Lecture 7: The CMOS Inverter (Power)

Visvesh S. Sathe

Based on material prepared by prof. Visvesh S. Sathe

ELECTRICAL & COMPUTER ENGINEERING

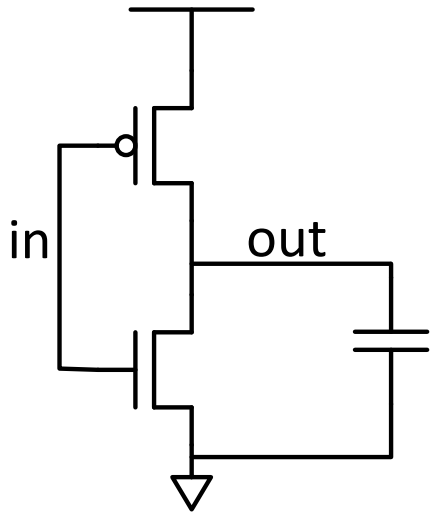
Acknowledgements

All class materials (lectures, assignments, etc.) based on material prepared by Prof. Visvesh S. Sathe, and reproduced with his permission

Visvesh S. Sathe
Associate Professor
Georgia Institute of Technology
https://psylab.ece.gatech.edu

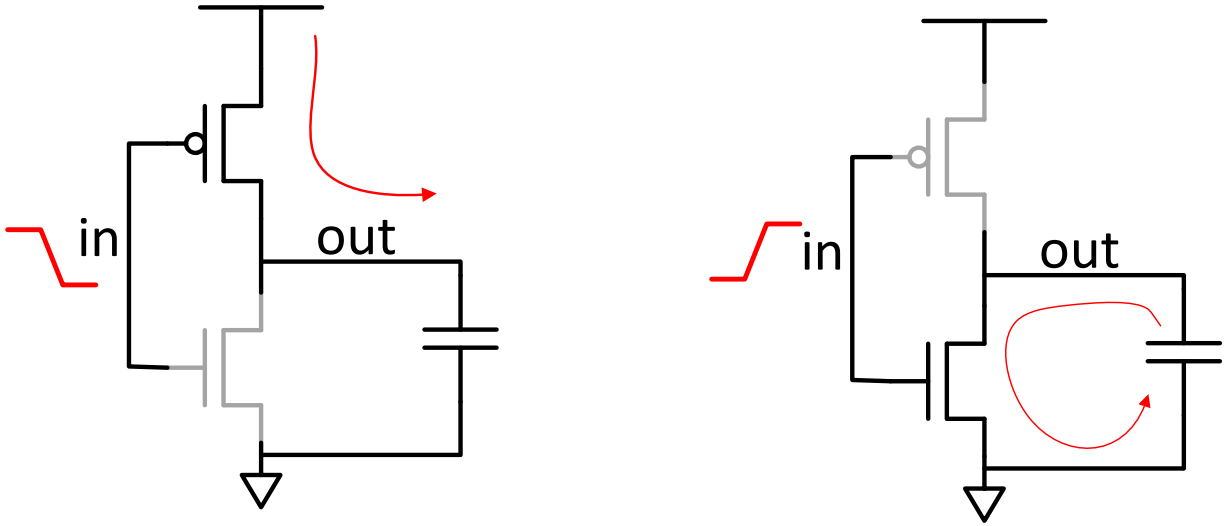
UW (2013-2022)
GaTech (2022-present)

Power/Energy Dissipation



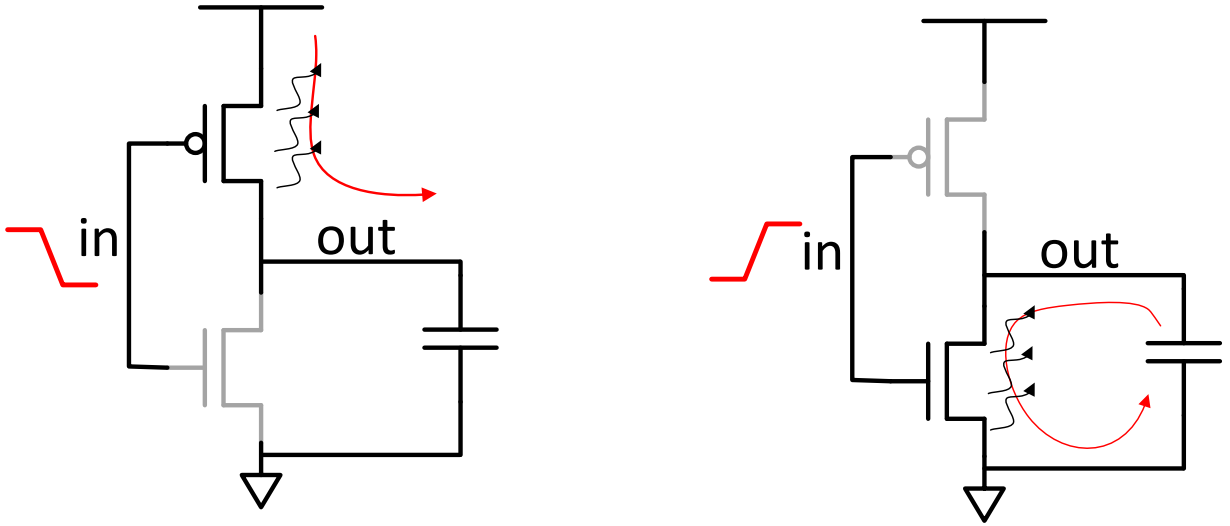
- Performing any computation dissipates some energy
- In CMOS, two major forms
  - Dynamic power dissipation (Occurs whenever signals switch)
  - Static power dissipation (Occurs regardless of switching activity)
- Going from gate-level to system-level power: $\sum p_i$

Dynamic Power Dissipation (Switching Loss)



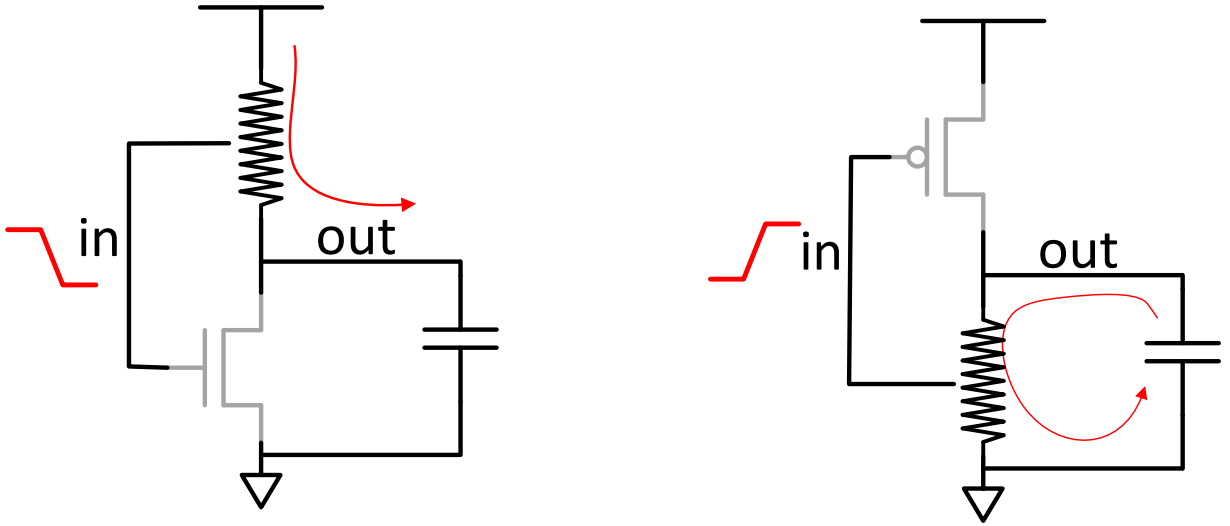
- Energy dissipated during charge/discharge
- Where does the dissipation occur?

Dynamic Power Dissipation (Switching Loss)



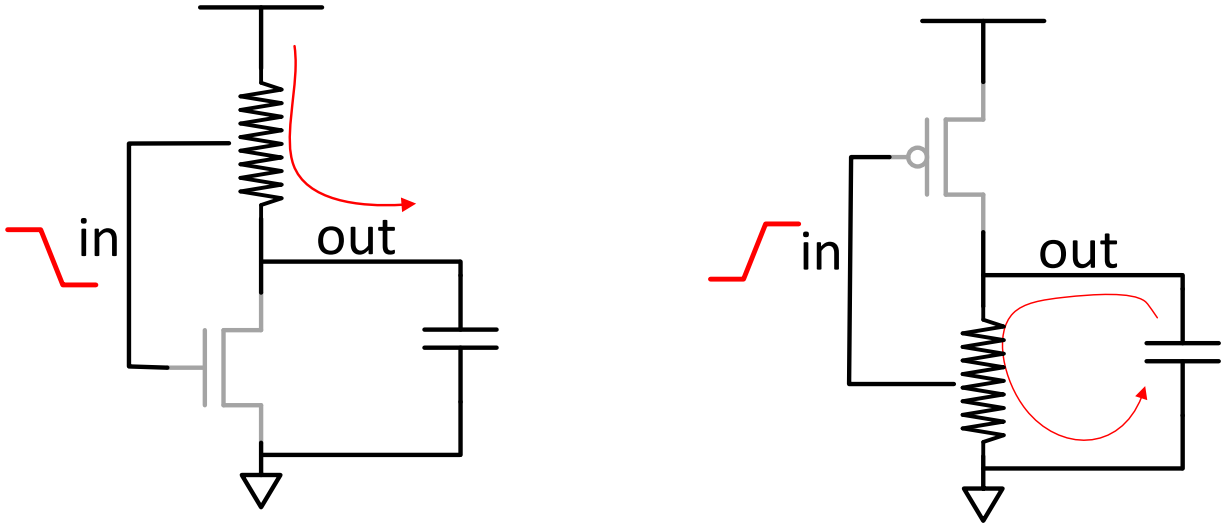
- Energy dissipated during charge/discharge
- Where does the dissipation occur? In the transistors,contacts,wire

Dynamic Power Dissipation (Switching Loss)



- Energy dissipated during charge/discharge
- Where does the dissipation occur? In the transistors,contacts,wire
- **IF** modeled as a resistor, per transition $E = \int_0^\infty I^2 R \, dt$

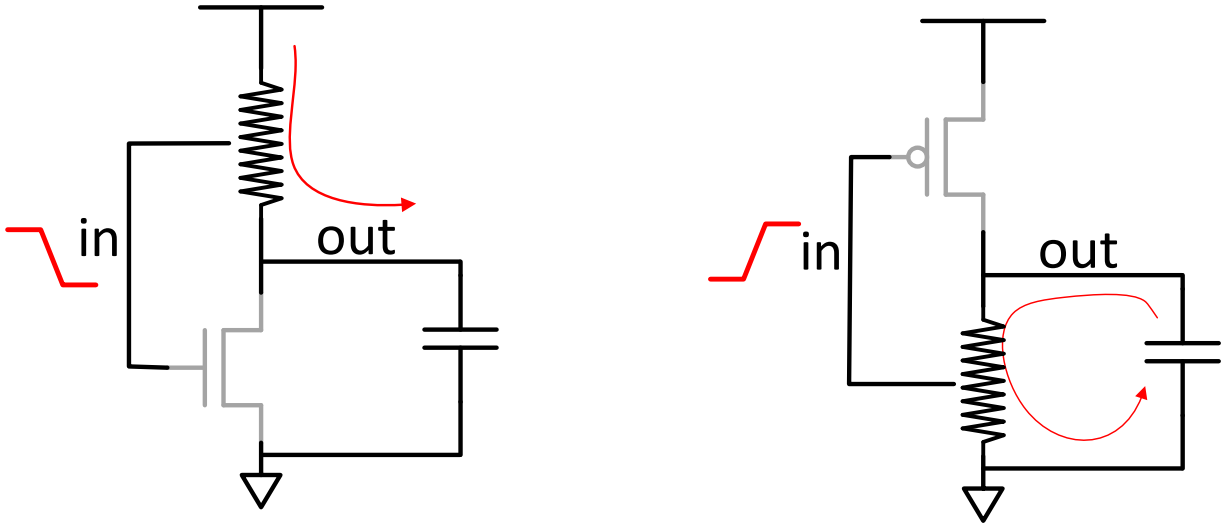
Dynamic Power Dissipation (Switching Loss)



- Energy dissipated during charge/discharge
- Where does the dissipation occur? In the transistors,contacts,wire
- **IF** modeled as a resistor, per transition $E = \int_0^\infty I^2 R \, dt$

  $= \int_0^\infty I^2 R \, dt$

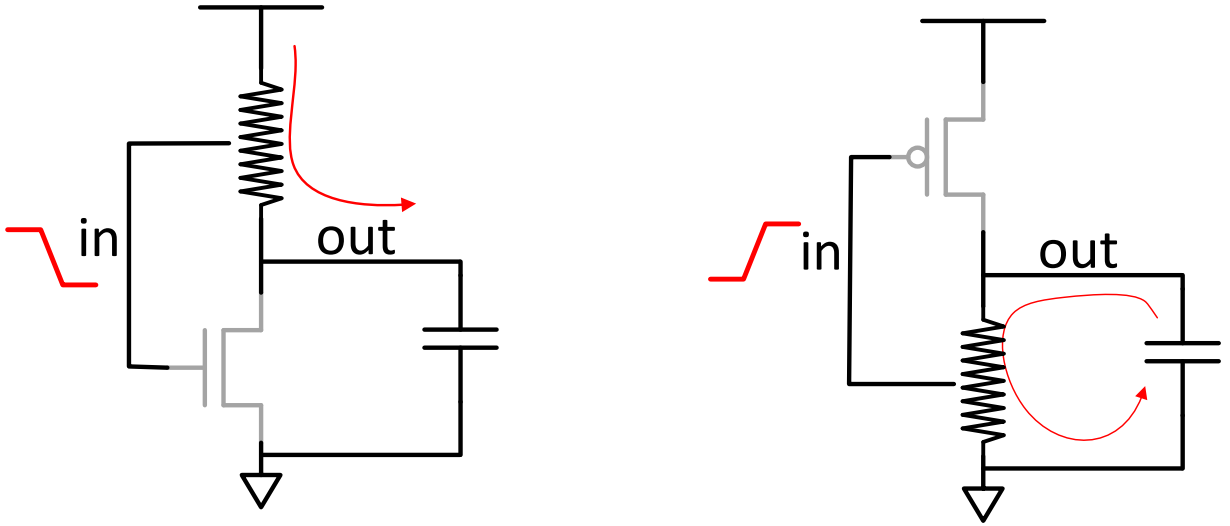
Dynamic Power Dissipation (Switching Loss)



- Energy dissipated during charge/discharge
- Where does the dissipation occur? In the transistors,contacts,wire
- **IF** modeled as a resistor, per transition $E = \int_0^\infty I^2 R\, dt$

  $= \int_0^\infty I^2 R\, dt = \int_0^\infty (\frac{V_{dd}}{R} e^{-\frac{t}{\tau}})^2 R\, dt$

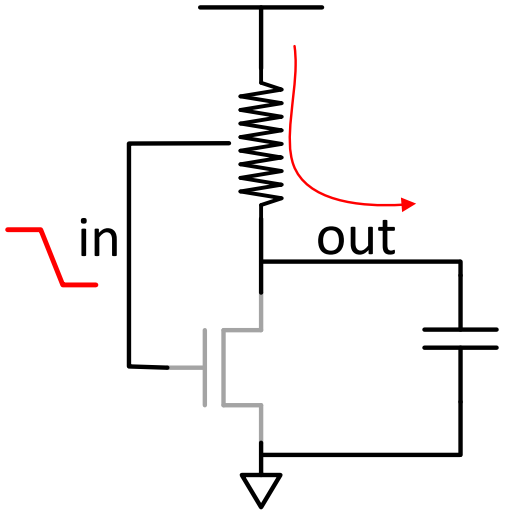
Dynamic Power Dissipation (Switching Loss)

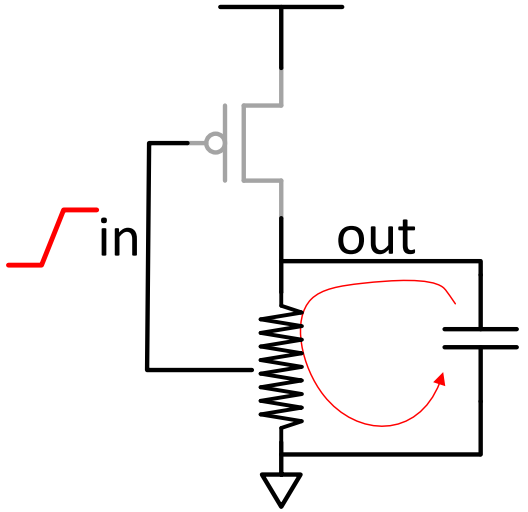


- Energy dissipated during charge/discharge
- Where does the dissipation occur? In the transistors,contacts,wire
- **IF** modeled as a resistor, per transition $E = \int_0^\infty I^2 R\, dt$

  $= \int_0^\infty I^2 R\, dt = \int_0^\infty (\frac{V_{dd}}{R} e^{-\frac{t}{\tau}})^2 R\, dt = \frac{V_{dd}^2}{R} [\frac{-\tau}{2} e^{\frac{-2t}{\tau}}]_0^\infty$

Dynamic Power Dissipation (Switching Loss)





- Energy dissipated during charge/discharge
- Where does the dissipation occur? In the transistors,contacts,wire
- **IF** modeled as a resistor, per transition $E = \int_0^\infty I^2 R\, dt$

$$= \int_0^\infty I^2 R\, dt = \int_0^\infty \left(\frac{V_{dd}}{R} e^{-\frac{t}{\tau}}\right)^2 R\, dt = \frac{V_{dd}^2}{R}\left[\frac{-\tau}{2} e^{\frac{-2t}{\tau}}\right]_0^\infty = \frac{1}{2} C V_{dd}^2$$

Dynamic Power Dissipation (Switching Loss)

- Resistor derivation gives the right answer…BUT
  - CMOS gates are not linear resistors
  - Current profile is not decaying exponential !!

Dynamic Power Dissipation (Switching Loss)

- Resistor derivation gives the right answer…BUT
  - CMOS gates are not linear resistors
  - Current profile is not decaying exponential !!
  - Simpler AND more accurate approach: Follow the charge
    - All* dissipation involved in charge/discharge of capacitance
    - Energy Delivered = Energy Dissipated + Change in Energy Stored
    - Track charge given by supply
    - Subtract away stored energy
    - Inverter example : $E_{diss} = V(CV) - 1/2CV^2 = 1/2CV^2$

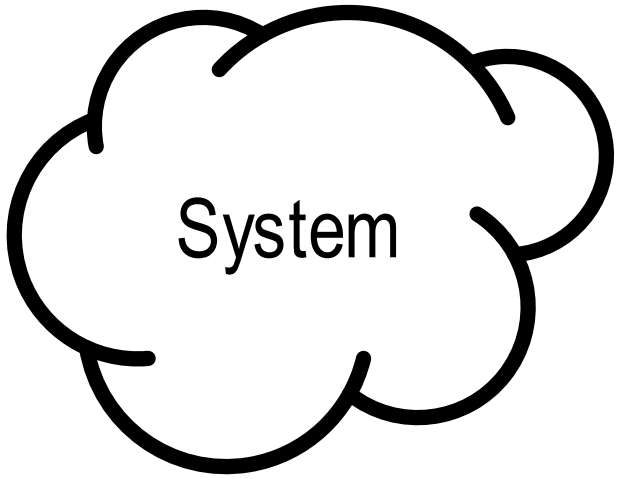
Dynamic Power Dissipation (Switching Loss)

- Resistor derivation gives the right answer…BUT
  - CMOS gates are not linear resistors
  - Current profile is not decaying exponential !!
  - Simpler AND more accurate approach: Follow the charge
    - All* dissipation involved in charge/discharge of capacitance
    - Energy Delivered = Energy Dissipated + Change in Energy Stored
    - Track charge given by supply
    - Subtract away stored energy
    - Inverter example : $E_{diss} = V(CV) - 1/2CV^2 = 1/2CV^2$
- Exercise: What happens during the fall?

Dynamic Power Dissipation (Switching Loss)

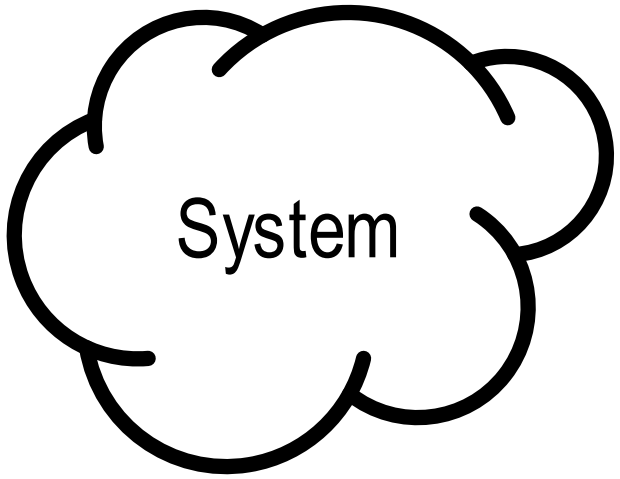
- Resistor derivation gives the right answer…BUT
  - CMOS gates are not linear resistors
  - Current profile is not decaying exponential !!
  - Simpler AND more accurate approach: Follow the charge
    - All* dissipation involved in charge/discharge of capacitance
    - Energy Delivered = Energy Dissipated + Change in Energy Stored
    - Track charge given by supply
    - Subtract away stored energy
    - Inverter example : $E_{diss} = V(CV) - 1/2CV^2 = 1/2CV^2$
- Exercise: What happens during the fall?
- Independent of:
  - Output rise time
  - Device threshold voltage

Middle-school Flashback



- Conservation of Energy is the only principle you need ..

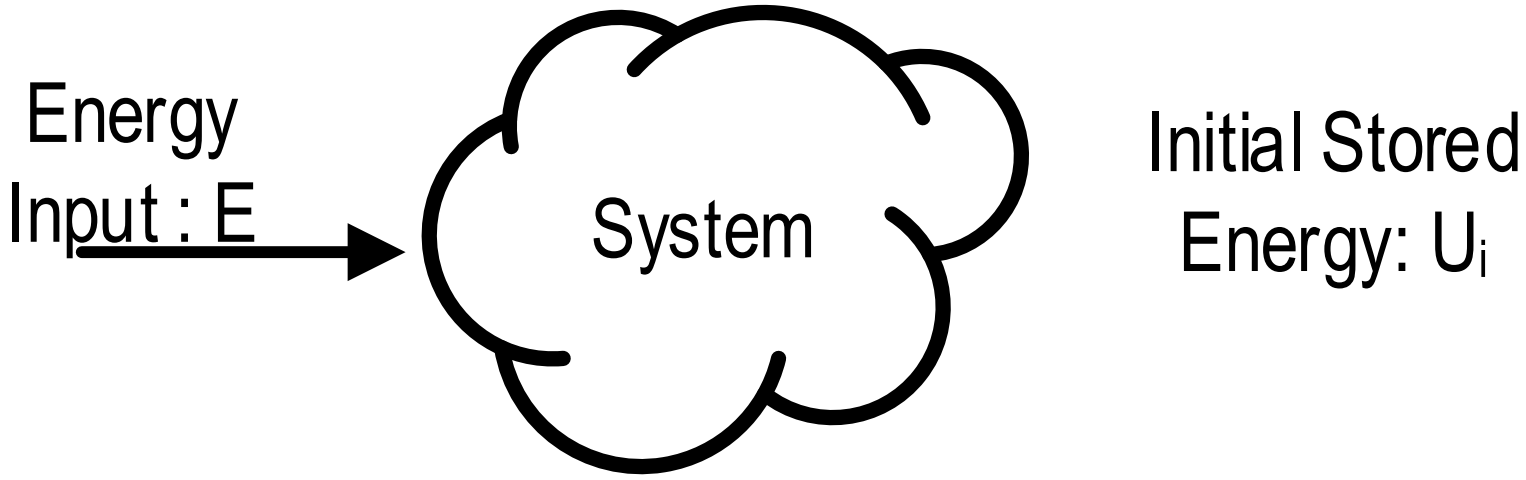
Middle-school Flashback



Initial Stored Energy: U_i

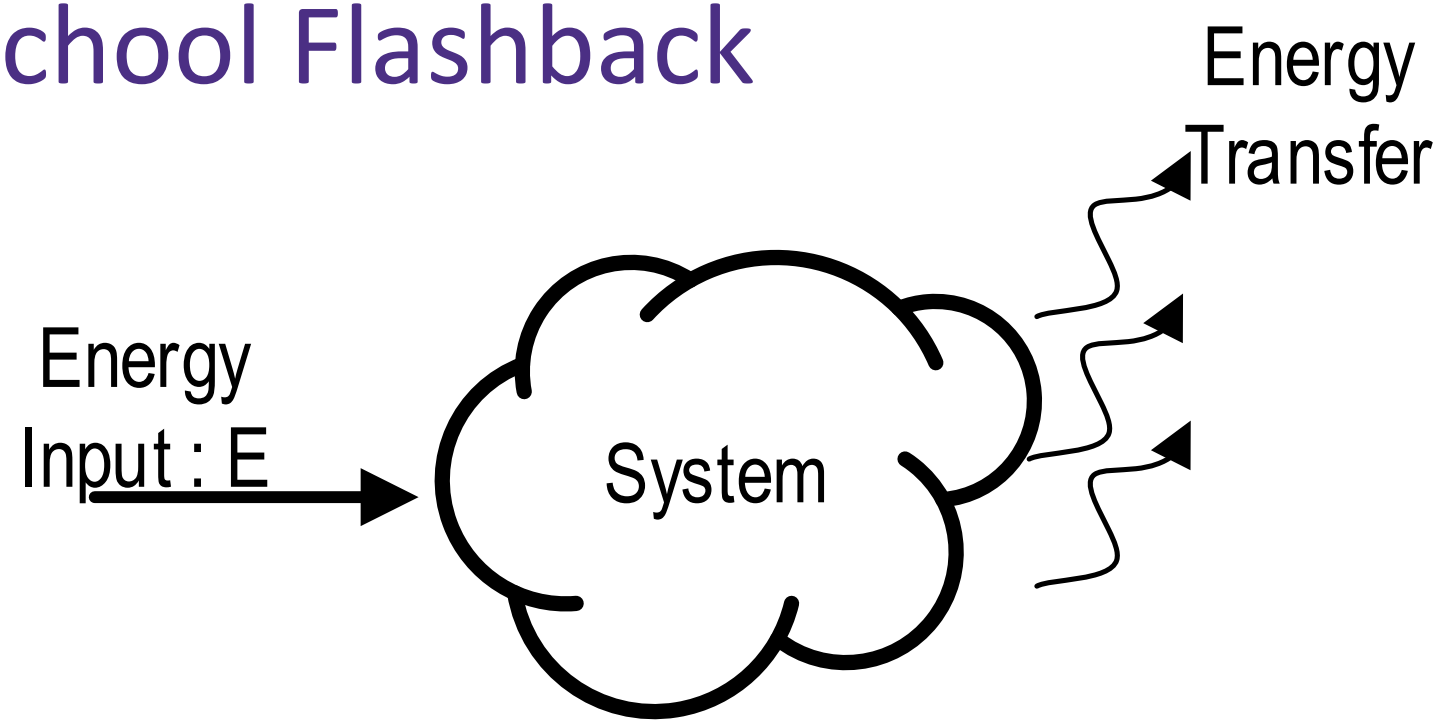
- Conservation of Energy is the only principle you need ..

Middle-school Flashback



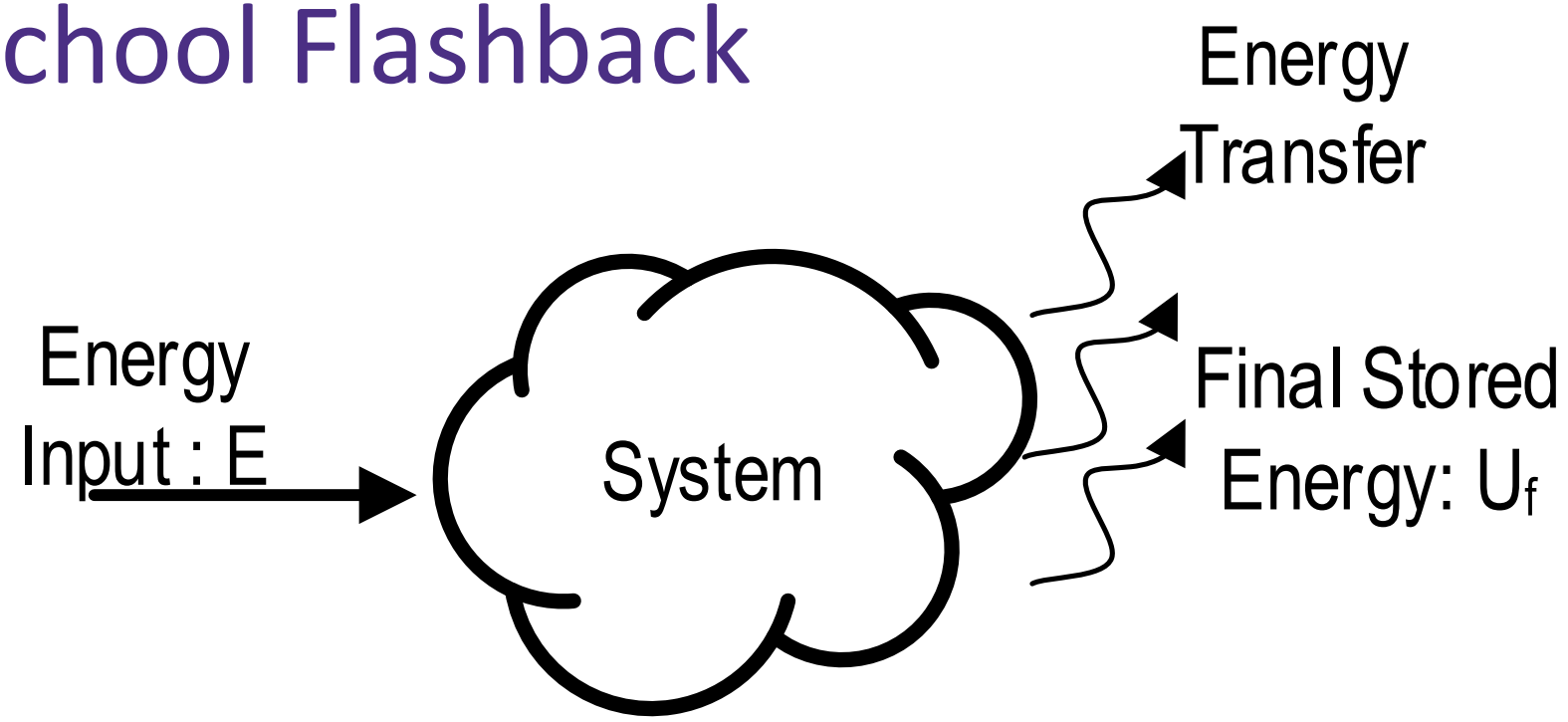
- Conservation of Energy is the only principle you need ..

Middle-school Flashback



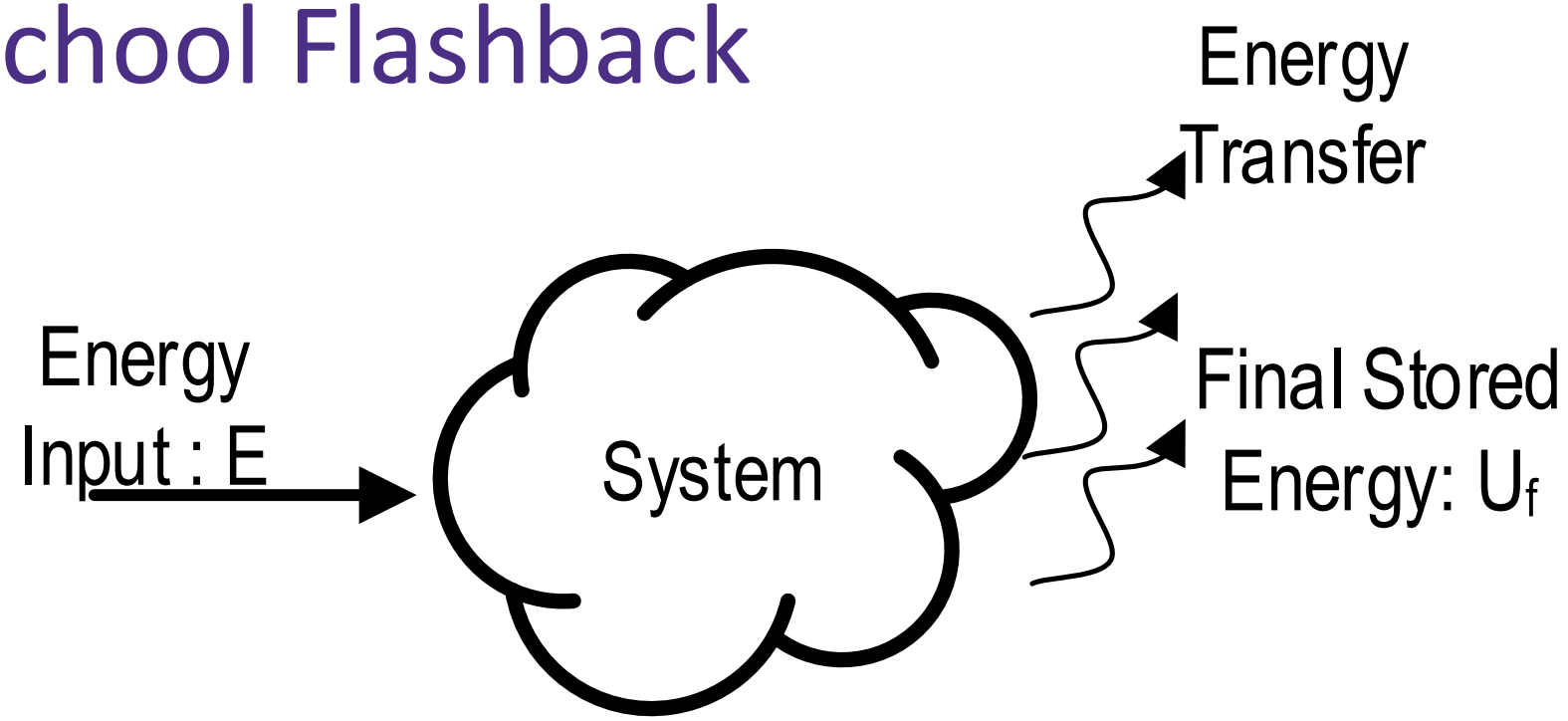
- Conservation of Energy is the only principle you need ..

Middle-school Flashback



- Conservation of Energy is the only principle you need ..

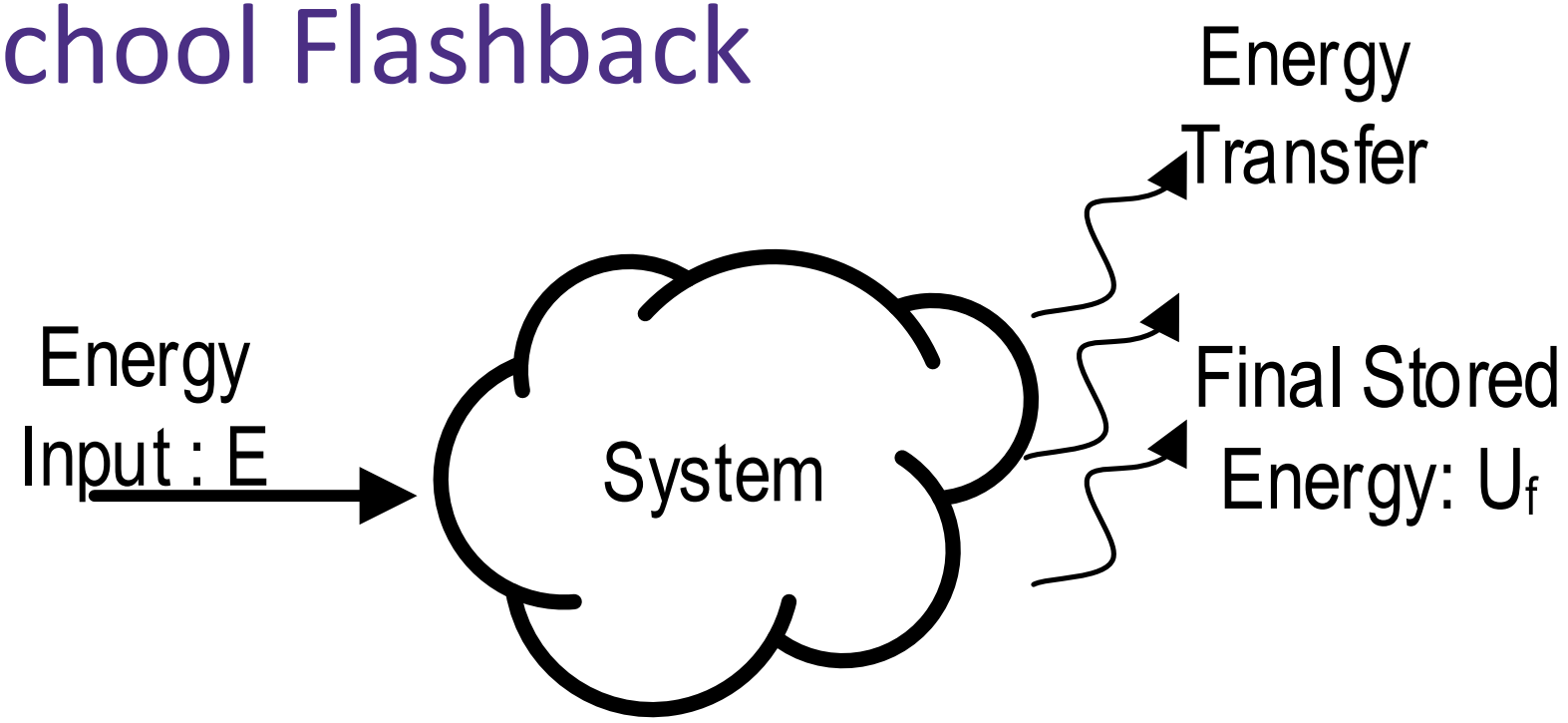
Middle-school Flashback



Conservation of Energy: $E - E_h = U_f - U_i$

- Conservation of Energy is the only principle you need ..

Middle-school Flashback

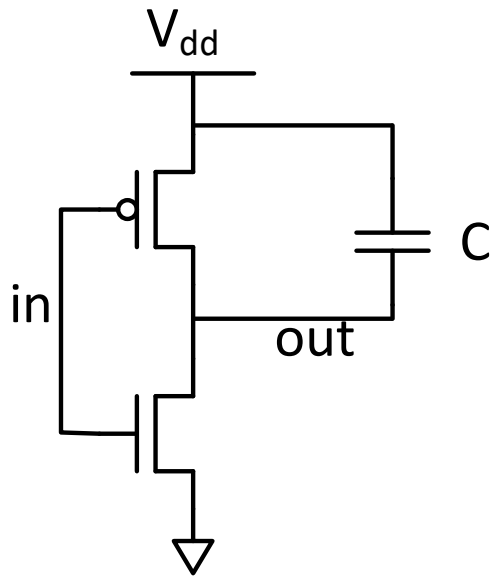


Conservation of Energy: $E - E_h = U_f - U_i$

$E - \Delta U = E_h$

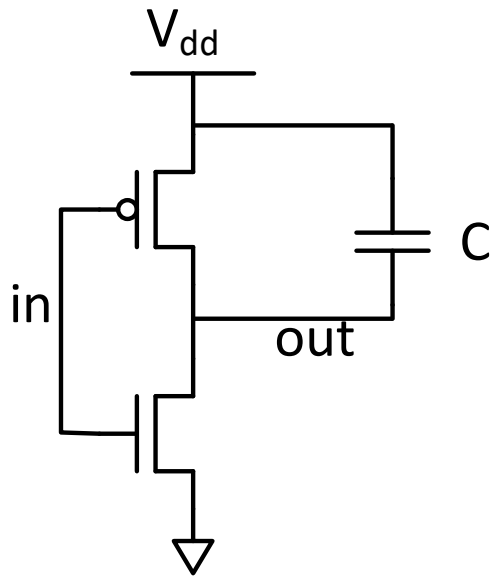
- Conservation of Energy is the only principle you need ..

Breakout Exercise



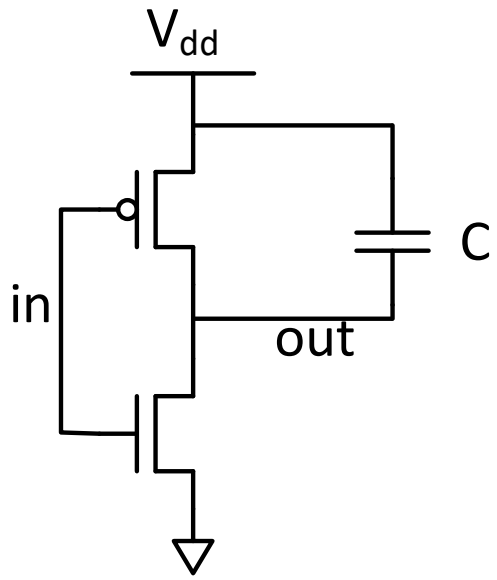
Parameter\Edge	Output Rise	Output Fall
E delivered by supply		
Power diss. In nmos		
Power diss. In pmos		
Total power diss.		

Breakout Exercise



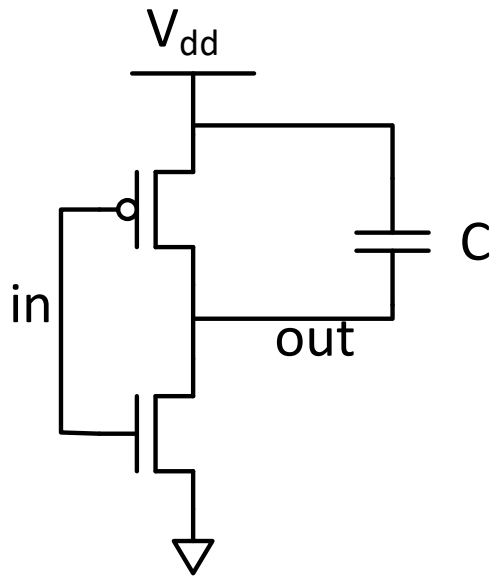
Parameter\Edge	Output Rise	Output Fall
E delivered by supply	0	
Power diss. In nmos		
Power diss. In pmos		
Total power diss.		

Breakout Exercise



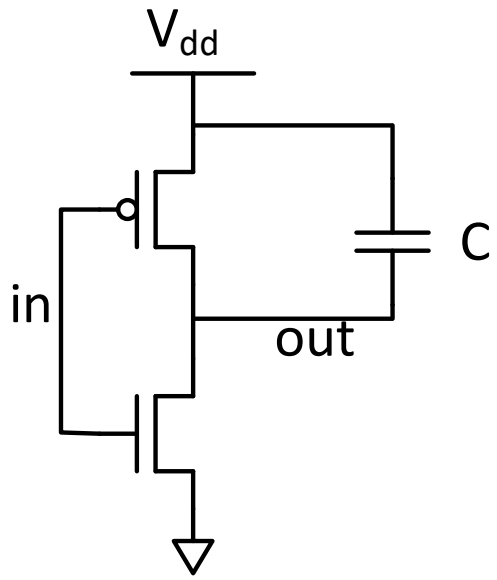
Parameter\Edge	Output Rise	Output Fall
E delivered by supply	0	CV^2_{dd}
Power diss. In nmos		
Power diss. In pmos		
Total power diss.		

Breakout Exercise



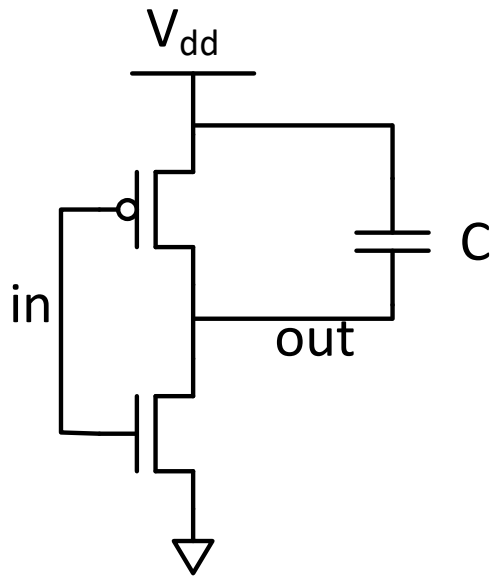
Parameter\Edge	Output Rise	Output Fall
E delivered by supply	0	CV^2_{dd}
Power diss. In nmos	0	
Power diss. In pmos		
Total power diss.		

Breakout Exercise



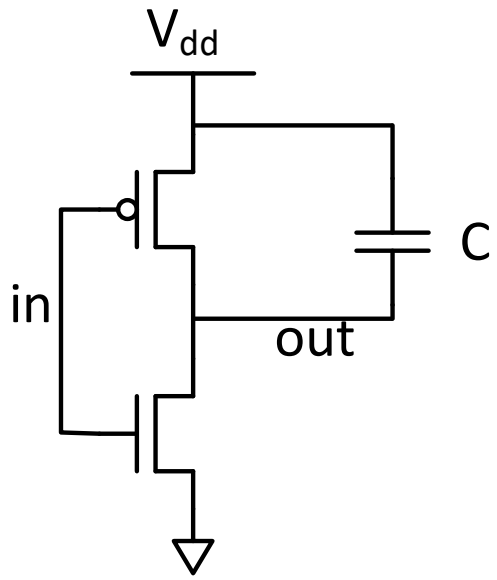
Parameter\Edge	Output Rise	Output Fall
E delivered by supply	0	CV^2_{dd}
Power diss. In nmos	0	$1/2CV_{dd}^2$
Power diss. In pmos		
Total power diss.		

Breakout Exercise



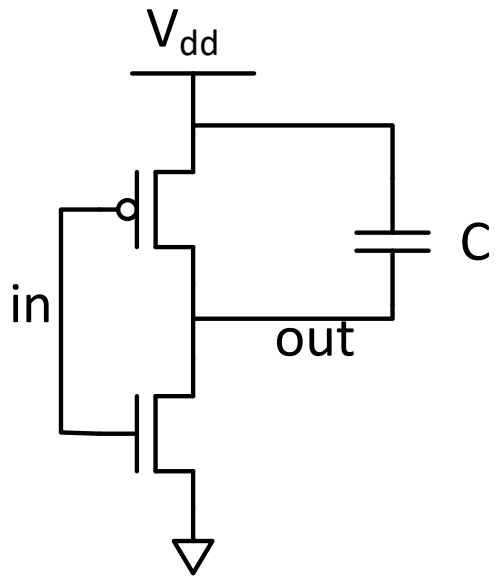
Parameter\Edge	Output Rise	Output Fall
E delivered by supply	0	CV^2_{dd}
Power diss. In nmos	0	$1/2CV_{dd}^2$
Power diss. In pmos	$1/2CV_{dd}^2$	
Total power diss.		

Breakout Exercise



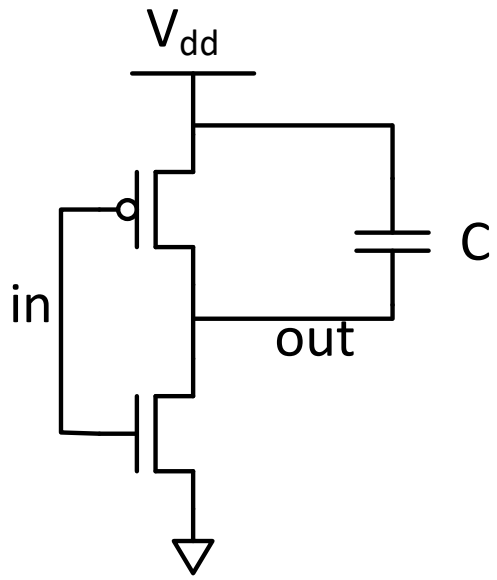
Parameter\Edge	Output Rise	Output Fall
E delivered by supply	0	CV^2_{dd}
Power diss. In nmos	0	$1/2CV_{dd}^2$
Power diss. In pmos	$1/2CV_{dd}^2$	0
Total power diss.		

Breakout Exercise



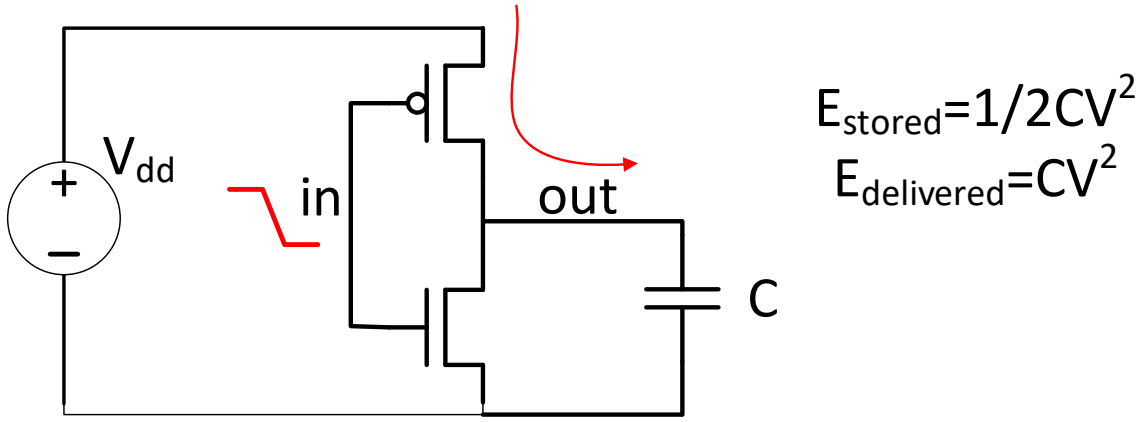
Parameter\Edge	Output Rise	Output Fall
E delivered by supply	0	CV^2_{dd}
Power diss. In nmos	0	$1/2CV_{dd}^2$
Power diss. In pmos	$1/2CV_{dd}^2$	0
Total power diss.	$1/2CV_{dd}^2$	

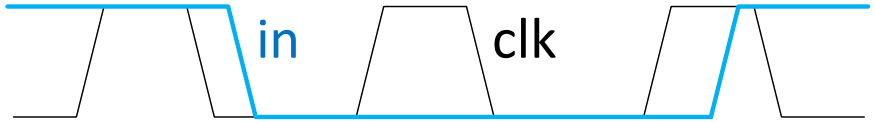
Breakout Exercise



Parameter\Edge	Output Rise	Output Fall
E delivered by supply	0	CV^2_{dd}
Power diss. In nmos	0	$1/2CV_{dd}^2$
Power diss. In pmos	$1/2CV_{dd}^2$	0
Total power diss.	$1/2CV_{dd}^2$	$1/2CV_{dd}^2$

Dynamic Power Dissipation (Switching Loss)





- $E_{diss} = 1/2CV_{dd}^2$
- Digital systems are clocked. Gates toggle 0 or more times every cycle
- Power = $\frac{\partial E}{\partial t} \approx \frac{E}{T} = E \cdot f = \frac{1}{2} sCV_{dd}^2 f$
  - s : switching-activity. Average number of net toggles in a compute cycle
  - f : switching frequency of the digital system
- Total system power
  - Track power of each net
  - Perform summation across all nets : $\sum \frac{1}{2} s_i C_i V_{dd}^2 f$

Switching Activity

- Often wrongly quoted as probability of switching event
- Expected number of “toggles” within one clock cycle
  - 0 to 1
  - 1 to 0
- Clock nets tend to have the highest switching activity at
- Is it possible to have higher switching activities?

- Spurious switching activity of ten referred to as glitching

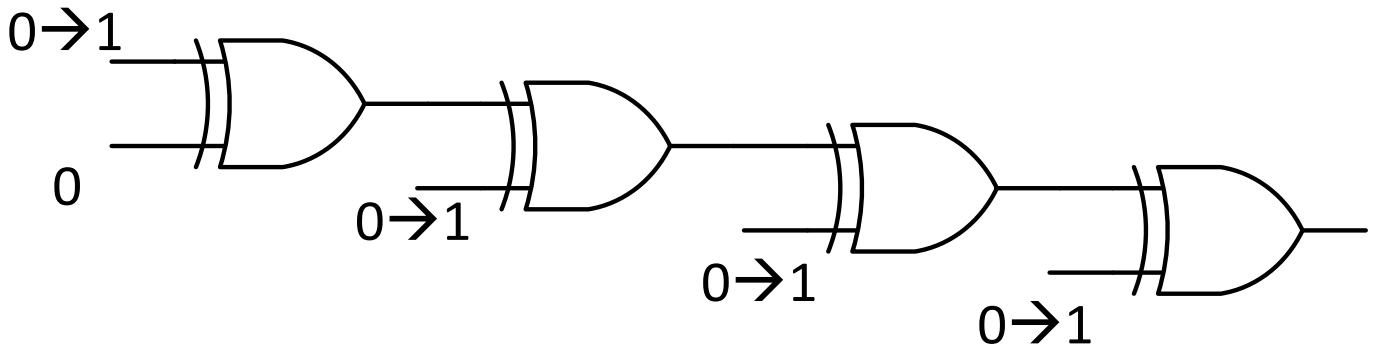
Switching Activity

- Often wrongly quoted as probability of switching event
- Expected number of “toggles” within one clock cycle
  - 0 to 1
  - 1 to 0
- Clock nets tend to have the highest switching activity at s=2
- Is it possible to have higher switching activities?

- Spurious switching activity of ten referred to as glitching

Switching Activity

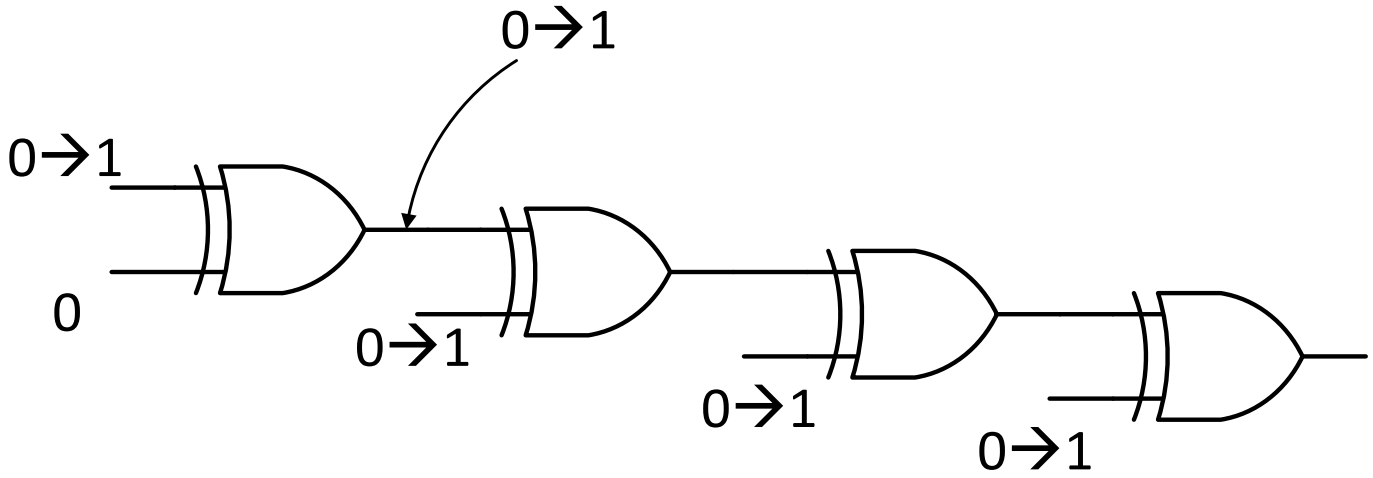
- Often wrongly quoted as probability of switching event
- Expected number of “toggles” within one clock cycle
  - 0 to 1
  - 1 to 0
- Clock nets tend to have the highest switching activity at s=2
- Is it possible to have higher switching activities?



- Spurious switching activity of ten referred to as glitching

Switching Activity

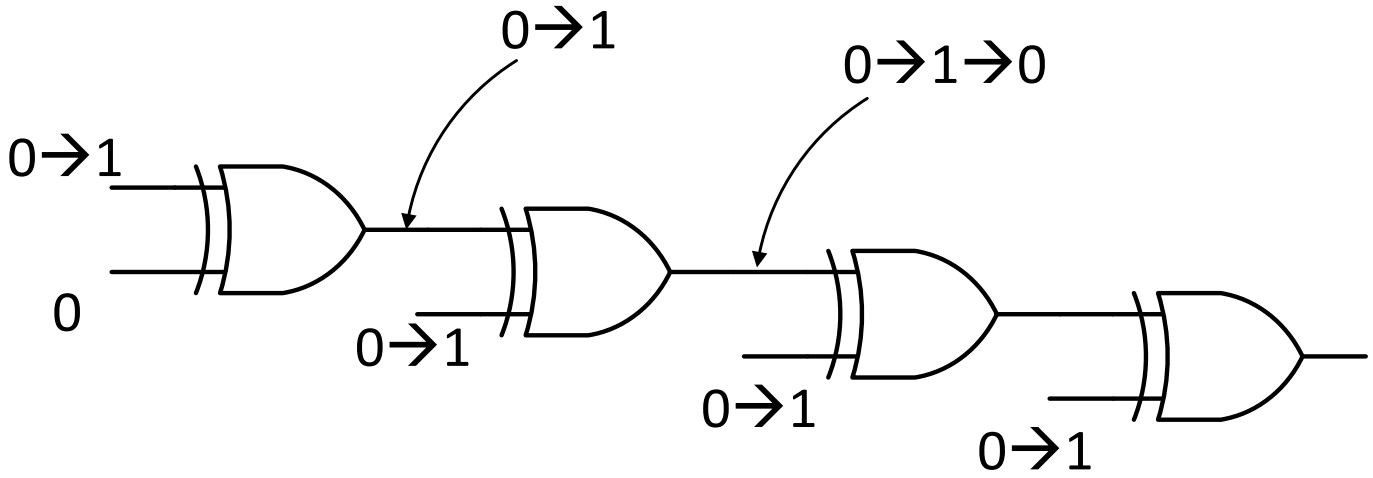
- Often wrongly quoted as probability of switching event
- Expected number of “toggles” within one clock cycle
  - 0 to 1
  - 1 to 0
- Clock nets tend to have the highest switching activity at s=2
- Is it possible to have higher switching activities?



- Spurious switching activity of ten referred to as glitching

Switching Activity

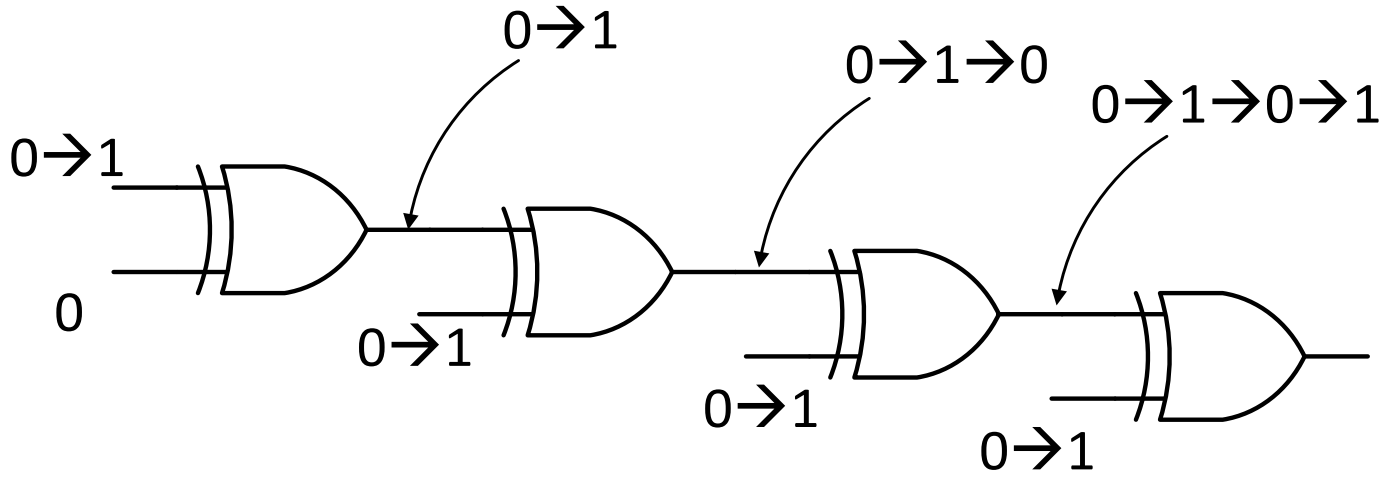
- Often wrongly quoted as probability of switching event
- Expected number of “toggles” within one clock cycle
  - 0 to 1
  - 1 to 0
- Clock nets tend to have the highest switching activity at s=2
- Is it possible to have higher switching activities?



- Spurious switching activity of ten referred to as glitching

Switching Activity

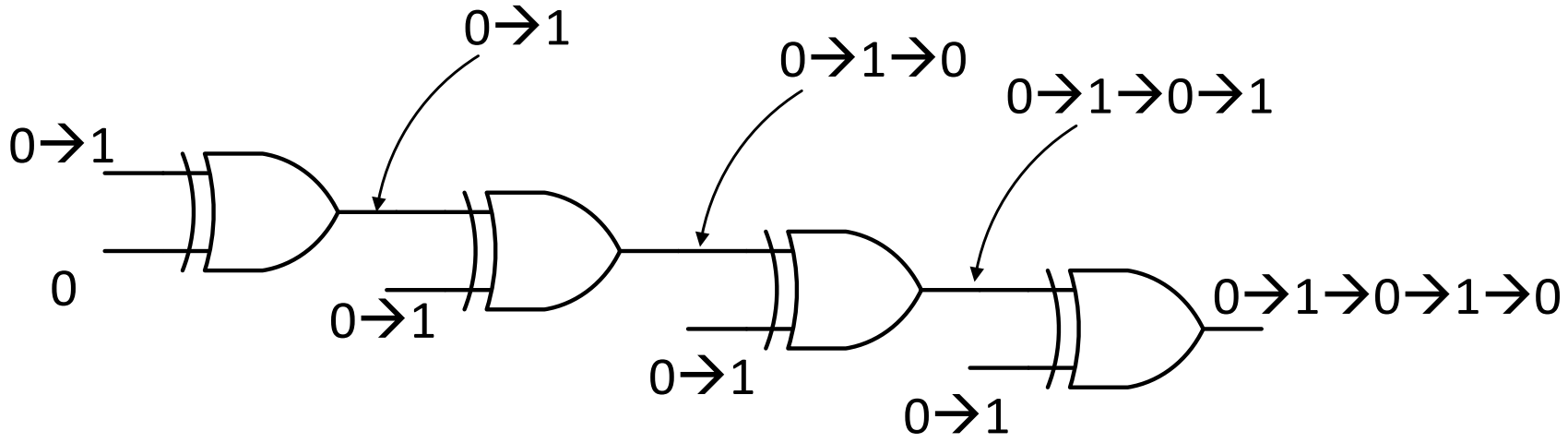
- Often wrongly quoted as probability of switching event
- Expected number of “toggles” within one clock cycle
  - 0 to 1
  - 1 to 0
- Clock nets tend to have the highest switching activity at s=2
- Is it possible to have higher switching activities?



- Spurious switching activity of ten referred to as glitching

Switching Activity

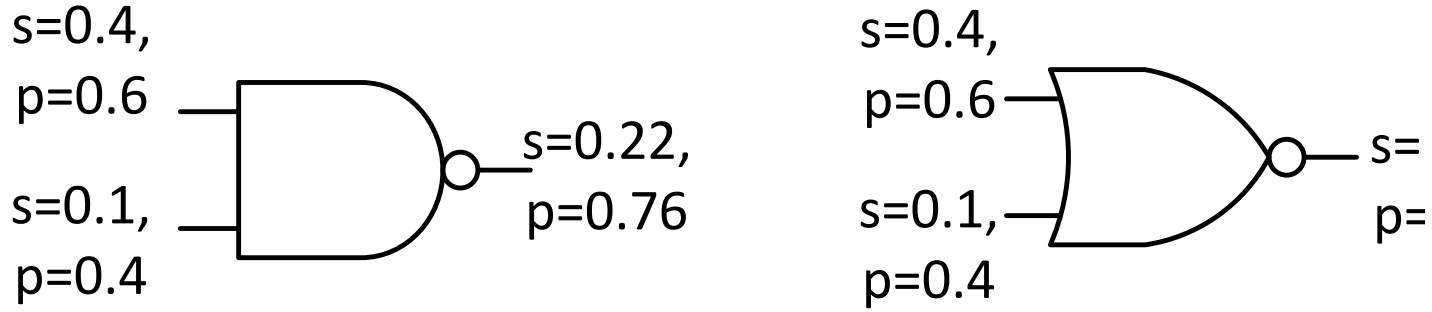
- Often wrongly quoted as probability of switching event
- Expected number of “toggles” within one clock cycle
  - 0 to 1
  - 1 to 0
- Clock nets tend to have the highest switching activity at s=2
- Is it possible to have higher switching activities?



- Spurious switching activity of ten referred to as glitching

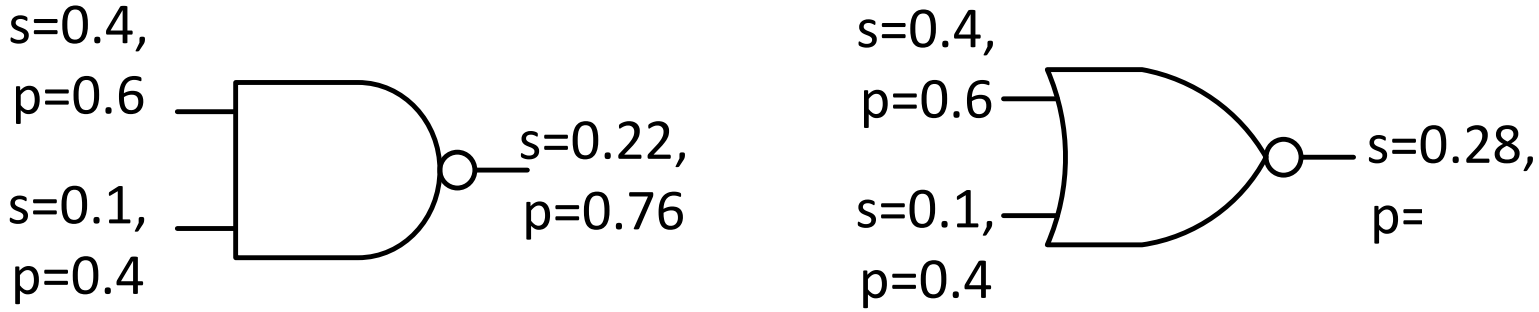
Switching activity (contd.)

- Glitches
  - Data arrives at inputs over a distribution of **time**
  - Gates that have "*balanced*" logical outputs tend to cause more glitches
- Switching activity estimate from output probability:
  - If P(out==1) = *p*
  - S = 2pq
  - Not a very good estimate
- Better estimate (still not precise)
  - Propagation of switching activity of a ***single event*** with ***independent inputs***
  - Track P(out==1) and S for each input
  - $S_{out} = \sum_i P(x_i \; transition \; causes \; switch) * \; S_i$



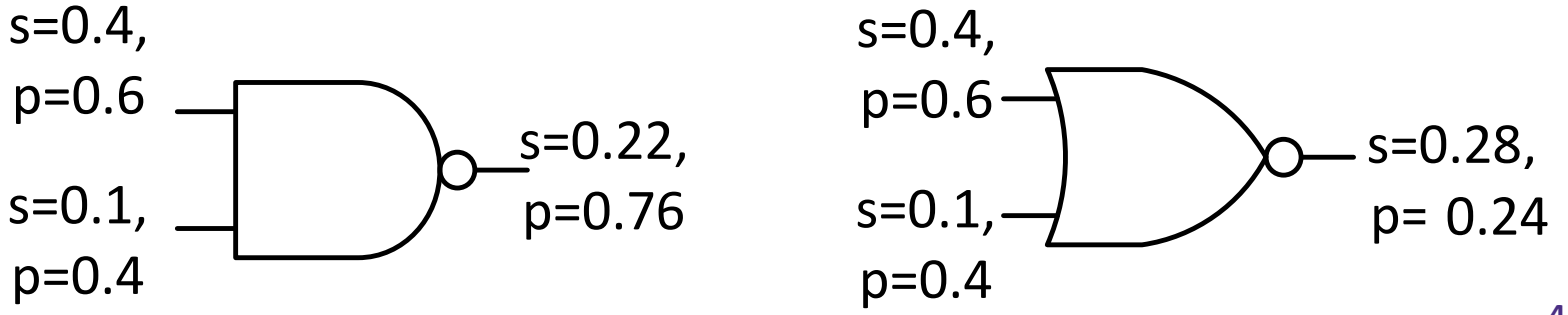
Switching activity (contd.)

- Glitches
  - Data arrives at inputs over a distribution of **time**
  - Gates that have "*balanced*" logical outputs tend to cause more glitches
- Switching activity estimate from output probability:
  - If P(out==1) = *p*
  - S = 2pq
  - Not a very good estimate
- Better estimate (still not precise)
  - Propagation of switching activity of a ***single event*** with ***independent inputs***
  - Track P(out==1) and S for each input
  - $S_{out} = \sum_i P(x_i \; transition \; causes \; switch) * \; S_i$

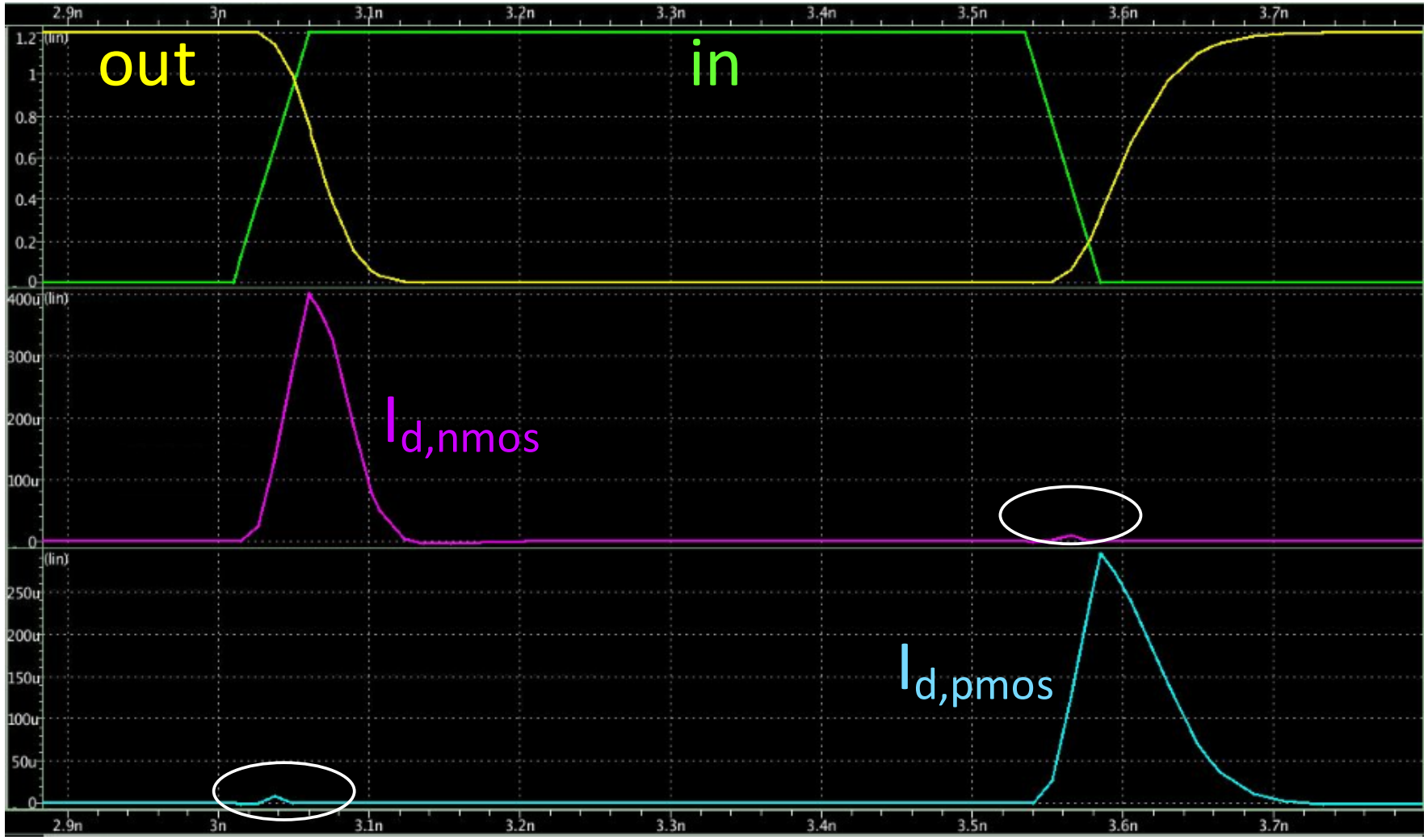


Switching activity (contd.)

- Glitches
  - Data arrives at inputs over a distribution of **time**
  - Gates that have "*balanced*" logical outputs tend to cause more glitches
- Switching activity estimate from output probability:
  - If P(out==1) = *p*
  - S = 2pq
  - Not a very good estimate
- Better estimate (still not precise)
  - Propagation of switching activity of a ***single event*** with ***independent inputs***
  - Track P(out==1) and S for each input
  - $S_{out} = \sum_i P(x_i \; transition \; causes \; switch) * \; S_i$

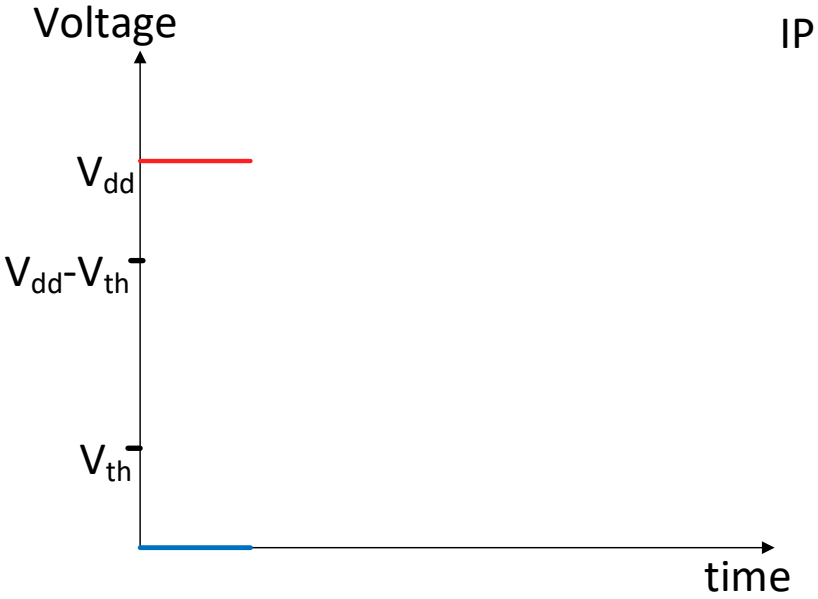


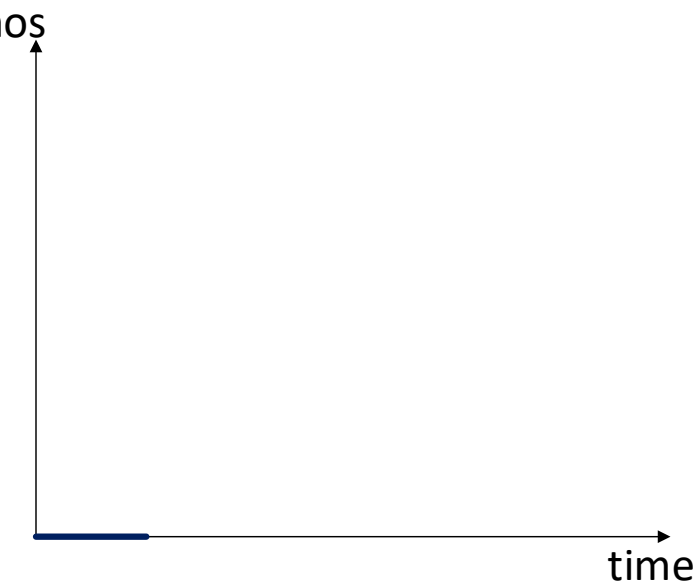
Dynamic Power Dissipation (Crossover Current)



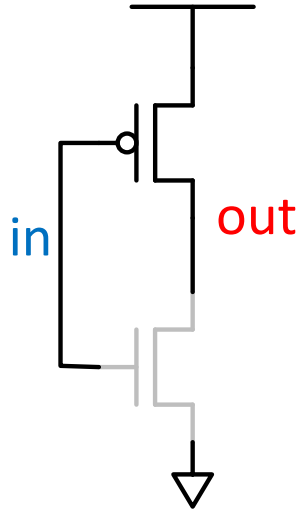
- A.k.a Crowbar current, Shoot-through current, Short-circuit current

Crowbar (aka shoot-through, crossover current)

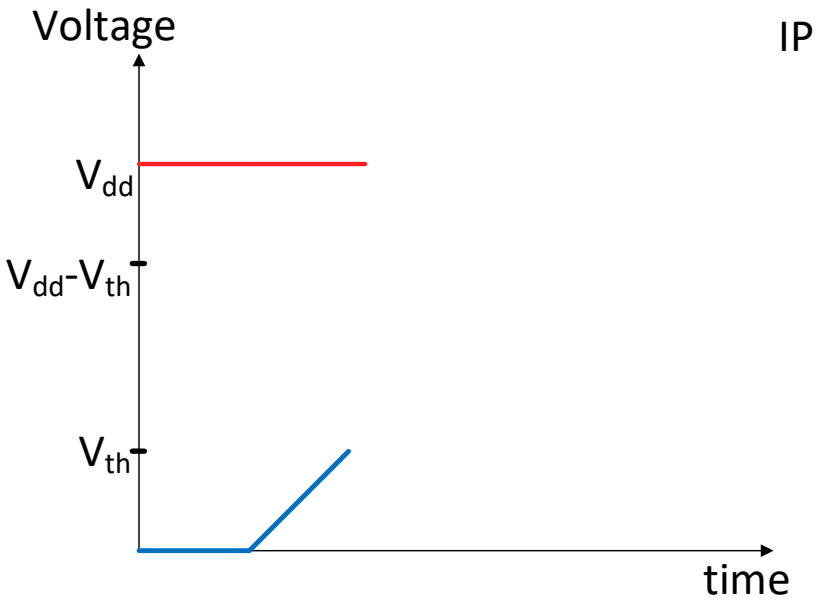


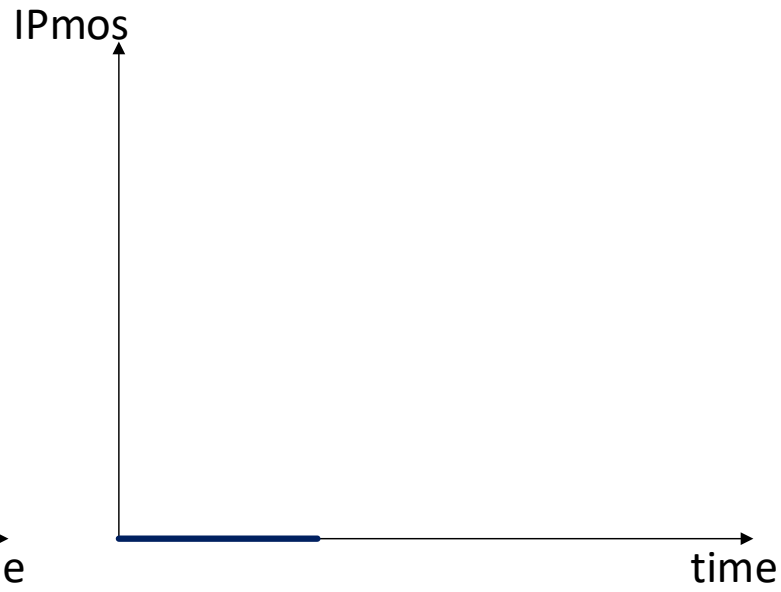


- Examine output discharge event
- Initially, in=0, out=vdd

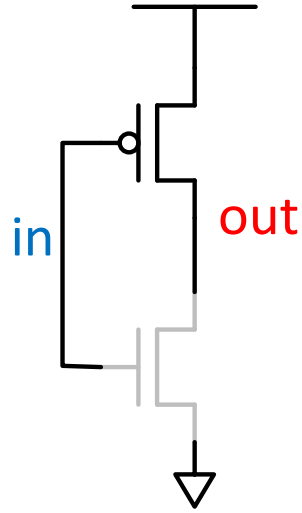


Crowbar (aka shoot-through, crossover current)

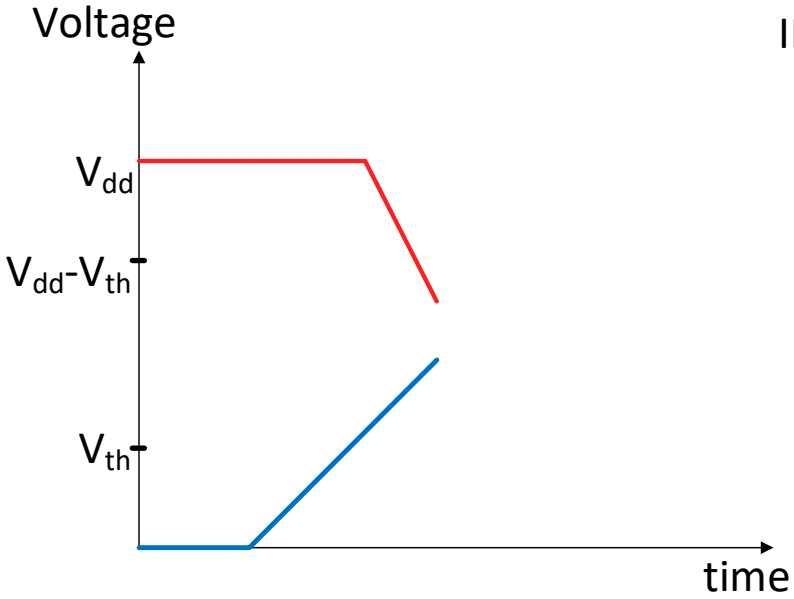


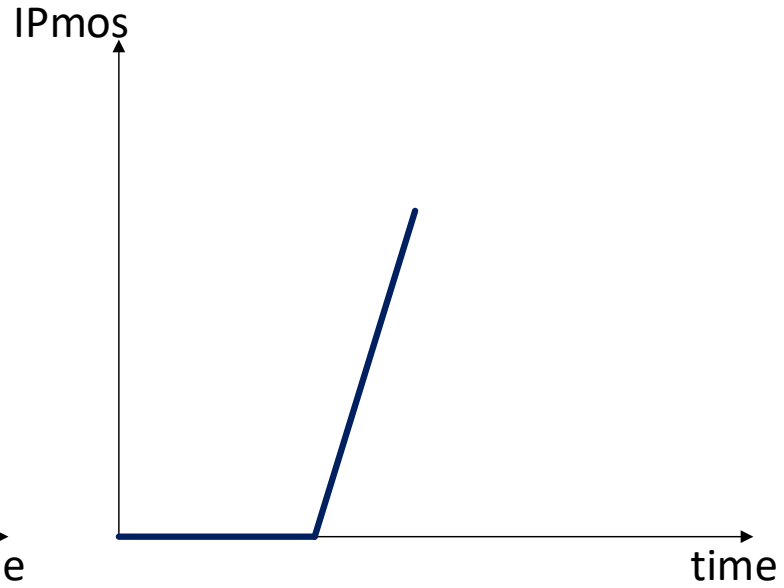


- Vin=Vth → Nmos off → out=vdd

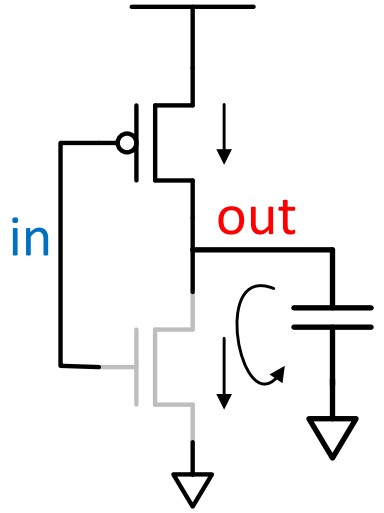


Crowbar (aka shoot-through, crossover current)

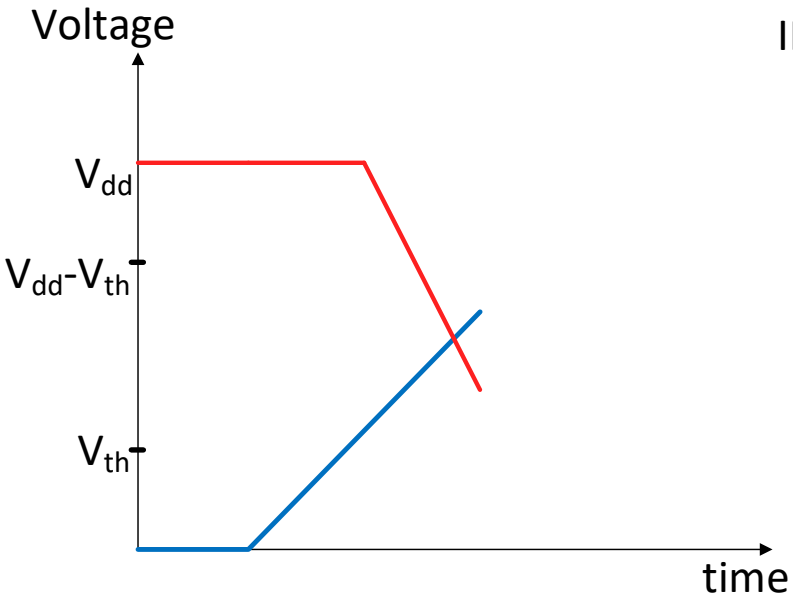


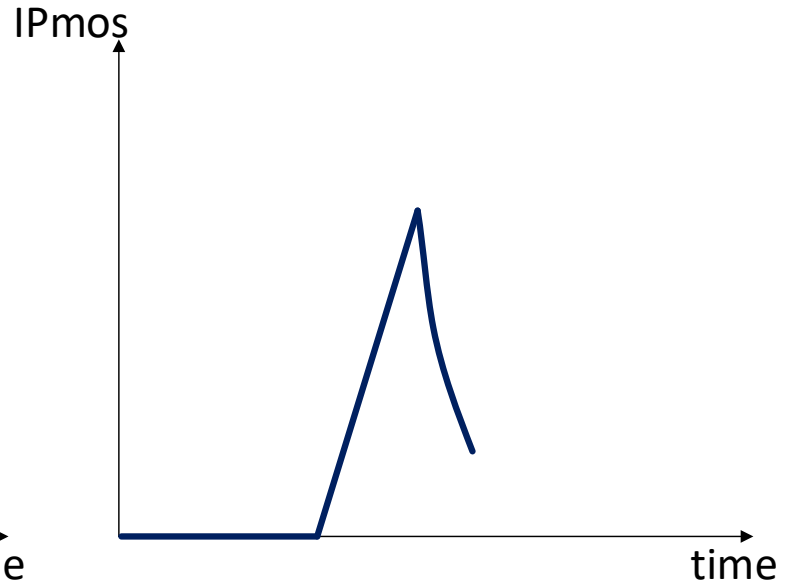


- Vin>Vth, Vout<Vdd
  - Nmos, Pmos both on
  - Nmos discharges load cap
  - Pmos overdrive ↓, Vds ↑: Overall I ↑

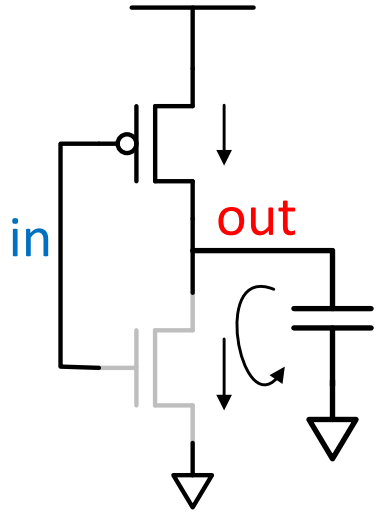


Crowbar (aka shoot-through, crossover current)

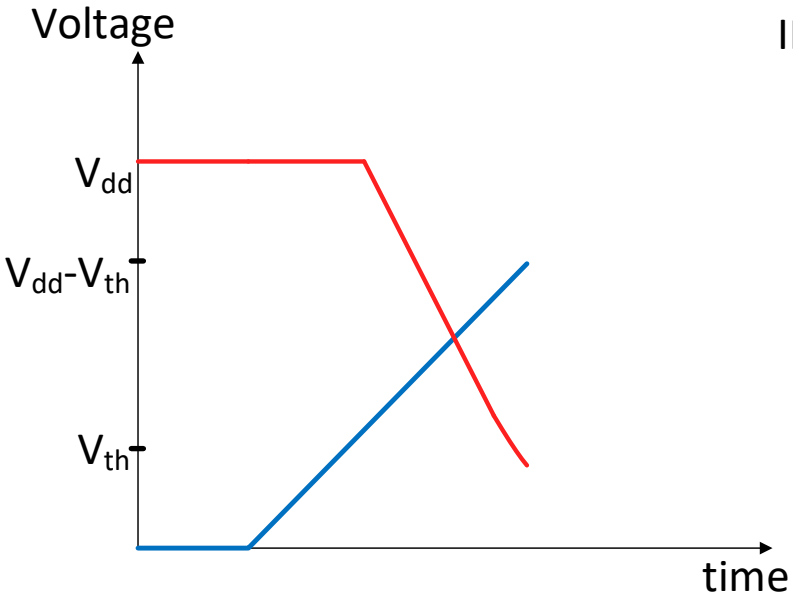


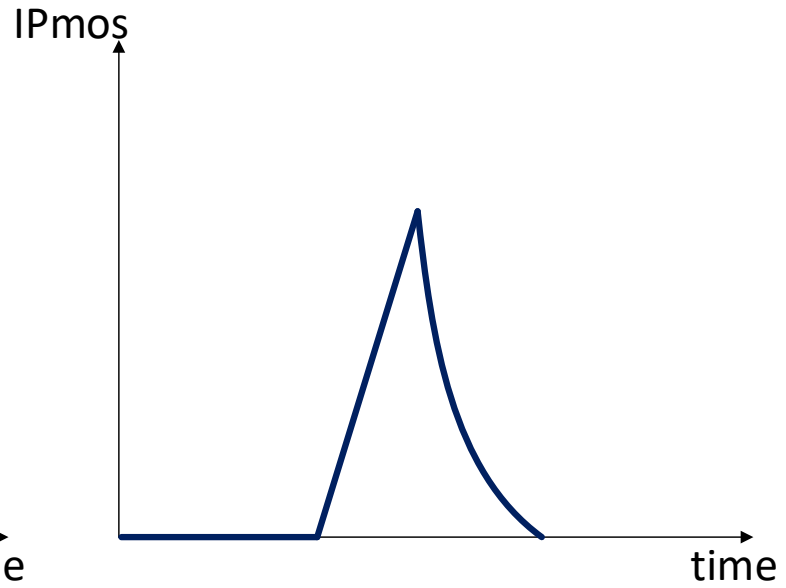


- Vin>Vth, Vout<Vdd
  - Nmos, Pmos both on
  - Nmos continues load cap discharge
  - Pmos overdrive ↓, Vds ↑: Overall I ↓

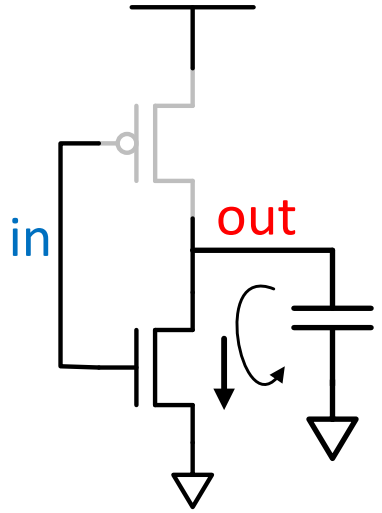


Crowbar (aka shoot-through, crossover current)

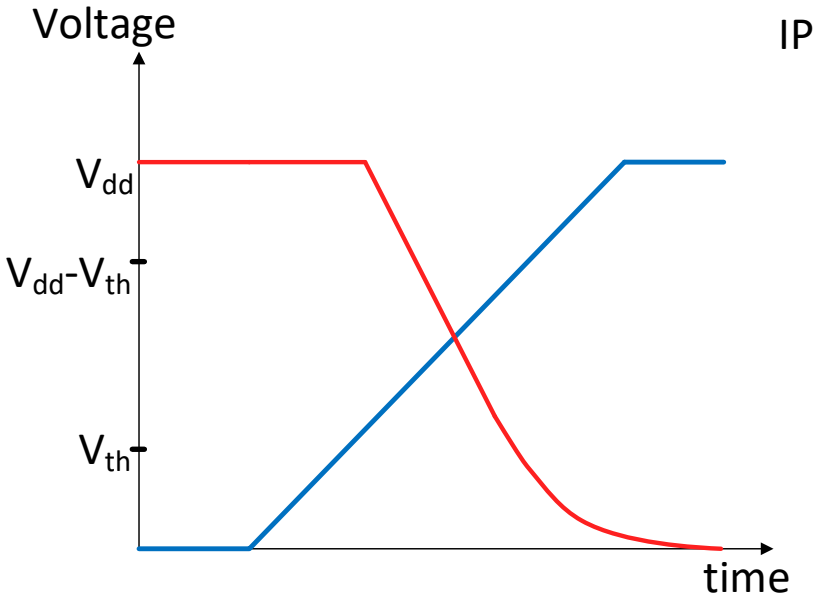


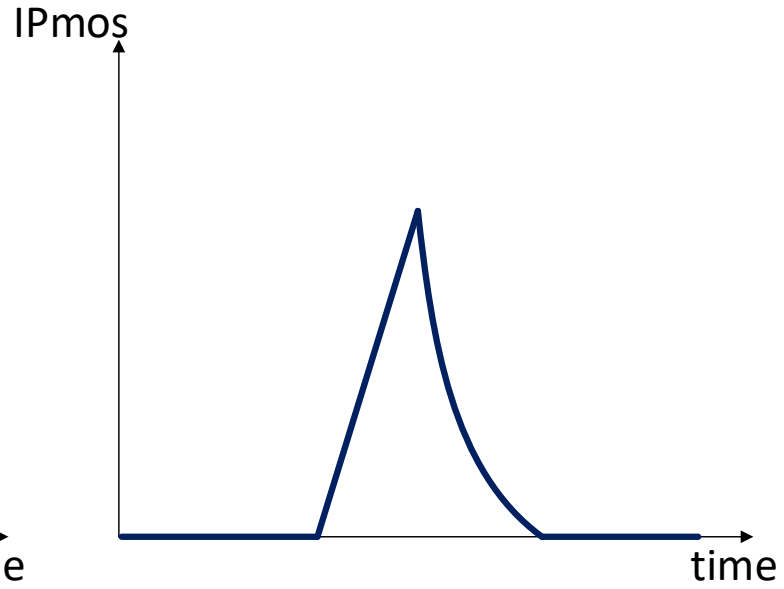


- Vin=Vdd-Vth
  - Nmos continues cap discharge
  - Pmos overdrive is 0 → off

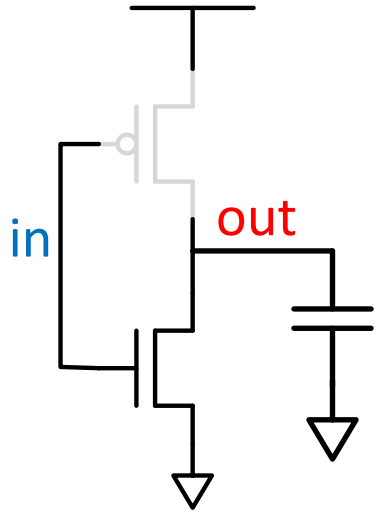


Crowbar (aka shoot-through, crossover current)

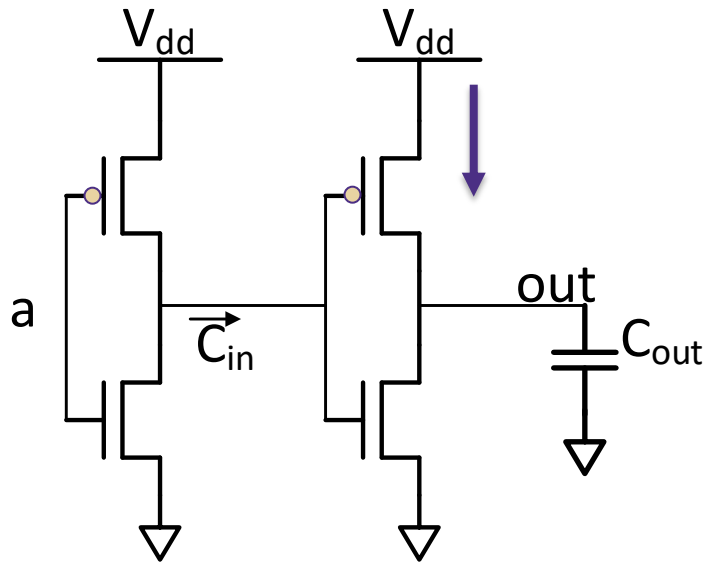




- Vin>Vth, Vout<Vdd
  - Nmos completes discharge
  - Pmos off

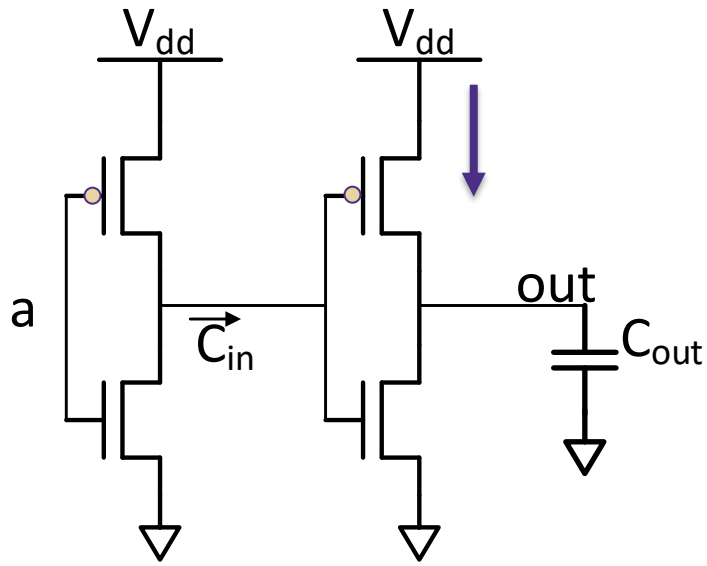


Dynamic Power Dissipation: Exercise



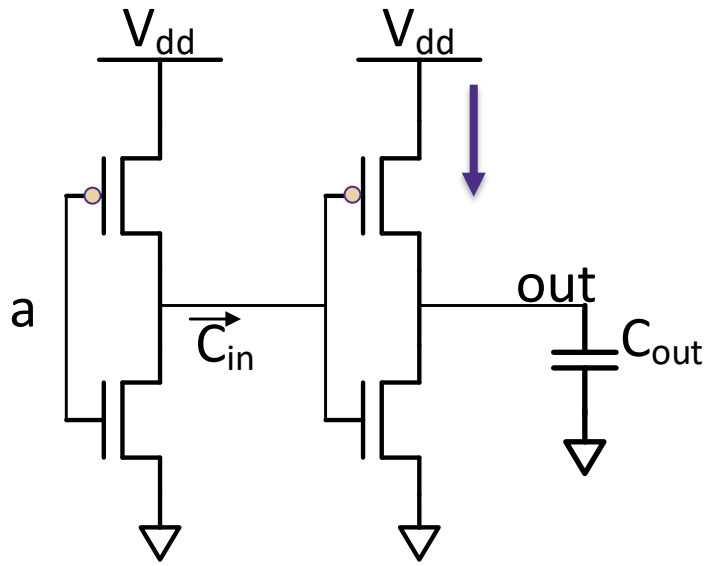
- Discuss the impact of the following on the absolute quantity of energy dissipation due to crowbar current in the second inverter (assume all other properties are constant)
  - V_{dd}
  - V_{th}
  - C_{in} (due to off-path loading)
  - C_{out}

Dynamic Power Dissipation: Exercise



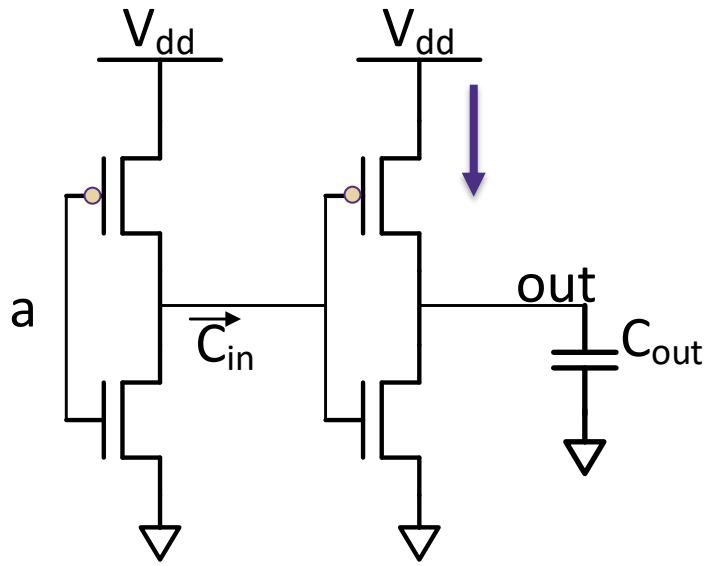
- Discuss the impact of the following on the absolute quantity of energy dissipation due to crowbar current in the second inverter (assume all other properties are constant)
  - V_{dd} (~k^2)
  - V_{th}
  - C_{in} (due to off-path loading)
  - C_{out}

Dynamic Power Dissipation: Exercise



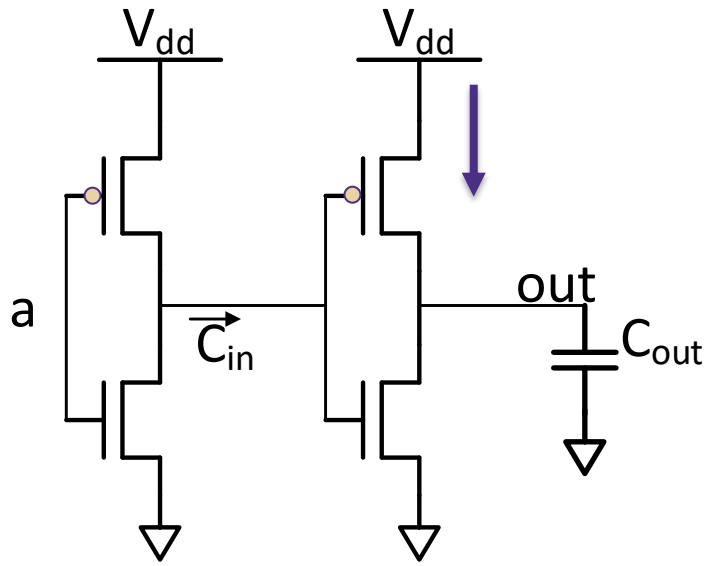
- Discuss the impact of the following on the absolute quantity of energy dissipation due to crowbar current in the second inverter (assume all other properties are constant)
  - V_{dd} (~k^2)
  - V_{th} (~-k)
  - C_{in} (due to off-path loading)
  - C_{out}

Dynamic Power Dissipation: Exercise



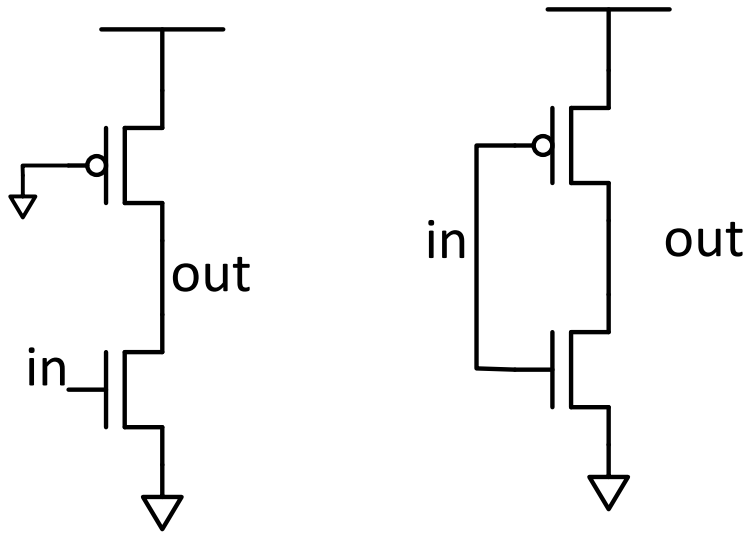
- Discuss the impact of the following on the absolute quantity of energy dissipation due to crowbar current in the second inverter (assume all other properties are constant)
  - V_{dd} (~k^2)
  - V_{th} (~-k)
  - C_{in} (due to off-path loading) (~k)
  - C_{out}

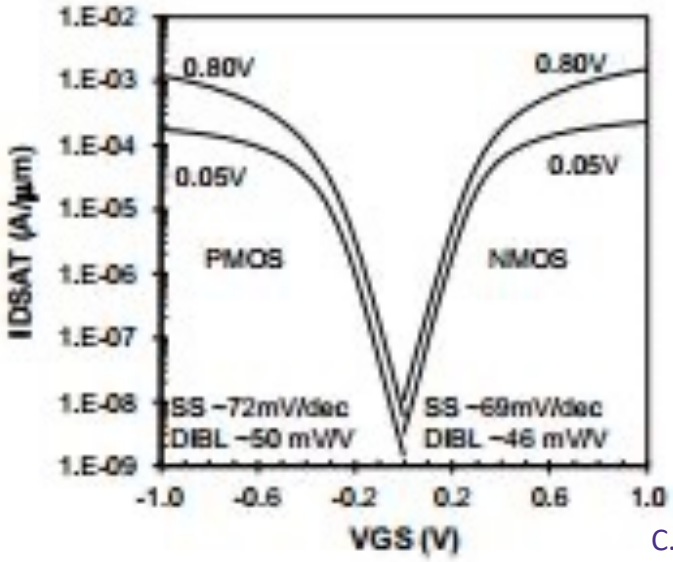
Dynamic Power Dissipation: Exercise



- Discuss the impact of the following on the absolute quantity of energy dissipation due to crowbar current in the second inverter (assume all other properties are constant)
  - V_{dd} (~k^2)
  - V_{th} (~-k)
  - C_{in} (due to off-path loading) (~k)
  - C_{out} (~1/k)

Static power dissipation

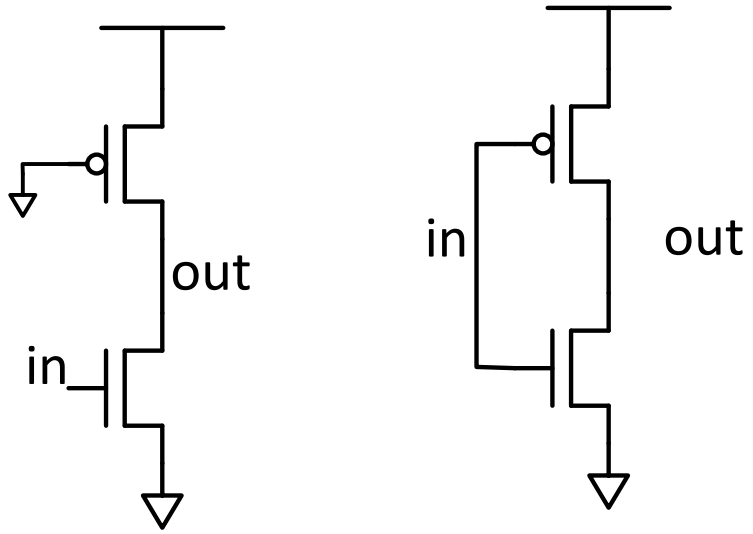


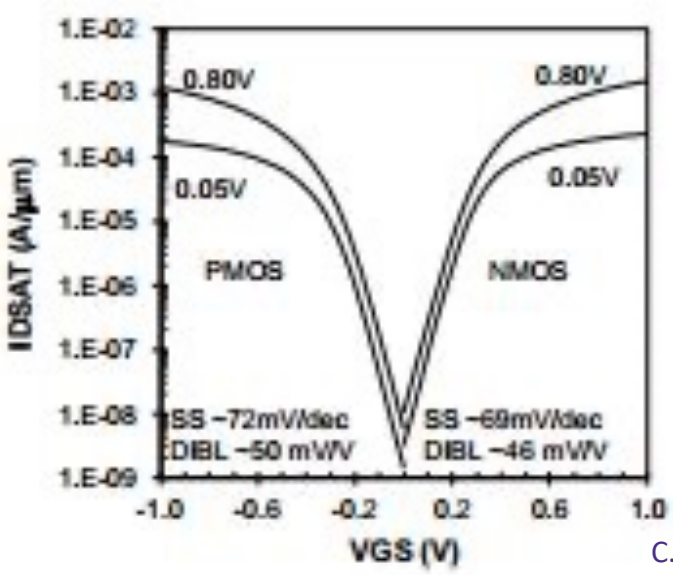


C.Auth (Intel)

- Power dissipation that does not depend on switching activity
- Often (almost always) state dependent
- Pseudo Nmos : When out = '0'
- 0^{th} order: CMOS ensures pullup-pulldown trees do not turn on simultaneously
- Still, devices leak : $I_{leak} = k10^{\frac{V_{gs}-V_{th}+\eta V_{ds}}{S}}(1 - e^{\frac{-V_{ds}}{V_T}})$
- S=sub-threshold swing, V_T = Thermal voltage

Static power dissipation





C.Auth (Intel)

- Sub-threshold Slope ideally 58mV/dec
  - ~90mV/dec more common among planar technologies
  - Tri-gate (Finfets) allow much lower swing
  - I_{on}/I_{off} in the order of 1000: Seems like a lot but you have LOTS of devices
- Other sources: Gate leakage, Junction leakage
- Leakage control drives V_{th} to remain high.
  - Will see how this affects power/performance

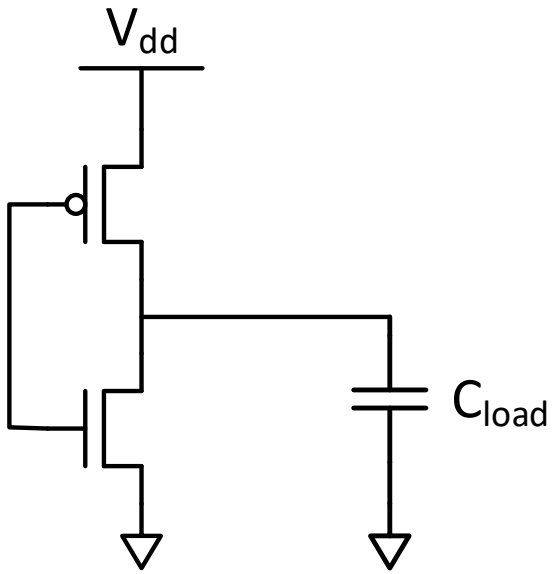
A Note on Energy vs. Power

- Energy : Work done to perform computation
- Power : Rate at which energy is dissipated

- E.g: Energy matters in…
  - Ultra Low Power (Energy scavenging systems): pJ per computation for viability
  - Battery operated systems : Battery life
  - Server machines : Wall-power

- E.g.: Power matters in …
  - Biomedical implants (e.g. neural/retinal) : Heat dissipation damages tissue
  - Hand-held devices : Surface temperature
  - Laptops/Desktops/Servers : Power-constrained performace

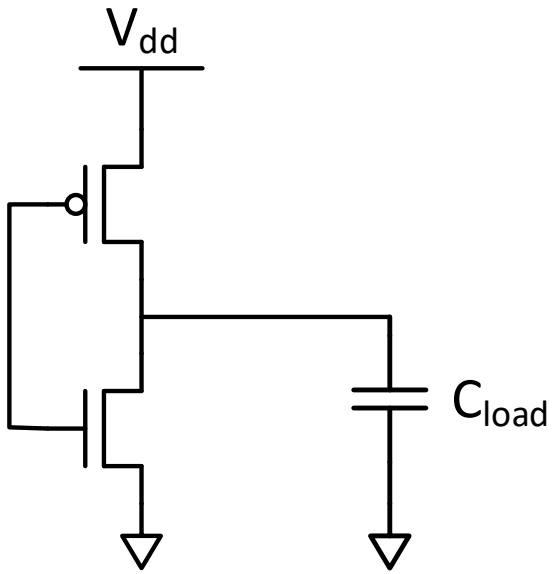
Energy Dissipation Example

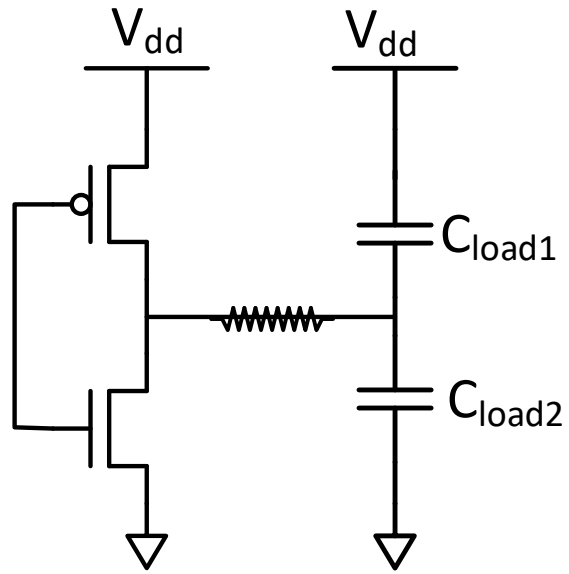


$$E_{dyn}=1/2sC_{load}V_{dd}^{2}$$

- Total energy dissipation
  - Energy dissipated during supply during rise = ?
  - Energy dissipated during fall = ?

Energy Dissipation Example

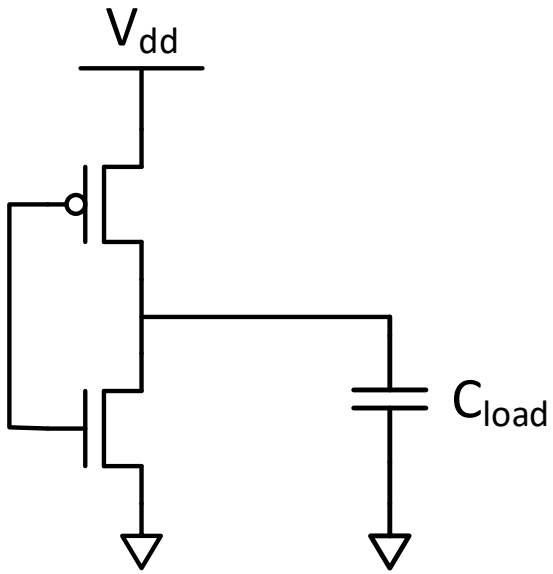




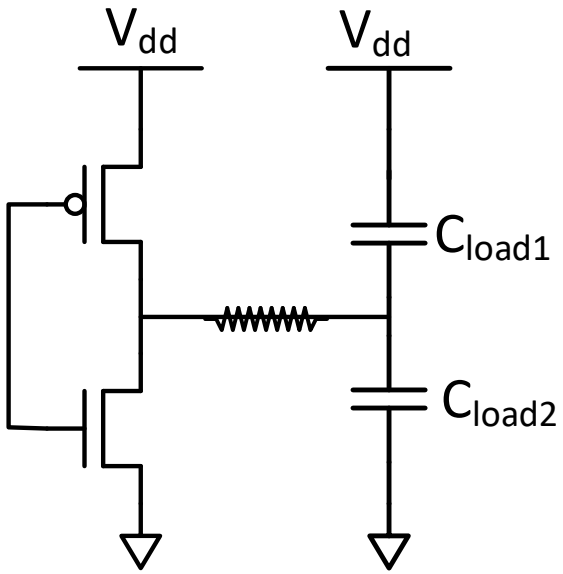
$E_{dyn}=1/2sC_{load}V_{dd}^{2}$

- Total energy dissipation
  - Energy dissipated during supply during rise = ?
  - Energy dissipated during fall = ?

Energy Dissipation Example



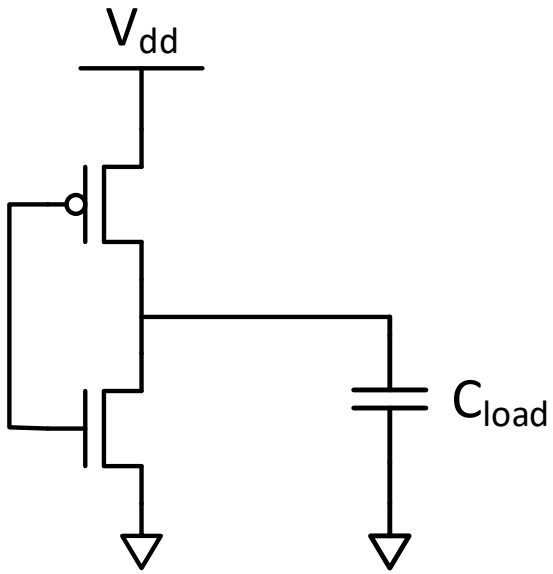
$E_{dyn}=1/2sC_{load}V_{dd}^{2}$



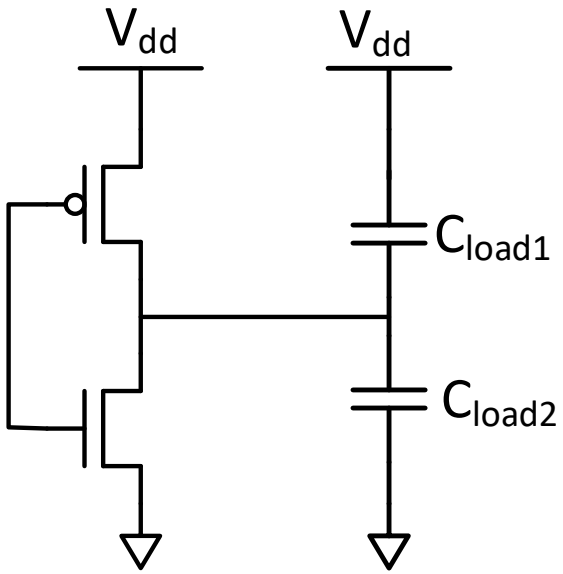
More typical,
Realistic scenario

- Total energy dissipation
  - Energy dissipated during supply during rise = ?
  - Energy dissipated during fall = ?

Energy Dissipation Example



$E_{dyn}=1/2sC_{load}V_{dd}^2$

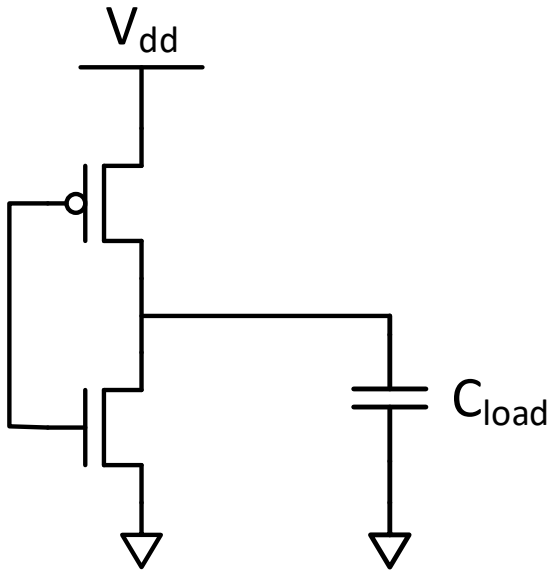


$E_{dyn}=1/2s(C_{load1}+C_{load2})V_{dd}^2$

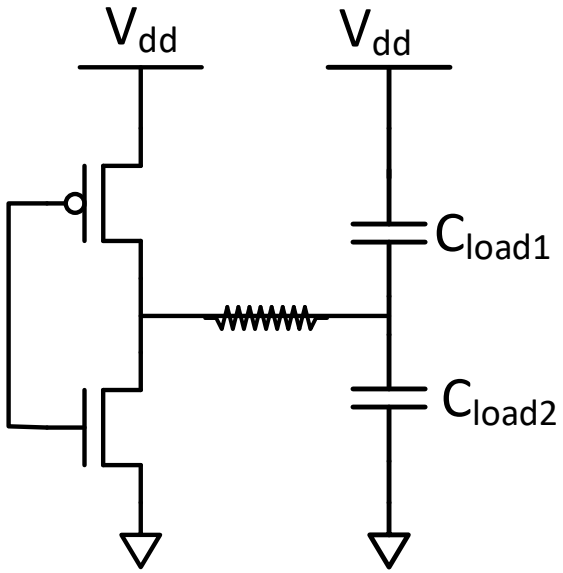
More typical,
Realistic scenario

- Total energy dissipation
  - Energy dissipated during supply during rise = ?
  - Energy dissipated during fall = ?

Energy Dissipation Example



$E_{dyn}=1/2sC_{load}V_{dd}^2$



$E_{dyn}=1/2s(C_{load1}+C_{load2})V_{dd}^2$

More typical,
Realistic scenario

- Total energy dissipation
  - Energy dissipated during supply during rise = ?
  - Energy dissipated during fall = ?

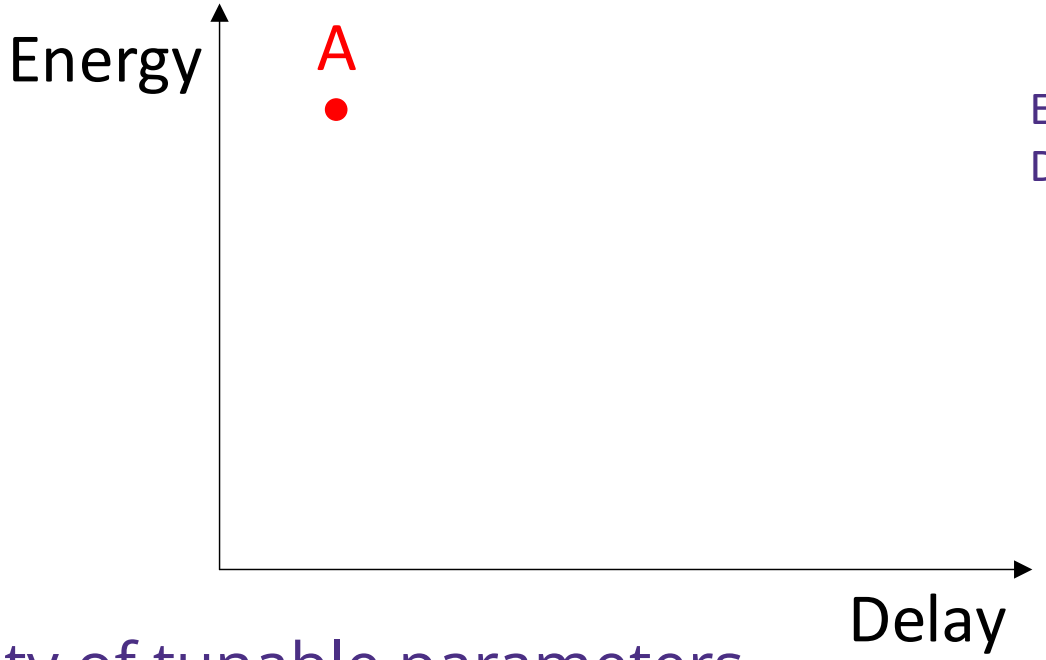
Energy Efficient Design

$$E = \frac{1}{2} sCV_{dd}^2 + V_{dd} I_{leak} T$$

$$\tau = \frac{kCV_{dd}}{\frac{1}{2}\beta(V_{dd} - V_{th})^{\alpha}}$$

- CMOS Low-power design: Reduce energy dissipation by
  - Effectively exploiting system-level structure
  - Reducing the terms in the energy equation, minimizing performance impact
- More on this topic late in the course
- Popular low-power techniques…
  - ↓s : Glitch removal, encoding, avoiding unnecessary computation
  - ↓C : Gate sizing, Memory hierarchy
  - ↓ V_{dd} : Operate the system(s) at the minimum voltage required at the given time

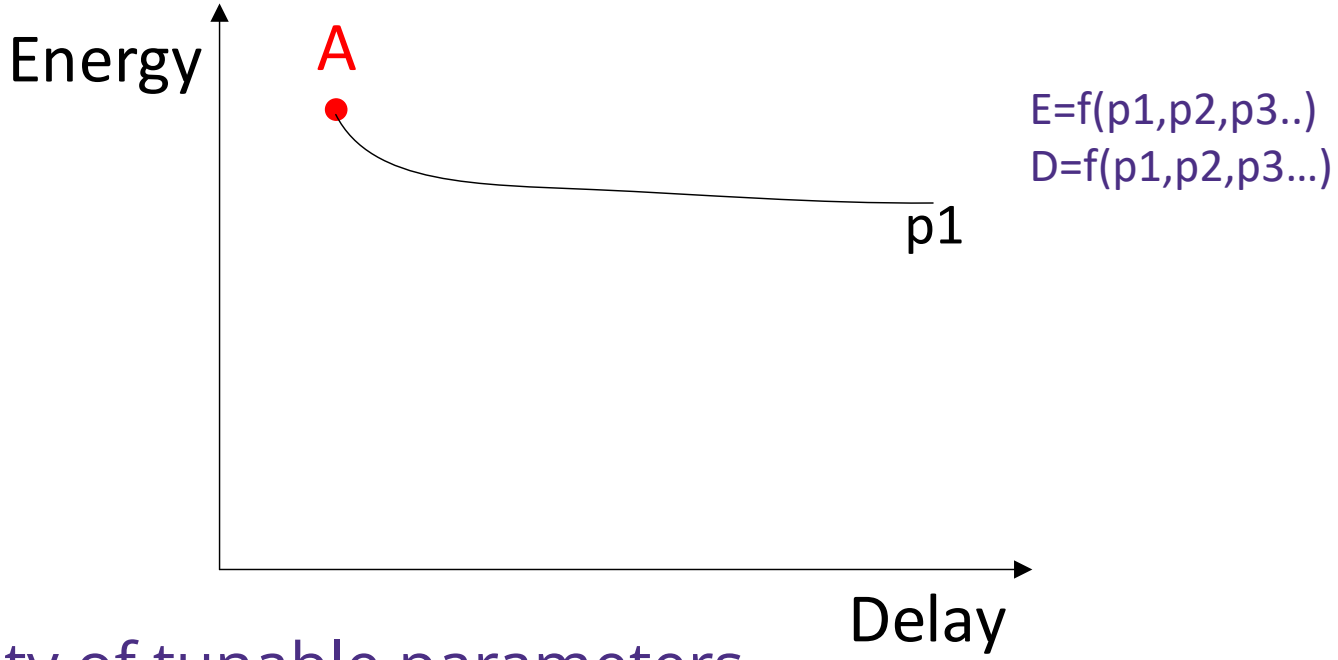
Energy-Delay tradeoff



E=f(p1,p2,p3..)
D=f(p1,p2,p3…)

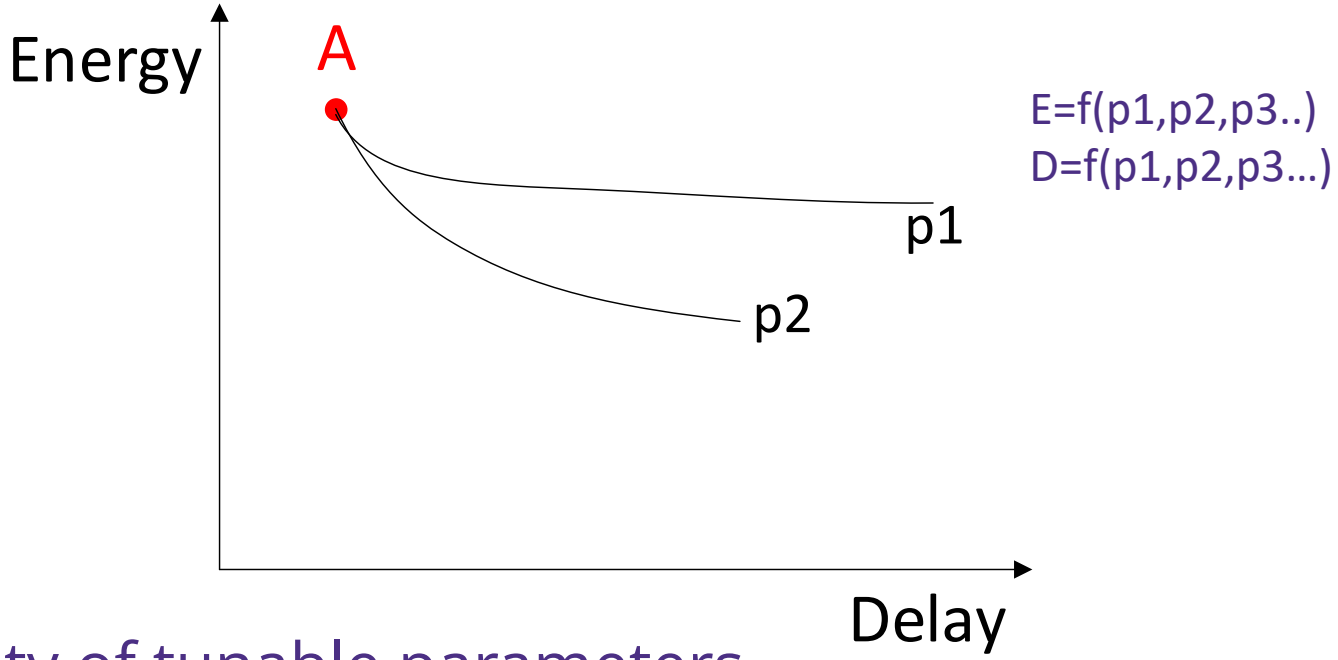
- Circuits offer a variety of tunable parameters
  - Sizing
  - Topology
  - Vdd
  - Vth
  - Implementation (Full-custom/SAPR)
  - Architecture/System-level (Pipelining, Parallel Processing, Clock-gating)

Energy-Delay tradeoff



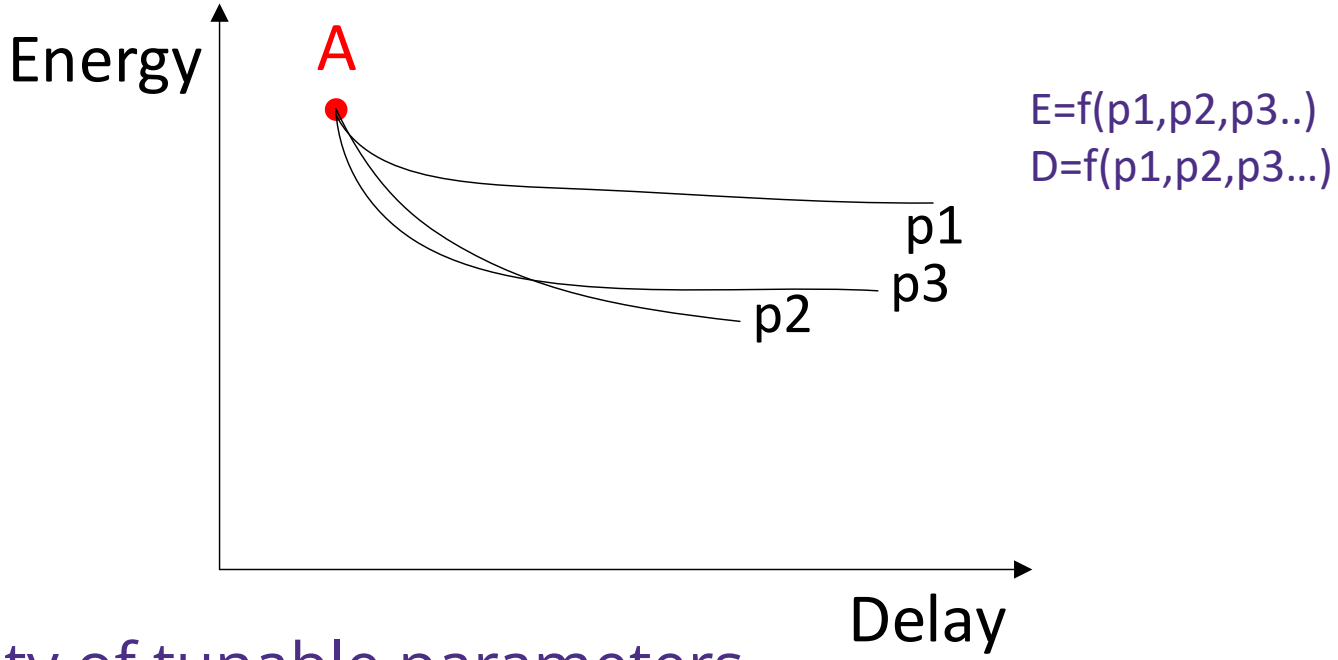
- Circuits offer a variety of tunable parameters
  - Sizing
  - Topology
  - Vdd
  - Vth
  - Implementation (Full-custom/SAPR)
  - Architecture/System-level (Pipelining, Parallel Processing, Clock-gating)

Energy-Delay tradeoff



- Circuits offer a variety of tunable parameters
  - Sizing
  - Topology
  - Vdd
  - Vth
  - Implementation (Full-custom/SAPR)
  - Architecture/System-level (Pipelining, Parallel Processing, Clock-gating)

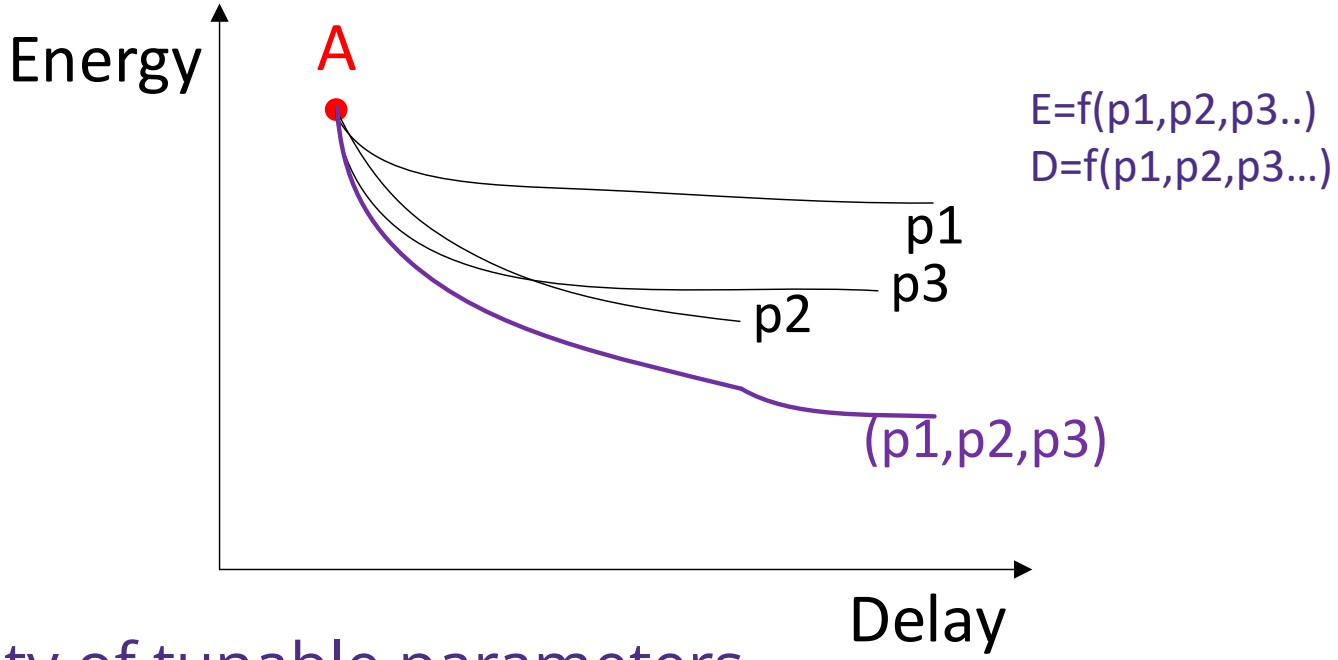
Energy-Delay tradeoff



E=f(p1,p2,p3..)
D=f(p1,p2,p3…)

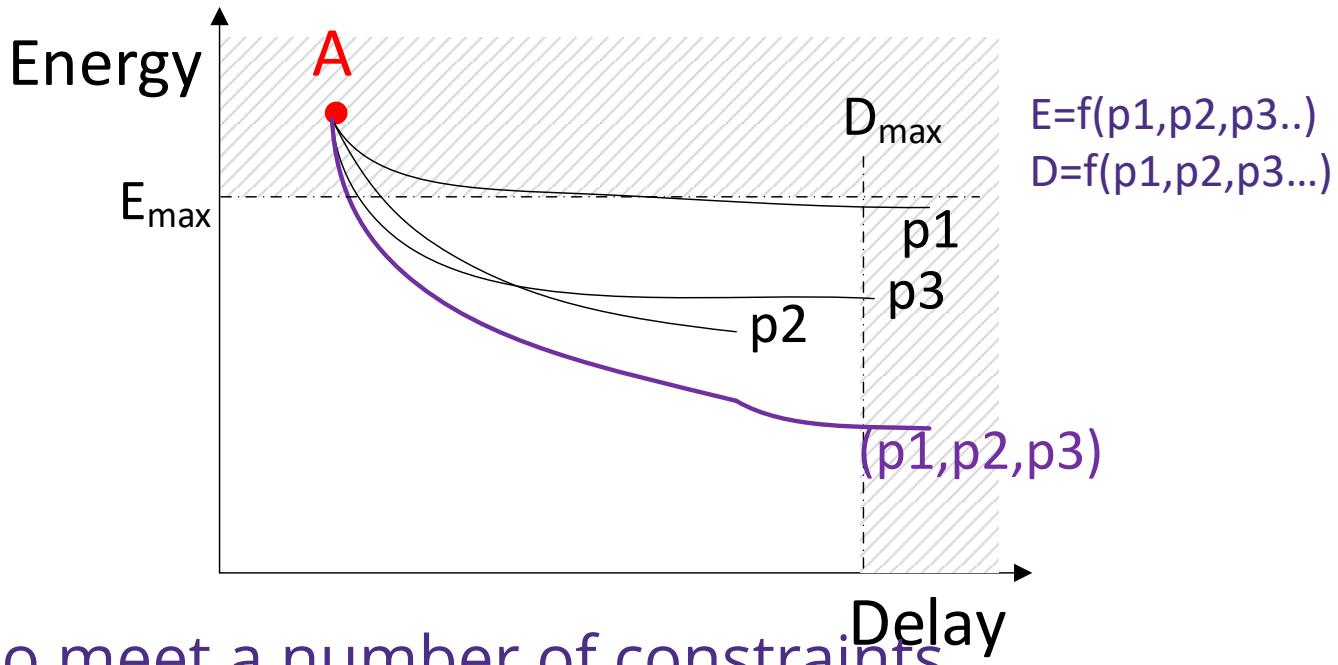
- Circuits offer a variety of tunable parameters
  - Sizing
  - Topology
  - Vdd
  - Vth
  - Implementation (Full-custom/SAPR)
  - Architecture/System-level (Pipelining, Parallel Processing, Clock-gating)

Energy-Delay tradeoff



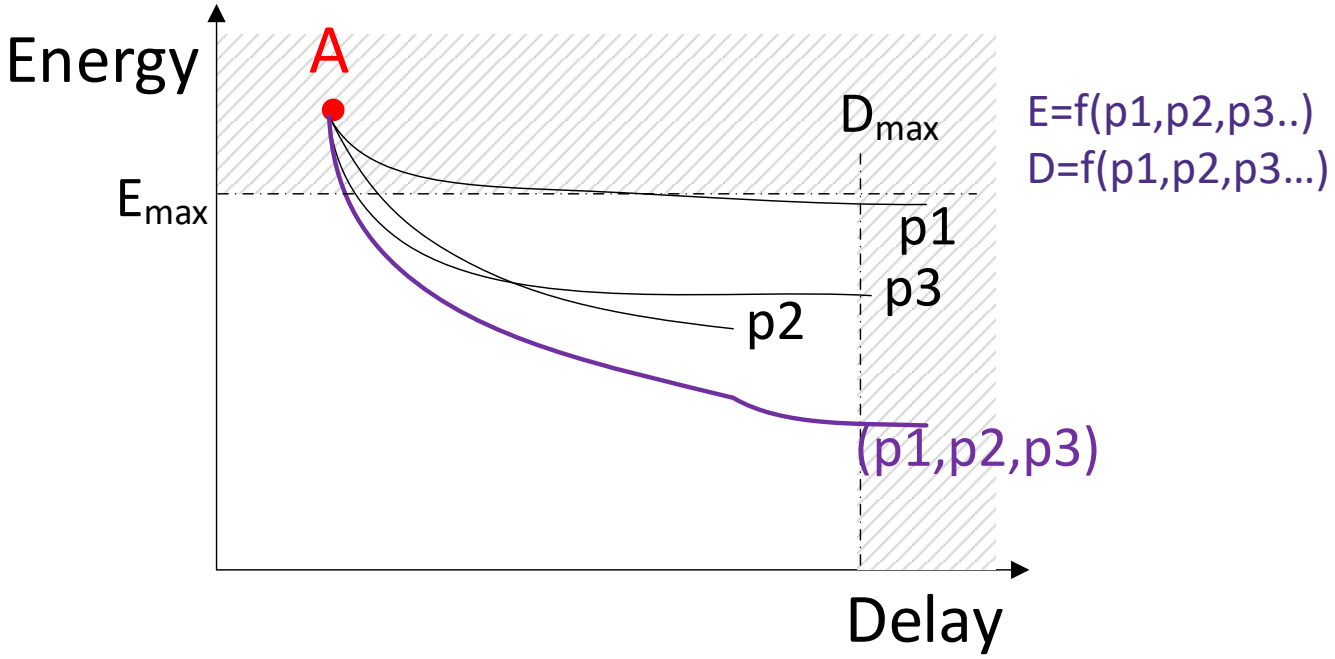
- Circuits offer a variety of tunable parameters
  - Sizing
  - Topology
  - Vdd
  - Vth
  - Implementation (Full-custom/SAPR)
  - Architecture/System-level (Pipelining, Parallel Processing, Clock-gating)

Energy-Delay tradeoff: Constraints



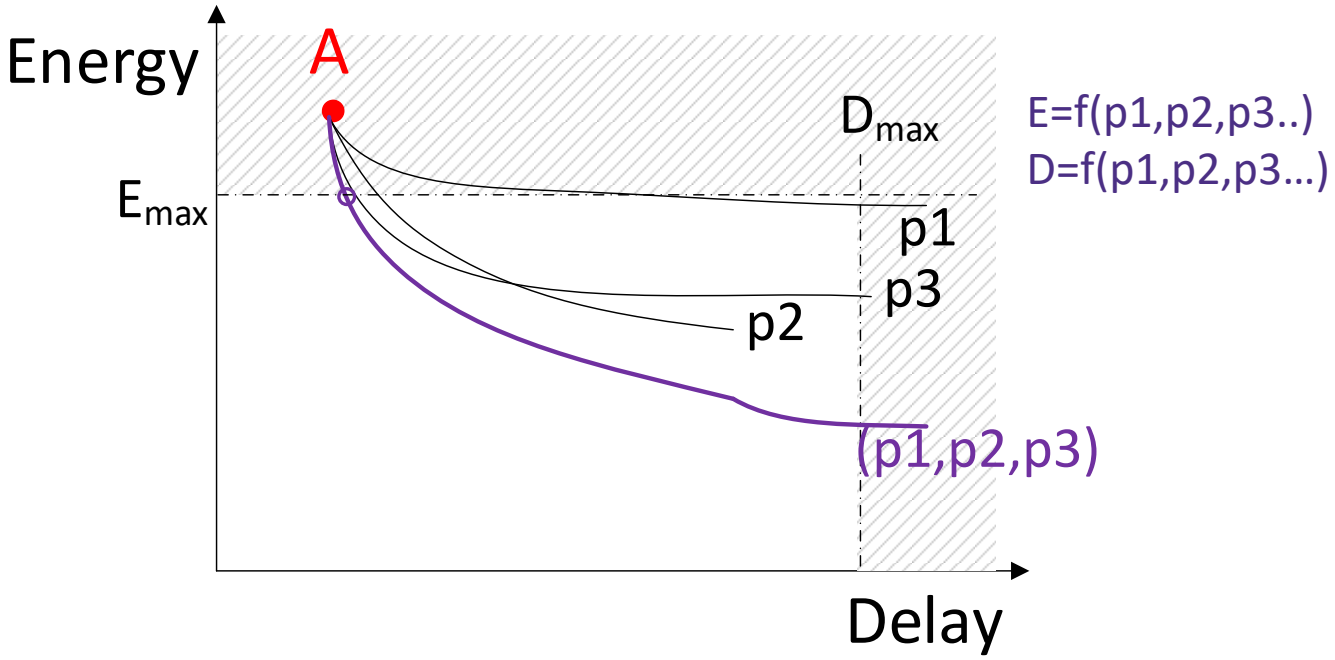
- Real systems need to meet a number of constraints
  - Design-Time
  - Energy
  - Power
  - Performance (→Frequency, →Delay)
  - System-level (Cost, Form factor, socket compatibility, board-compatibility)
- Constraints reduce parameter feasibility space

Energy-Delay tradeoff: Constraints



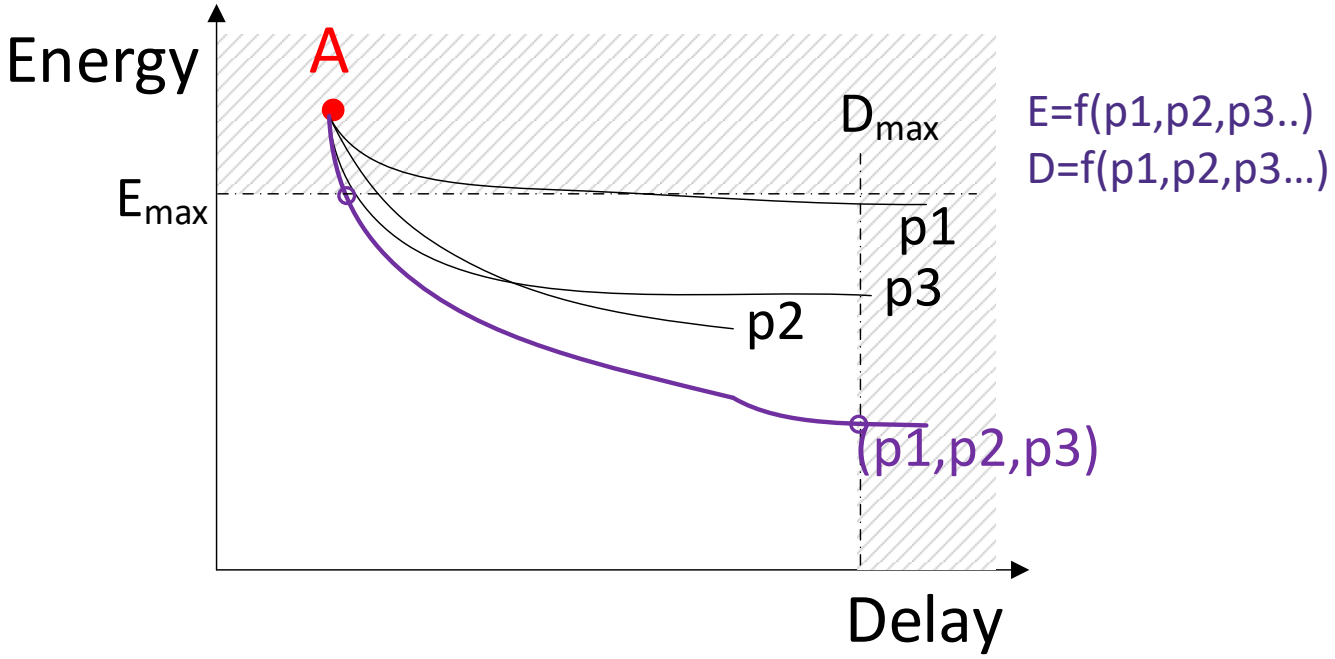
- Vary parameters (in the right combination) to achieve ***pareto-optimality*** between objectives
  - Not possible to reduce *E* with the given parameters without increasing *D*
- Points of interest
  - Dmin @ E=Emax
  - Emin @ D=Dmax

Energy-Delay tradeoff: Constraints



- Vary parameters (in the right combination) to achieve ***pareto-optimality*** between objectives
  - Not possible to reduce *E* with the given parameters without increasing *D*
- Points of interest
  - Dmin @ E=Emax
  - Emin @ D=Dmax

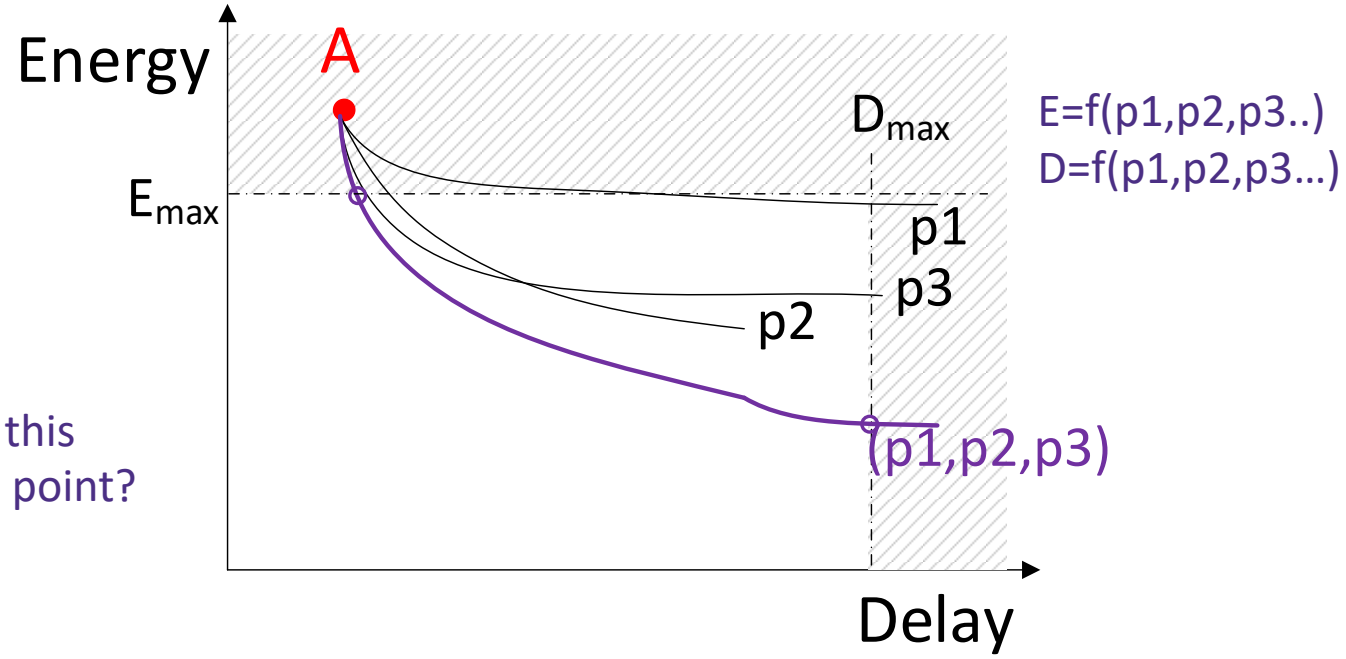
Energy-Delay tradeoff: Constraints



- Vary parameters (in the right combination) to achieve ***pareto-optimality*** between objectives
  - Not possible to reduce *E* with the given parameters without increasing *D*
- Points of interest
  - Dmin @ E=Emax
  - Emin @ D=Dmax

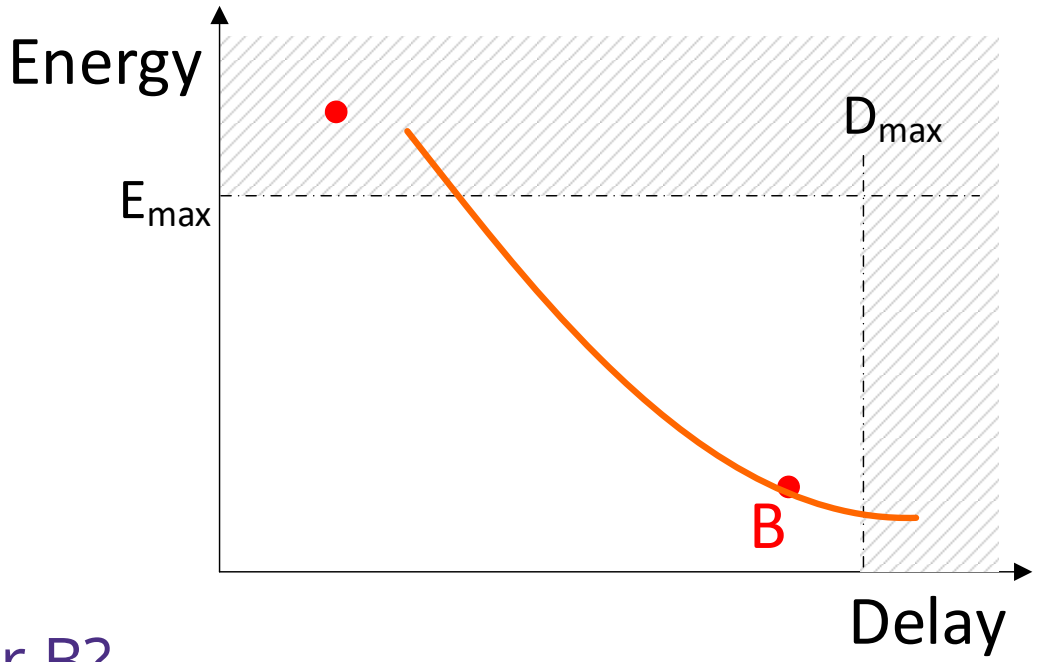
Energy-Delay tradeoff: Constraints

How do I trace this Pareto-optimal point?



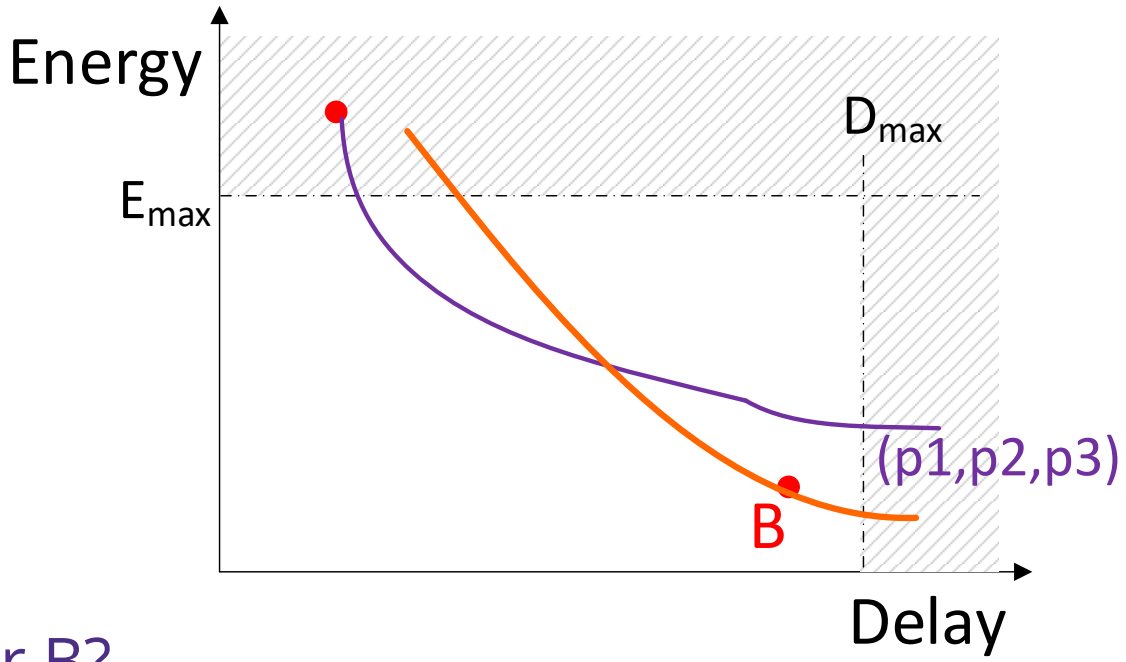
- Vary parameters (in the right combination) to achieve ***pareto-optimality*** between objectives
  - Not possible to reduce *E* with the given parameters without increasing *D*
- Points of interest
  - Dmin @ E=Emax
  - Emin @ D=Dmax

Note on Design Comparison



- Which is better A or B?
  - Not enough information
  - Depends on the specific constraints and objective function
  - Often need either pareto optimal plots, or comparison of optimal instances of each design that meets objectives
- Constraints affect feasibility of parameters (differently for designs)

Note on Design Comparison



- Which is better A or B?
  - Not enough information
  - Depends on the specific constraints and objective function
  - Often need either pareto optimal plots, or comparison of optimal instances of each design that meets objectives
- Constraints affect feasibility of parameters (differently for designs)

Reading assignment

- Required: W&H 5.2 – 5.2.5
- Optional: W&H 5.2.6