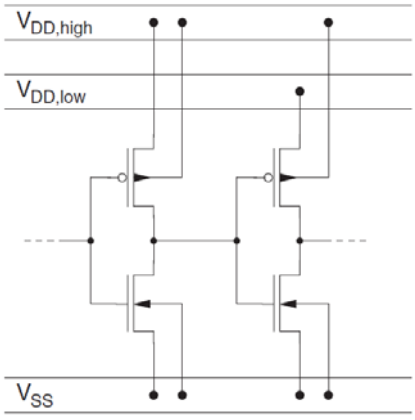


Zhang et.al., IEEE ISLPED, 1998

Implementation of a Multi-V_{DD} Scheme

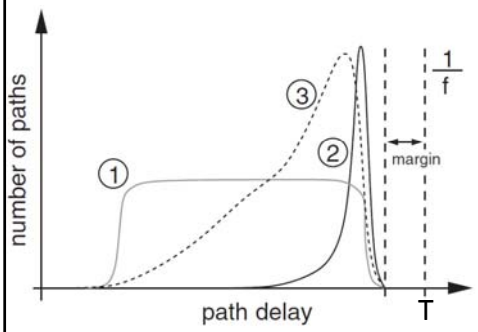
(a)



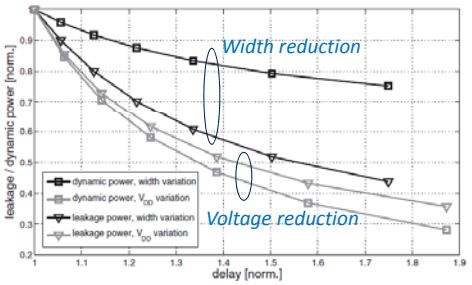


(b)

The Power Saving Capability



Case of an inverter driving 10fF load.



Circuits with the distribution (2) can be the result of a timing driven place and route tool with power optimization. These tools reduce the transistors' widths to gain in power.

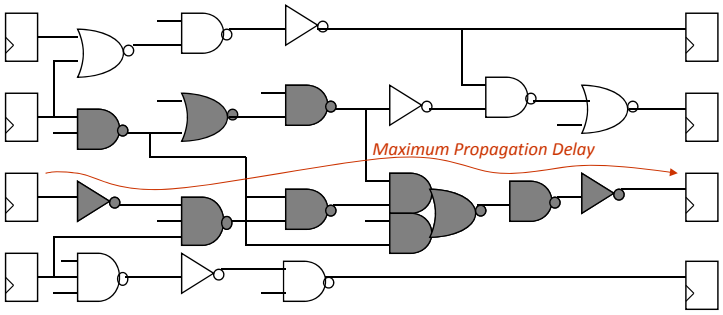
However, which is the best strategy? To reduce the transistors' widths in the sub-critical paths using a single power supply in the whole design or to reduce the power supply for the gates of the sub-critical paths leaving the transistors' widths unaltered?

The second approach seems to provide better results !

Multiple Threshold Voltage Technologies

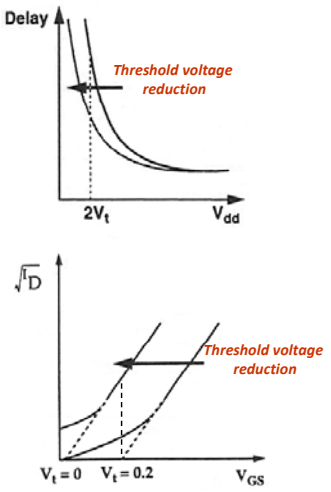
5

Global use of a lower than V_{DD} power supply voltage in the design in order to reduce the dynamic energy consumption.

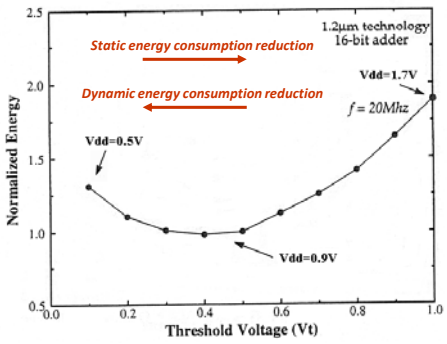


The shaded logic blocks have been designed using transistors with lower threshold voltage than the nominal technology threshold voltage, in order to retain the speed performance of the design.

Threshold Voltage Impact on Energy



Fixed Throughput @ 20MHz



Chandrakasan et.al., Proceedings of IEEE, 83(4) 1995

V_{th} Reduction Issues

- Static power consumption – increment of the weak-inversion (sub-threshold) leakage current.
- Circuit operation susceptibility on the V_{th} variations.

Demand for:

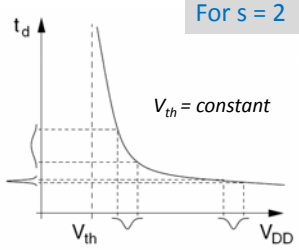
- *increased threshold voltage V_{th} at the idle (stand-by) state*
- *adaptable threshold voltage V_{th} at the normal mode of operation by adjusting the substrate bias.*

Dynamic & static power consumption reduction techniques

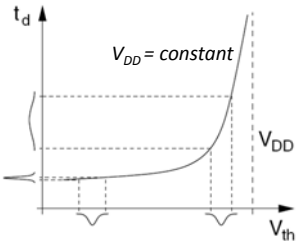
	Active	Stand-by
Multiple V_{th}	Dual-V_{th}	MTCMOS
Variable V_{th}	V_{th} hopping	VTCMOS
Multiple V_{DD}	Dual-V_{DD}	Reverse V_{GS} switch
Variable V_{DD}	V_{DD} hopping	

Impact of V_{DD} & V_{th} in Propagation Delay

For s = 2



$$t_d = \frac{C \cdot V_{DD}}{k \cdot (V_{DD} - V_{th})^2}$$



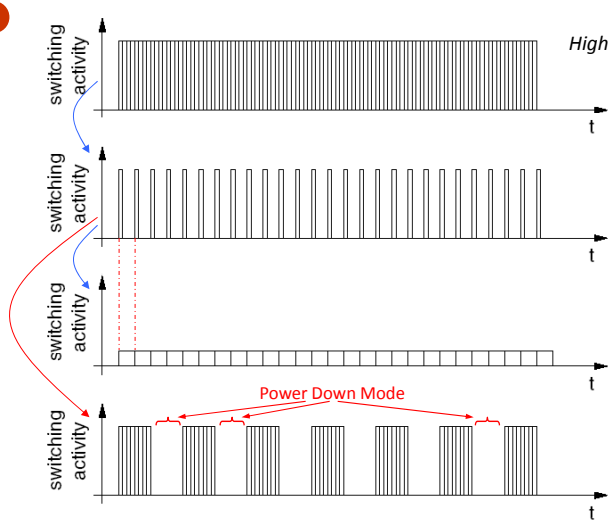
$$\frac{\partial t_d}{\partial V_{DD}} = -\frac{C}{k} \cdot \frac{V_{DD} + V_{th}}{(V_{DD} - V_{th})^3}$$

$$\frac{\partial t_d}{\partial V_{th}} = \frac{C}{k} \cdot \frac{2V_{DD}}{(V_{DD} - V_{th})^3}$$

The susceptibility of the propagation delay on V_{DD} and V_{th} variations is increased as V_{th} is increased and V_{DD} is decreased.
The propagation delay variations is significantly increased for high V_{th} and low V_{DD} levels.
This effect must be seriously considered in designs where multi threshold voltage and/or multi power supply techniques are used.

Dynamic Voltage-Frequency Scaling – DVFS

6



High load & high performance operation
f_{max} *and* V_{DDmax}

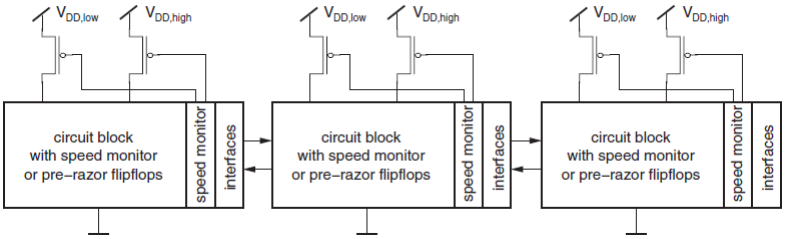
Low load operation
f_{min} *and* V_{DDmax}

Low dynamic power operation
f_{min} *and* V_{DDmin}

Low static power operation
f_{max} *and* V_{DDmax}

The Pair of Sources Concept

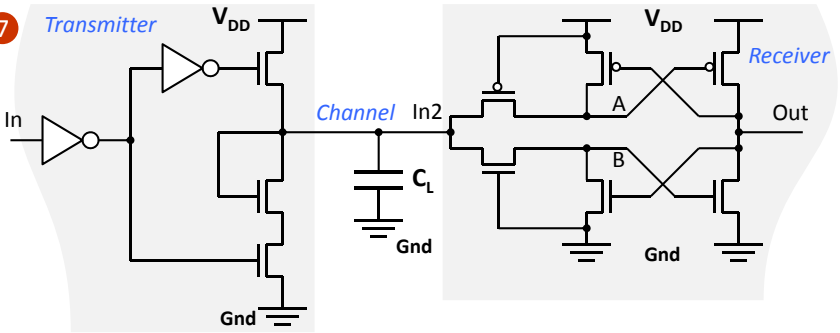
Local power supply adjustment

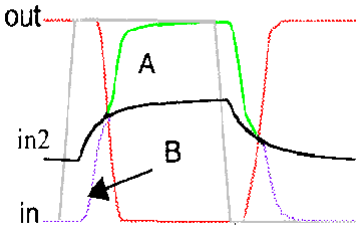


Each circuit block can operate either with the high-V_{DD} for high performance or with the low-V_{DD} for low power.

Yield improvement option: the whole circuit may work in the low power mode (low-V_{DD}) except possibly this circuit block(s) where the performance is affected by local variations (slow block(s)). In that case the circuit block is fed by the high-V_{DD} while the overall power consumption remains low.

Signal Swing Reduction





The signal driver (transmitter) generates signals with swing between [V_{thn}, V_{DD}–V_{thn}].

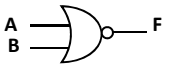
The sense amplifier (receiver) recovers the signal swing between [0, V_{DD}].

$$P = f \cdot C_L \cdot (V_{DD} - 2V_{thn})^2$$

Zhang et.al., IEEE ISLPED, 1998

Logic Family Selection

8



A	B	F
0	0	1
0	1	0
1	0	0
1	1	0

In case that:

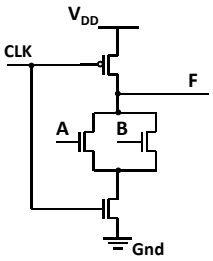
p(A=1)=1/2 & p(B=1)=1/2

$\Rightarrow$

p(F=1)=1/4

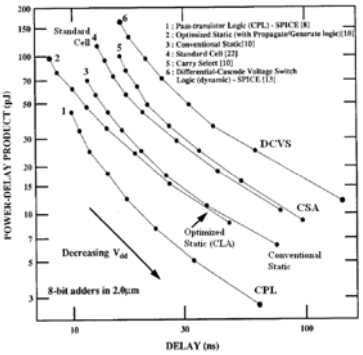
&

$p_{0\rightarrow1}$ = p(F=0)=3/4



NOR Dynamic Gate

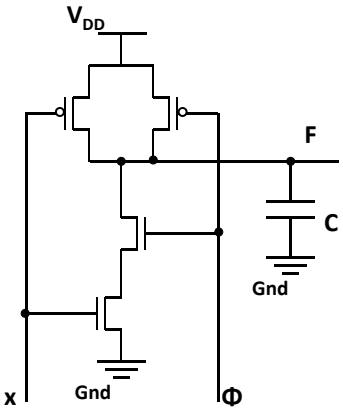
$C_{EFF}=3/4C_L$

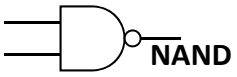


9

Energy Recovery

Energy Consumption (I)





The required charge Q in order to charge the output node F (x="1", Φ="0") is given by the next equation :

$$Q = CV_{DD} \qquad (1)$$

The required energy from the power supply is :

$$E = QV_{DD} = CV^2_{DD} \qquad (2)$$

Considering that the conducting pMOS transistor and the capacitance C form an RC network, the charging current as a function of time is:

$$i(t) = \frac{dq}{dt} = \frac{V_{DD}}{R} e^{-\frac{t}{RC}} \qquad (3)$$

Energy Consumption (II)

The instantaneous power dissipation in the pMOS transistor is:

$$P(t) = i^2(t)R = \frac{V^2_{DD}}{R} e^{-\frac{2t}{RC}} \qquad (4)$$

The energy dissipation within a time duration T in order to charge the output node, is given by:

$$E_{diss-p} = \int_0^T P(t)dt = \int_0^T \frac{V^2_{DD}}{R} e^{-\frac{2t}{RC}} dt = \frac{1}{2} CV^2_{DD} \left(1 - e^{-\frac{2T}{RC}}\right) = \frac{1}{2} CV^2_{DD}\Big|_{T\to\infty} \qquad (5)$$

Consequently, the energy stored on C is provided by the following difference:

$$E_{bit} = E - E_{diss-p} = CV^2_{DD} - \frac{1}{2} CV^2_{DD} = \frac{1}{2} CV^2_{DD} \qquad (6)$$

Energy Consumption (III)

In the next transition of the clock signal Φ (x="1", Φ="1") the stored energy on C will change from E_{bit} to 0 and the difference between the initial and the final energy is dissipated on the equivalent resistances of the nMOS transistors that discharge the output node. Thus, the total energy dissipation in a clock cycle is:

$$E_{cycle} = E_{diss-p} + E_{diss-n} = \frac{1}{2}CV_{DD}^2 + \frac{1}{2}CV_{DD}^2 = CV_{DD}^2 = 2E_{bit} = E \qquad (7)$$

Consequently, the energy that is recovered back to the power supply within a clock cycle is zero (0):

$$E_{rtn} = 0$$

Energy Recovery Principle (I)

The previous discussion shows that during the charging of a node, the current asymptotically approaches 0 and the energy dissipation asymptotically approaches $CV^2_{DD}/2$. After a time duration of 3RC the output voltage is at the 95% of its final voltage, the current is less than 5% of its initial value and the energy dissipation is at the 97.5% of its final value. Thus, after another 3RC time duration the circuit has almost reached its final state, so that for a total time interval of 6RC the transition is considered completed.

Let us replace the power supply with a "step" voltage, where for the time interval from 0 to 3RC the voltage level is $V_{DD}/2$ and from 3RC to 6RC the voltage level is V_{DD}. Then, the energy dissipation will be:

$$E_{diss-p} = \int_0^{3RC} P(t)dt + \int_{3RC}^{6RC} P(t)dt \cong \frac{1}{2}C\left(\frac{V_{DD}}{2}\right)^2 + \frac{1}{2}C\left(\frac{V_{DD}}{2}\right)^2 = \frac{1}{4}CV_{DD}^2 \qquad (8)$$

Energy Recovery Principle (II)

According to the previous analysis, "stepwise" voltage power supplies with more than two steps can be used in order to reduce further the energy dissipation.

The next idea, is to consider a power supply that continuously varies (from $0 \rightarrow V_{DD}$) so that the charging current remains constant throughout the entire process:

$$i(t) = \frac{Q}{T} = \frac{CV}{T} \qquad (9)$$

Then, the charging energy dissipation will be:

$$E_{diss-p} = \int_0^T i^2(t) R dt = R \int_0^T \left(\frac{CV_{DD}}{T} \right)^2 dt = \frac{RC}{T} CV_{DD}^2 \qquad (10)$$

Energy Recovery Principle (III)

The same situation can take place during the discharging phase by continuously varying the Gnd power supply from $V_{DD} \rightarrow 0$, so that the current remains constant. Thus, again the energy dissipation on the nMOS transistor will be:

$$E_{diss-n} = \frac{RC}{T} CV_{DD}^2 \qquad (11)$$

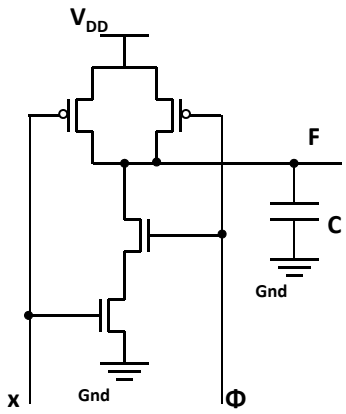
Consequently, the total energy dissipation turns to be :

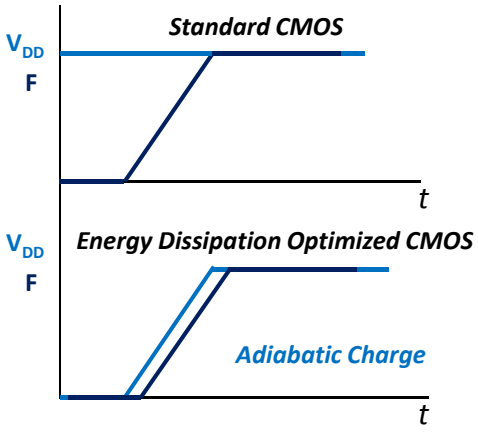
$$E_{diss} = E_{diss-p} + E_{diss-n} = 2 \frac{RC}{T} CV_{DD}^2 \qquad (12)$$

By selecting a proper T, the total energy dissipation can be less than the energy delivered within a cycle when constant power supplies are used : $E=2E_{bit}=CV^2_{DD}$. The difference in the energy is:

$$\Delta E = CV_{DD}^2 - 2 \frac{RC}{T} CV_{DD}^2 > 0 \Big|_{T>2RC} \qquad (13)$$

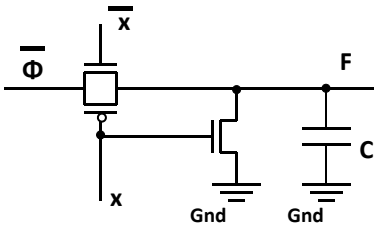
Energy Recovery Principle (IV)





The Adiabatic CMOS Technique

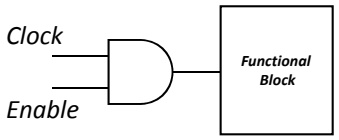
Adiabatic NOT gate



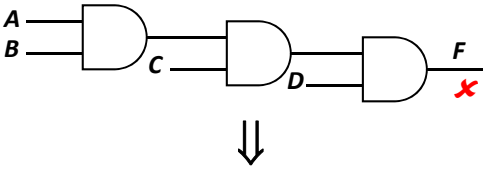
A possible solution in order to exploit the energy recovery principle is to replace the power supply V_{DD} with the clock signal Φ. This is a *power-clock*.

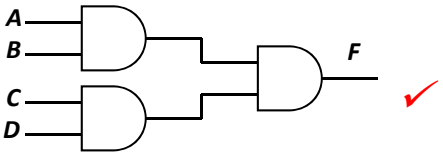
Provided that the power-clock is dynamically adjusted to the constant current requirement an adiabatic charging is achieved and the energy dissipation follows equation (10).

General Low Power Techniques

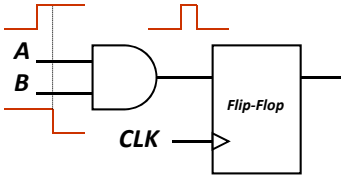


Clock Gating



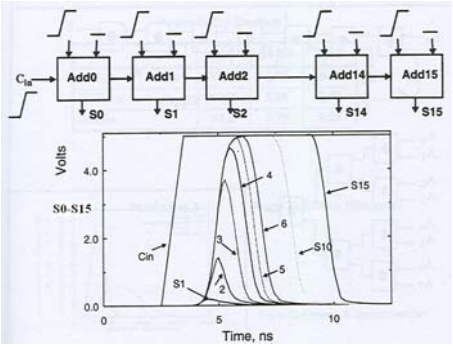


Balanced signal paths

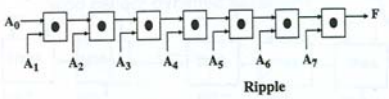


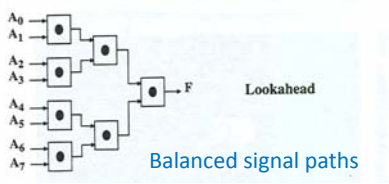
Data Gating

The Adder Example



Glitches or dynamic hazards highly contribute in the dynamic power consumption.



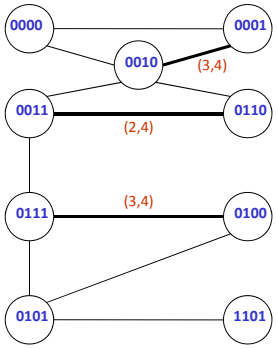


Balanced signal paths

The carry look-ahead design technique for an adder, drastically reduces the presence of glitches with respect to a ripple carry design and consequently the energy dissipation but also improves the speed performance!

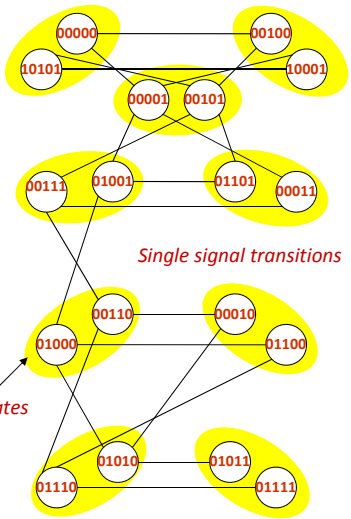
FSM State Assignment

11 The target is to reduce the signal transitions



States duplication
(cost = 1 memory element)

⇒



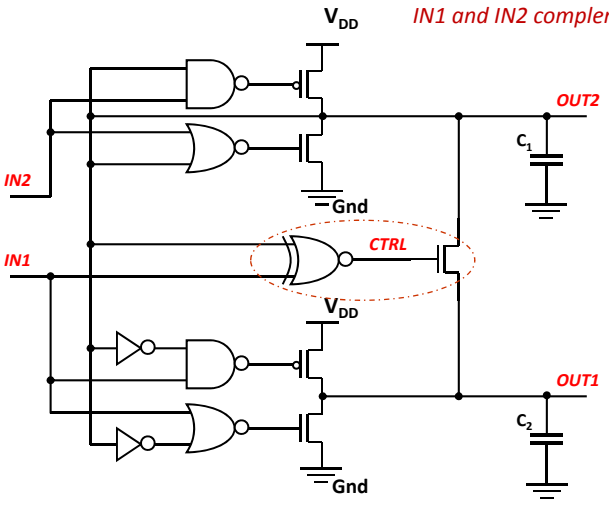
Single signal transitions

Equivalent States

Skias et.al., IEEE ICECS, 2000

Charge Recycling in Signal Drivers

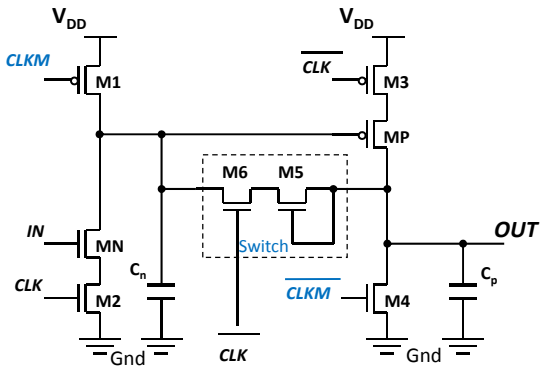
12

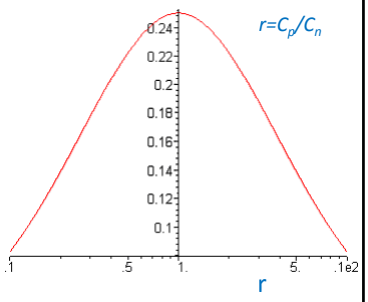


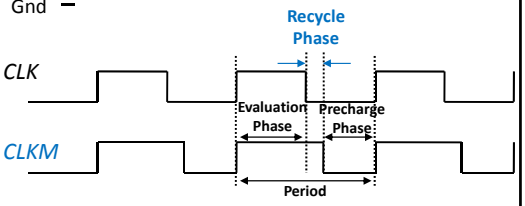
IN1 and IN2 complementary signals.

Bitzaros et.al., IEEE ISCAS, 1997

Charge Recycling in NORA Logic







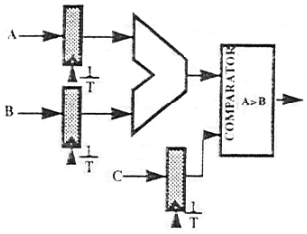
Limniotis et.al., IEEE TCAS-I, 53(12), 2006

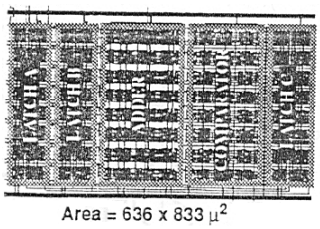
13

Architecture Level Low-Power Design

Architecture Level Options

$f_{ref} = \frac{1}{T}$

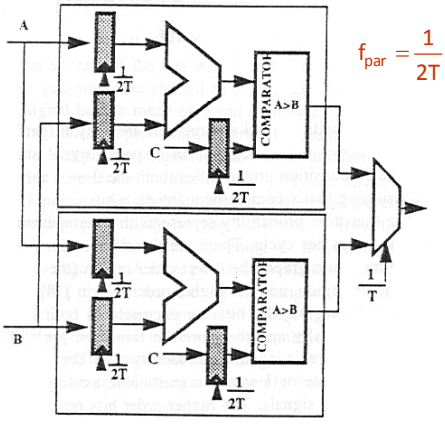




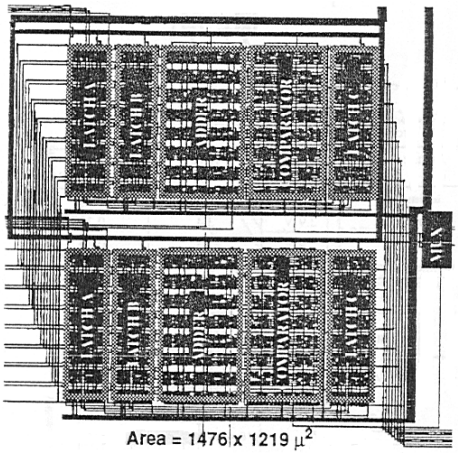
Area = 636 x 833 μ^2

- *Critical path delay: $T = T_{adder} + T_{comp} = 25ns \Rightarrow f_{ref} = 1/T = 40MHz$*
- *Total switched capacitance = C_{ref}*
- *$V_{DD} = V_{ref} = 5V$*
- *Power: $P = f_{ref} C_{ref} V^2_{ref}$*

Parallel Architectures



$f_{par} = \frac{1}{2T}$



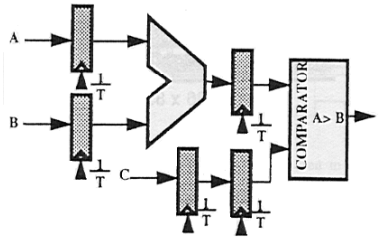
Area = 1476 x 1219 μ^2

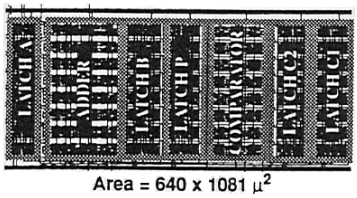
- *$f_{par} = 1/2T = f_{ref}/2$*
- *$C_{par} = 2.15 C_{ref}$*
- *$V_{par} = V_{ref}/1.7$*

Power: $P = f_{ref}/2 \cdot 2.15 C_{ref} \cdot (V_{ref}/1.7)^2 = 0.36 P_{ref}$

keeping the throughput constant!

Pipeline Architectures





- $f_{par}=1/T=f_{ref}$
- $C_{par} = 1.1C_{ref}$
- $V_{par} = V_{ref}/1.7$

Power: $P=f_{ref} \cdot 1.1C_{ref} \cdot (V_{ref}/1.7)^2=0.37P_{ref}$

keeping the throughput constant!

Chandrakasan et.al., IEEE J. Solid-State Circuits, 27(4) 1992

Algorithm Selection

Video Compression

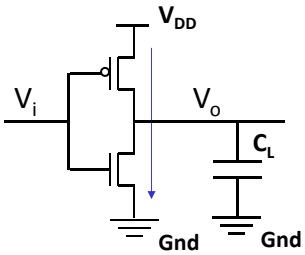
DCT Algorithm	Multiplies (8x8)	Additions (8x8)	Implemented by
Brute Force	4096	4096	-
Row-Col DCT	1024	1024	Bell core (16x16)
Chen's Algorithm	256	416	Telettra
Lee's Algorithm	192	464	SGS - Thompson
Feig's Algorithm (scaled DCT)	54	462	IBM (GP computer)

Algorithm complexity

By reducing the number of tasks (circuit activity) the energy dissipation is reduced.

Short Circuit Power Consumption

Short-Circuit Power Consumption



During the transition time from Gnd to V_{DD} and vice-versa and as input voltage V_i is within the interval

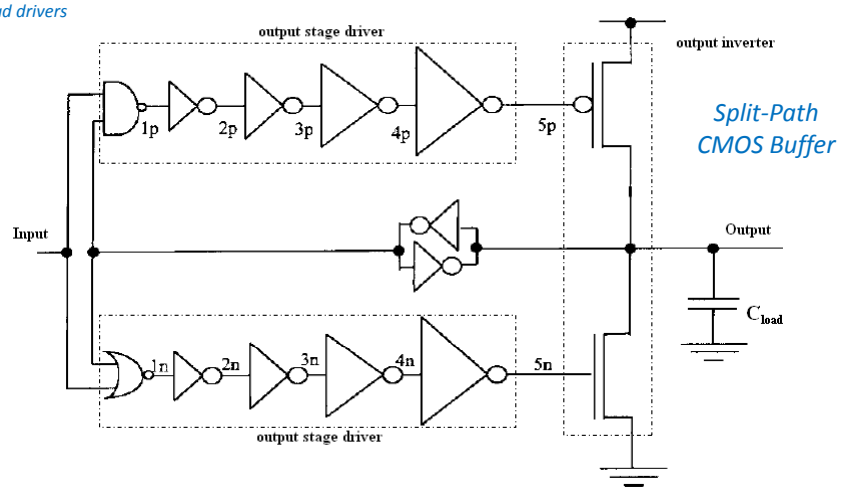
$V_{tn} < V_i < V_{DD} - |V_{tp}|$

a direct current from V_{DD} to the Gnd exists.

Short Circuit Power Consumption Reduction

Short-Circuit Consumption Reduction

High load drivers



Split-Path CMOS Buffer

General principle: *Signal rise and fall times must be equal!*

References

- "*Low Power Design Methodologies*," J. Rabaey and M. Pedram, Springer, 1997.
- "*Power Management of Digital Circuits in Deep Sub-Micron CMOS Technologies*," S. Henzler, Springer, 2006.
- "*Low-Power Digital VLSI Design*," A. Bellaouar and M. Elmasry, Kluwer Academic Publishers, 1996.
- "*A 1.5V Full-Swing Bootstrapped CMOS Large Capacitive-Load Driver Circuit Suitable for Low-Voltage CMOS VLSI*," J. Lou and J. Kuo, IEEE Journal of Solid-State Circuits, vol. 32, no. 1, pp. 119-121, 1997.
- "*A 0.9-V, 150-MHz, 10-mW, 4 mm , 2-D Discrete Cosine Transform Core Processor with Variable Threshold-Voltage (VT) Scheme*," T. Kuroda et.al., IEEE Journal of Solid-State Circuits, vol. 31, no. 11, pp. 1770-1779, 1996.
- "*Low-Power CMOS Digital Design*," A. Chandrakasan, S. Sheng and R. Brodersen, IEEE Journal of Solid-State Circuits, vol. 27, no. 4, pp. 473-484, 1992.
- "*Minimazing Power Consumption in Digital CMOS Circuits*," A. Chandrakasan, R. Brodersen, Proceedings of the IEEE, vol. 83, no. 4, pp. 498-523, 1995.