2. Overview of

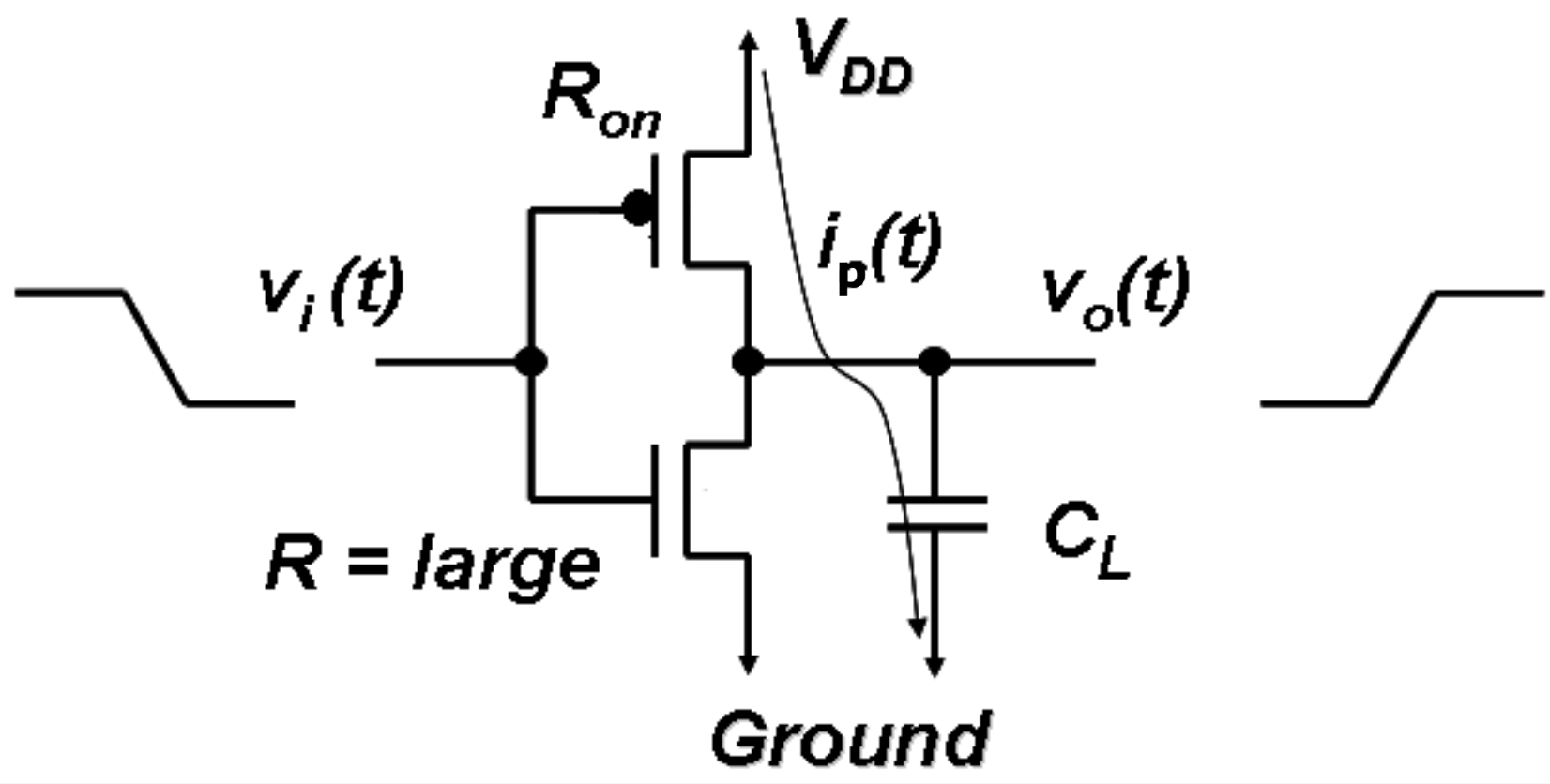
Power Dissipation in CMOS

Components of Power Dissipation

- Dynamic Component ($\mathbf{P_{dyn}}$)
  - Switching power ($\mathbf{P_{sw}}$)
    - Logic activity
    - Glitches
  - Short-circuit power ($\mathbf{P_{sc}}$)
- Static Component ($\mathbf{P_{stat}}$)
  - Leakage power
- $\boldsymbol{P_{total} = P_{dyn} + P_{stat}}$

  $\boldsymbol{= (P_{sw} + P_{sc}) + P_{stat}}$

Switching Power : P_{sw}



Switching Power : P_{sw}

$$P_{sw} = \frac{1}{T}\int_0^{T/2} i_N(t)\, V_{\mathrm{Out}}\, dt + \frac{1}{T}\int_{T/2}^{T} i_P(t)(V_{DD} - V_{\mathrm{Out}})dt$$

Since $i_N(t) = C_L dV_{\mathrm{Out}}/dt$ and analogously for $i_P(t)$,

$$P_{sw} = \frac{C_L}{T}\int_0^{V_{DD}} V_{\mathrm{Out}}\ dV_{\mathrm{Out}} + \frac{C_L}{T}\int_{V_{DD}}^{0} (V_{DD} - V_{\mathrm{Out}})d(V_{DD} - V_{\mathrm{Out}})$$

$$\boxed{P_{sw} = \frac{C_L {V_{DD}}^2}{T}}$$

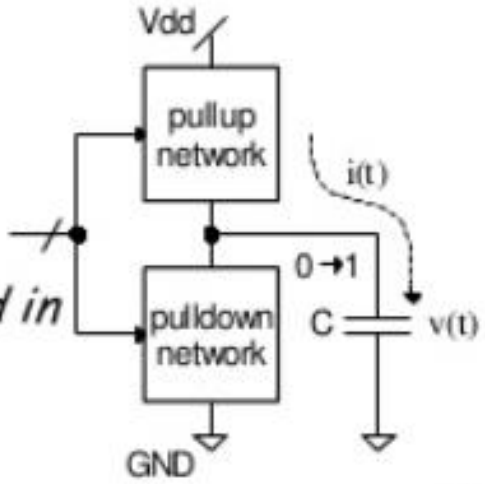
Switching Power : P_{sw}

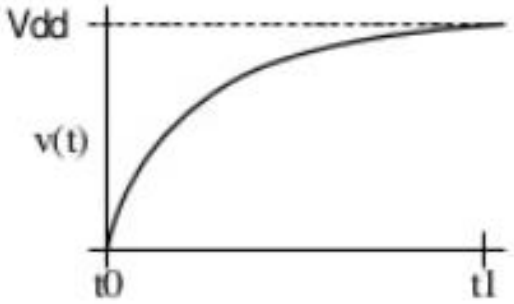
- Gate output rising transition
  - Energy dissipated in pMOS transistor = $C_L V^2_{DD}/2$
  - Energy stored in capacitor = $C_L V_{DD}^2/2$

Switching Power : P_{sw}

Switching Energy:
energy used to
switch a node

Calculate energy dissipated in pullup:





$$E_{sw} = \int_{t_0}^{t_1} P(t)dt = \int_{t_0}^{t_1} (V_{dd} - v) \cdot i(t)dt = \int_{t_0}^{t_1} (V_{dd} - v) \cdot c(dv/dt)\,dt =$$

$$= cV_{dd} \int_{t_0}^{t_1} dv - c\int_{t_0}^{t_1} v \cdot dv = cV_{dd}^{\,2} - 1/2cV_{dd}^{\,2} = \boxed{1/2\,cV_{dd}^{\,2}}$$

Energy supplied *Energy stored* *Energy dissipated*

An equal amount of energy is dissipated on pulldown

Switching Power : P_{sw}

- Gate output falling transition
  - Energy dissipated in nMOS transistor = $C_L V^2_{DD}/2$
- Energy dissipated per transition = $C_L V^2_{DD}$

$\therefore$ Total Average Power dissipation:

$$\mathbf{P_{sw} = C_L V^2_{DD} / T}$$

Lowering Switching Power:

Capacitance:
Function of fan-out, wire length and transistor sizes.

Clock Frequency:
Increasing

$$P_{sw} = C_L V_{DD}^2 f_{CLK} \alpha$$

Supply voltage:
Dropping with
Successive generations

Activity factor:
How often, on an average, does a node switch?

Switching Activity Factor

Type of Logic Function: NOR vs. XOR

Example: Static 2-input NOR Gate

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

Assume signal probabilities

$p_{A=1} = 1/2$

$p_{B=1} = 1/2$

Then transition probability

$p_{0\rightarrow1} = p_{Out=0} \times p_{Out=1}$

$= 3/4 \times 1/4 = 3/16$

If inputs switch every cycle

$\alpha_{0\rightarrow1} = 3/16$

Switching Activity Factor

Type of Logic Function: NOR vs. XOR

Example: Static 2-input XOR Gate

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

Assume signal probabilities

$p_{A=1} = 1/2$

$p_{B=1} = 1/2$

Then transition probability

$p_{0\to1} = p_{Out=0} \times p_{Out=1}$

$= 1/2 \times 1/2 = 1/4$

If inputs switch in every cycle

$\alpha_{0\to1} = 1/4$

Dynamic Power Consumption is Data Dependent

Dynamic 2-input NOR Gate

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

Assume signal probabilities

$P_{A=1} = 1/2$

$P_{B=1} = 1/2$

Then transition probability

$P_{0\rightarrow1} = P_{out=0} \times P_{out=1}$

$= 3/4 \times 1 = 3/4$

Switching activity always higher in dynamic gates!

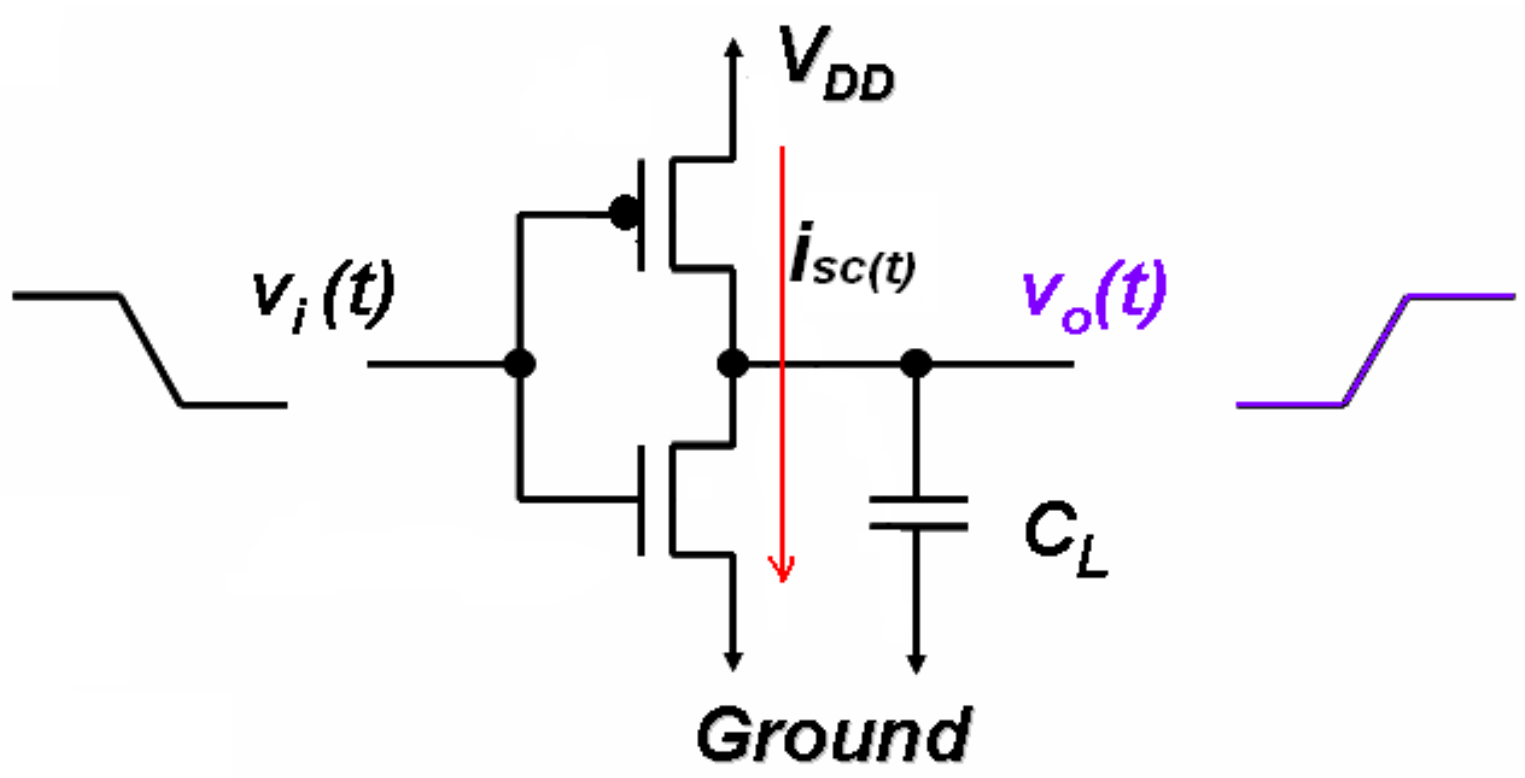
$P_{0\rightarrow1} = P_{out=0}$

Components of Power Dissipation

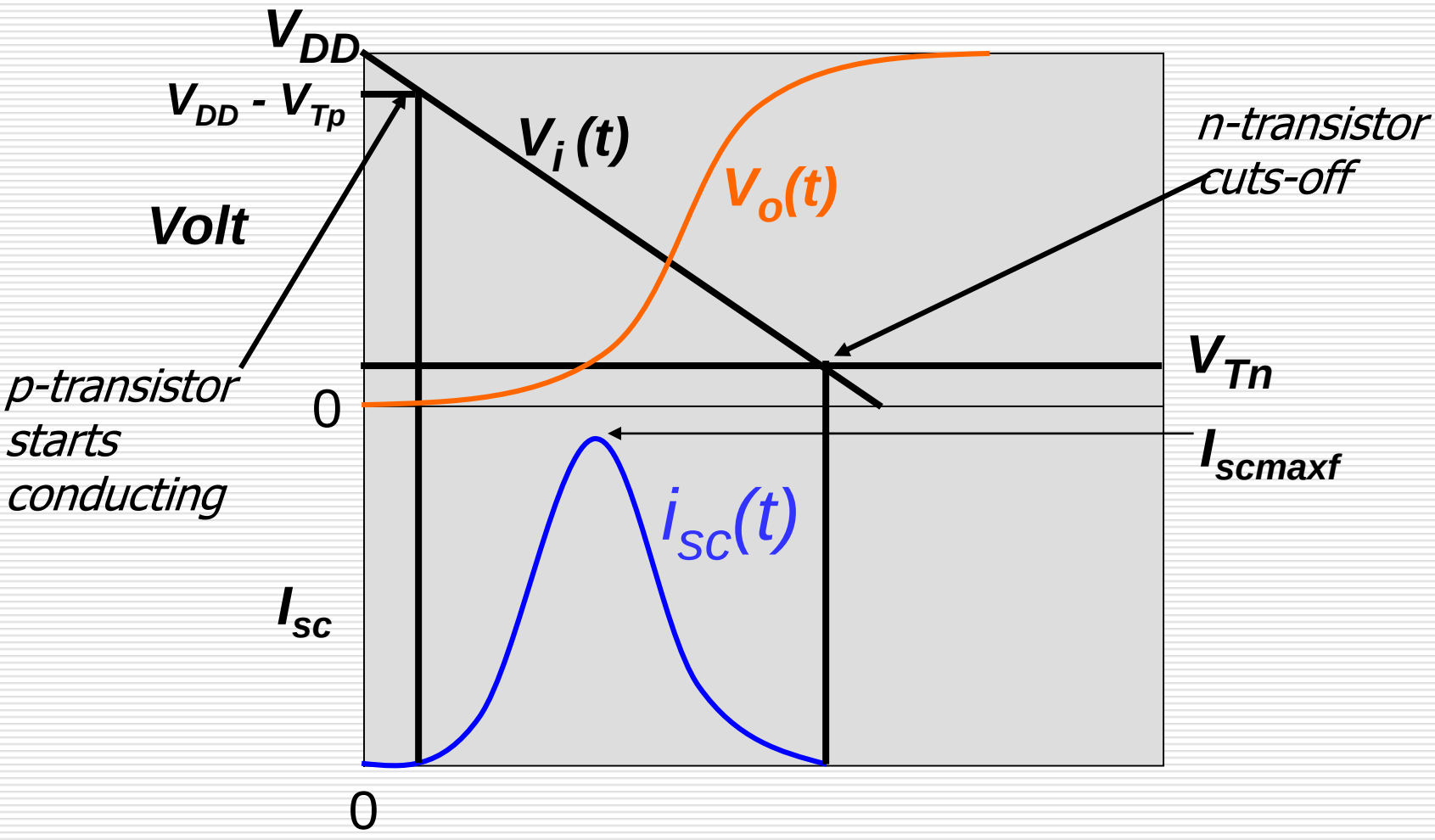
- Dynamic Component ($\mathbf{P_{dyn}}$)
  - Switching p.d. ($\mathbf{P_{sw}}$)
    - Logic activity
    - Glitches
  - **Short-circuit p.d. ($\mathbf{P_{sc}}$)**
- Static Component ($\mathbf{P_{stat}}$)
  - Leakage p.d.

$$P_{total} = P_{dyn} + P_{stat}$$
$$= (P_{sw} + P_{sc}) + P_{stat}$$

Short Circuit Power of a Transition: P_{sc}



Short Circuit Current, $i_{sc}(t)$



Short Circuit Power

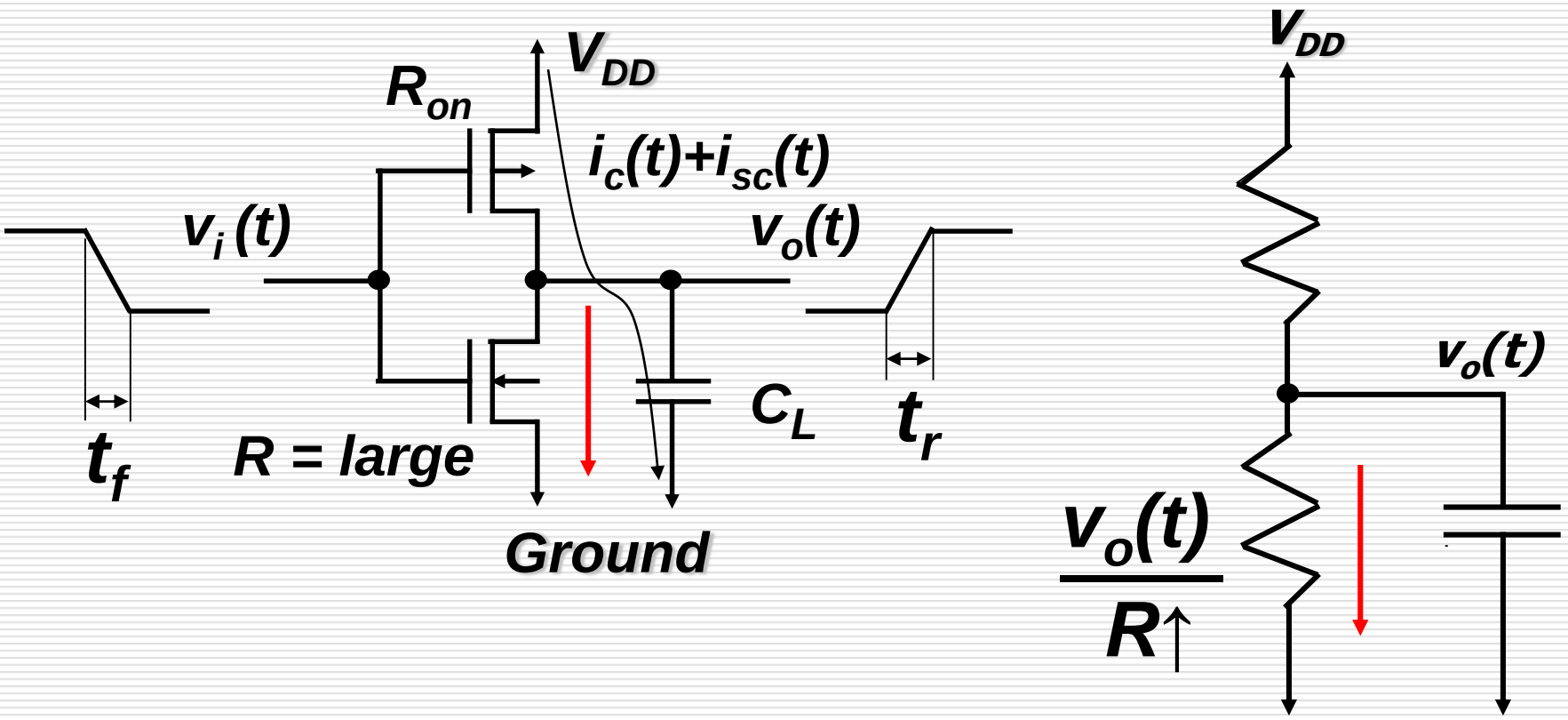
For an unloaded inverter,

$$P_{SC} = \frac{\beta}{12}(V_{DD} - 2V_{TH})^3 \frac{\tau}{T}$$

Where, τ = input rise time = input fall time

- Increases with the size (or gain, β) of transistors
- Increases with rise and fall times of input
- Decreases and eventually becomes zero when V_{DD} is scaled down but the threshold voltages are not scaled down.
- Decreases with load capacitance, C_L
- Largest when $C_L = 0$

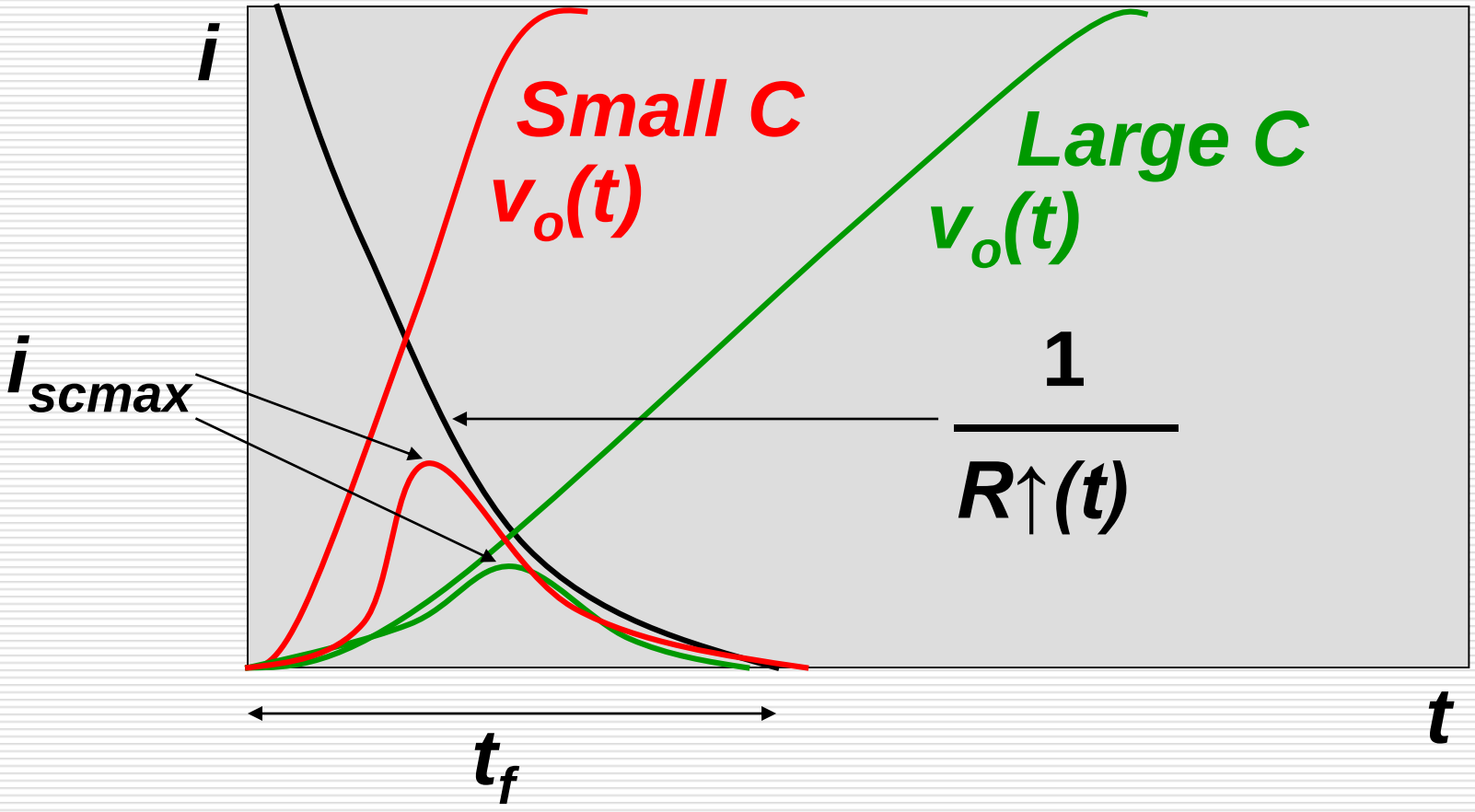
P_{SC}, Rise Time and Capacitance



i_{sc}, Rise Time and Capacitance

$$I_{sc}(t) = \frac{v_o(t)}{R\uparrow(t)} = \frac{V_{DD}\left[1 - exp\left(\frac{-t}{R\downarrow(t)\,C}\right)\right]}{R\uparrow(t)}$$

i_{scmax}, Rise Time and Capacitance



Short-Circuit Power, P_{sc}:

- If the inverter is lightly loaded,
  - the output rise-and-fall times become shorter than the input rise-and-fall times
  - the short circuit dissipation increases to become comparable to switching dissipation
- Therefore to minimize dissipation, an inverter should be designed in such a way that the input rise-and-fall times are about equal to the output rise-and-fall times

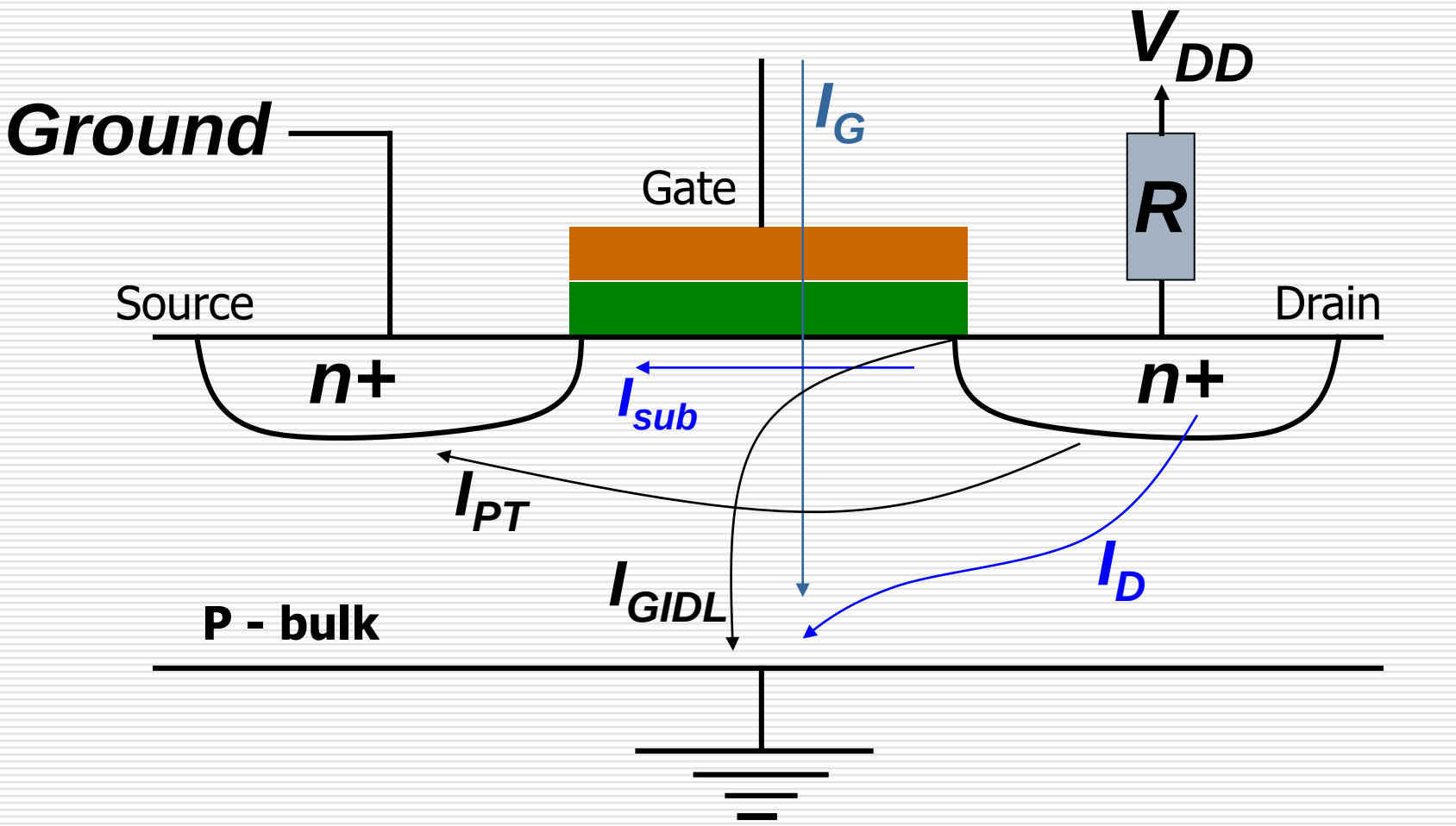
Summary: Short-Circuit Power

- Short-circuit power is consumed by each transition (increases with input transition time)
- Reduction requires that gate output transition should not be faster than the input transition (faster gates can consume more short-circuit power)
- Increasing the output load capacitance reduces short-circuit power
- Scaling down of supply voltage with respect to threshold voltages reduces short-circuit power; completely eliminated when $V_{DD} \leq |V_{tp}| + V_{tn}$.

Components of Power

- Dynamic
  - Signal transitions
    - Logic activity
    - Glitches
  - Short-circuit
- Static
  - Leakage

Leakage Power:



nMOS Transistor

Leakage Current Components

- Subthreshold leakage, I_{sub}
- Reverse bias pn junction leakage, I_D
- Gate induced drain leakage, I_{GIDL} due to tunneling at the gate-drain overlap
- Drain source punch-through, I_{PT} due to short channel and high drain-source voltage
- Gate tunneling, I_G through thin oxide; *may become significant with scaling*

1. Subthreshold Leakage, I_{sub}

- Occurs when $V_{GS} < V_{TH}$ (weak inversion), where minority carrier concentration is small, but not zero
- Subthreshold conduction is dominated by the diffusion current
- This leakage current is the dominant component in the modern device OFF-state leakage due to the low V_{TH} that is used

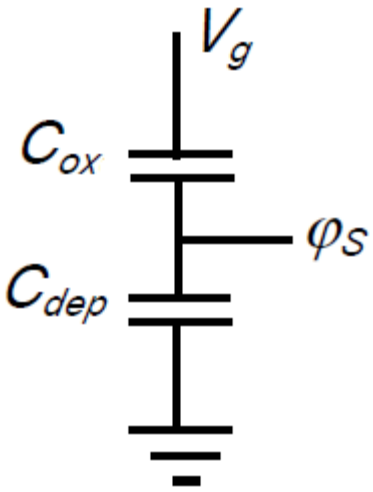
Subthreshold Leakage, I_{sub}

a) For Long Channel Devices:

$$I_{sub} = \mu_0 C_{ox} (W/L) v_T^2 \exp\{(V_{GS} - V_{TH}) / \eta v_T\}$$

μ_0: *zero bias carrier surface mobility*
C_{ox}: *gate oxide capacitance per unit area*
L: *channel length*
W: *gate width*
$v_T = kT/q$: *thermal voltage*
η: *a technology parameter*

Where, $\eta = 1 + \frac{C_{dep}}{C_{ox}}$



$$C_{dep} = \frac{\epsilon_{si}}{w_d}, \quad C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}.$$

where ϵ_{ox} and ϵ_{si} denote the dielectric constants of the oxide and silicon, resp, w_d is the depletion width under the *channel and* t_{ox} is the gate oxide thickness.

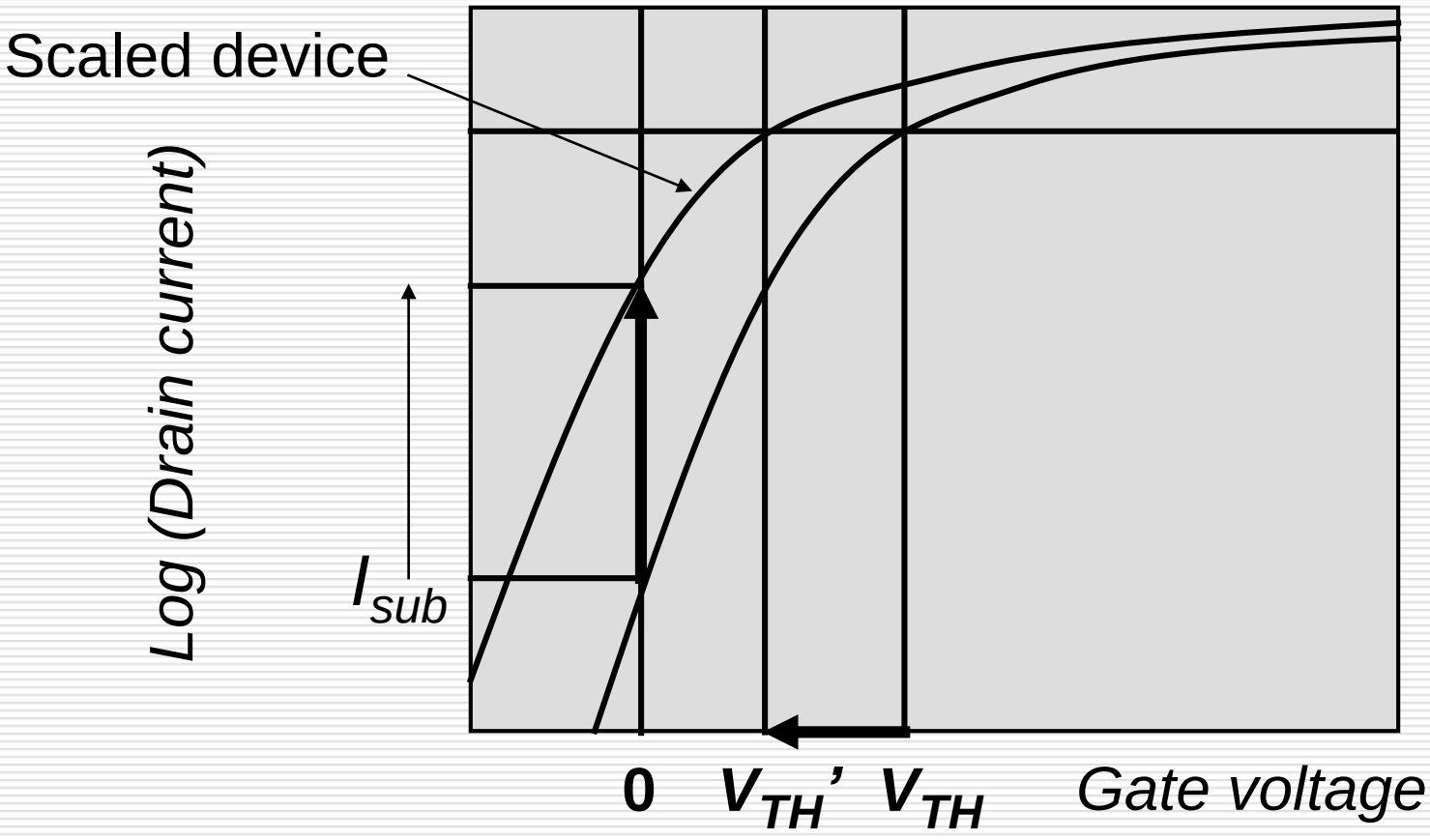
b) For Short Channel Devices:

$$I_{sub} = \mu_0 C_{ox}(W/L)v_T^2 \exp\{(V_{GS} - V_{TH} + nV_{DS})/\eta v_T\}$$

V_{DS} = *drain to source voltage*
η: *a proportionality factor*
n: a DIBL constant

W. Nebel and J. Mermet (Editors), *Low Power Design in Deep Submicron Electronics*, Springer, 1997, Section 4.1 by J. Figueras, pp. 81-104

Increased Subthreshold Leakage



Techniques to Reduce I_{sub}

- For given W and L, there are two ways to minimize I_{off}
- The first is to choose a large V_{th}. This is not desirable because a large V_{th} reduces I_{on} and therefore increases the gate delays
- The preferable way is to reduce the η. That can be done by increasing C_{ox}, i.e. using a thinner t_{ox}, and by decreasing C_{dep}, i.e. increasing W_{dep}
- One more way to reduce I_{off} is to operate the transistors at a lower temperature.
- This above approach is valid in principle but adds considerable cost for cooling.

2. Reverse Bias pn Junction Leakage, I_D

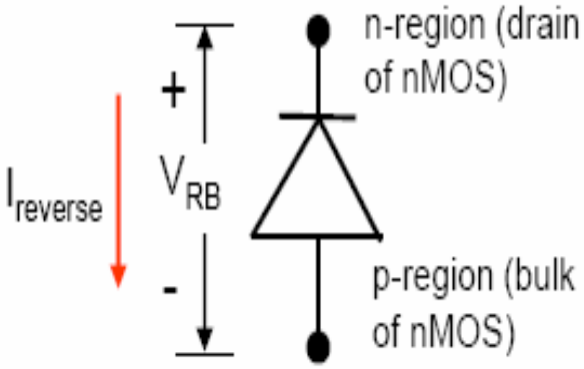
- This has 2 components:
  - Minority carrier diffusion/drift near the edge of the depletion region
  - Due to electron-hole pair generation in the depletion region of the reverse bias junction
- I_D is a function of junction area and doping concentration and strongly correlates with temperature

$$I = I_S \left(e^{V_D/(nV_T)} - 1 \right),$$

- Where I is the diode current,
- I_S is the reverse bias saturation current (or scale current),
- V_D is the voltage across the diode,
- V_T is the thermal voltage,
- n is the *ideality factor*, also known as the *quality factor* or sometimes *emission coefficient.*

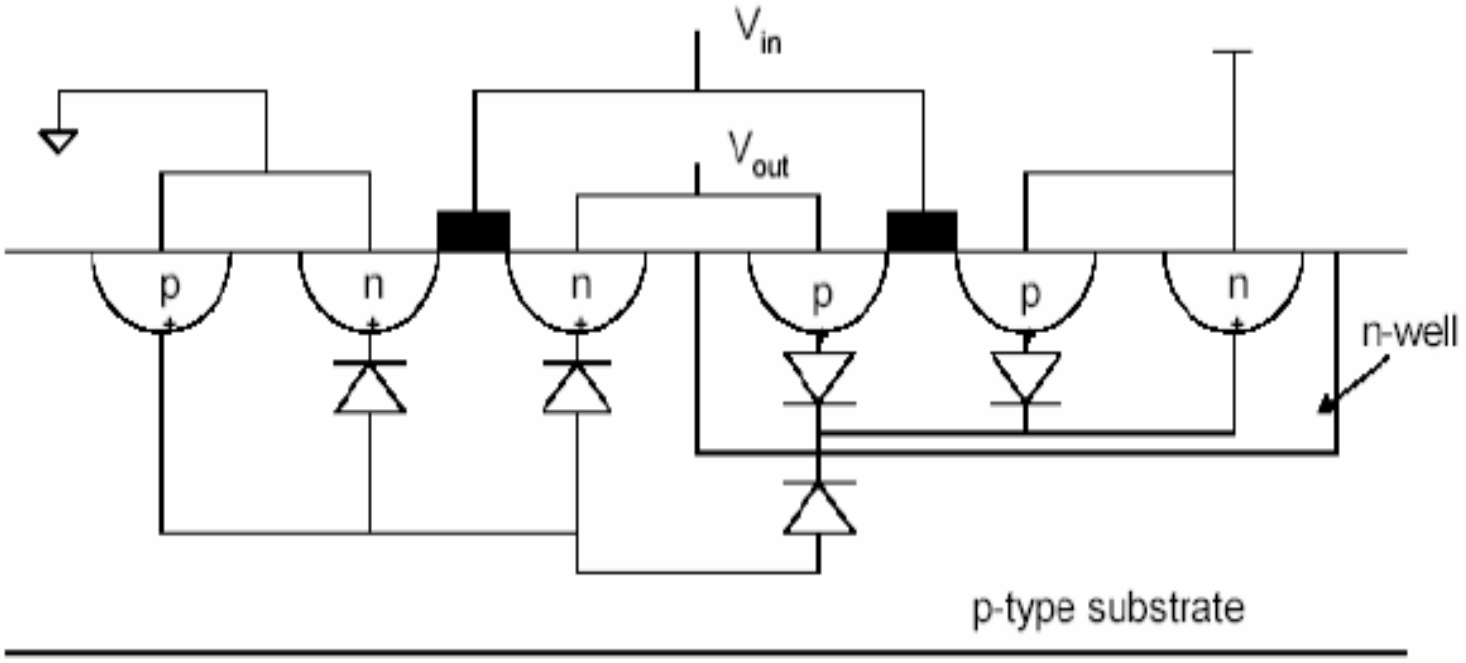
- The reverse leakage current of a pn-junction is expressed by

$$I_{reverse} = A \cdot J_s \cdot (1 - e^{-\frac{V_{RB}}{n\vartheta_T}})$$

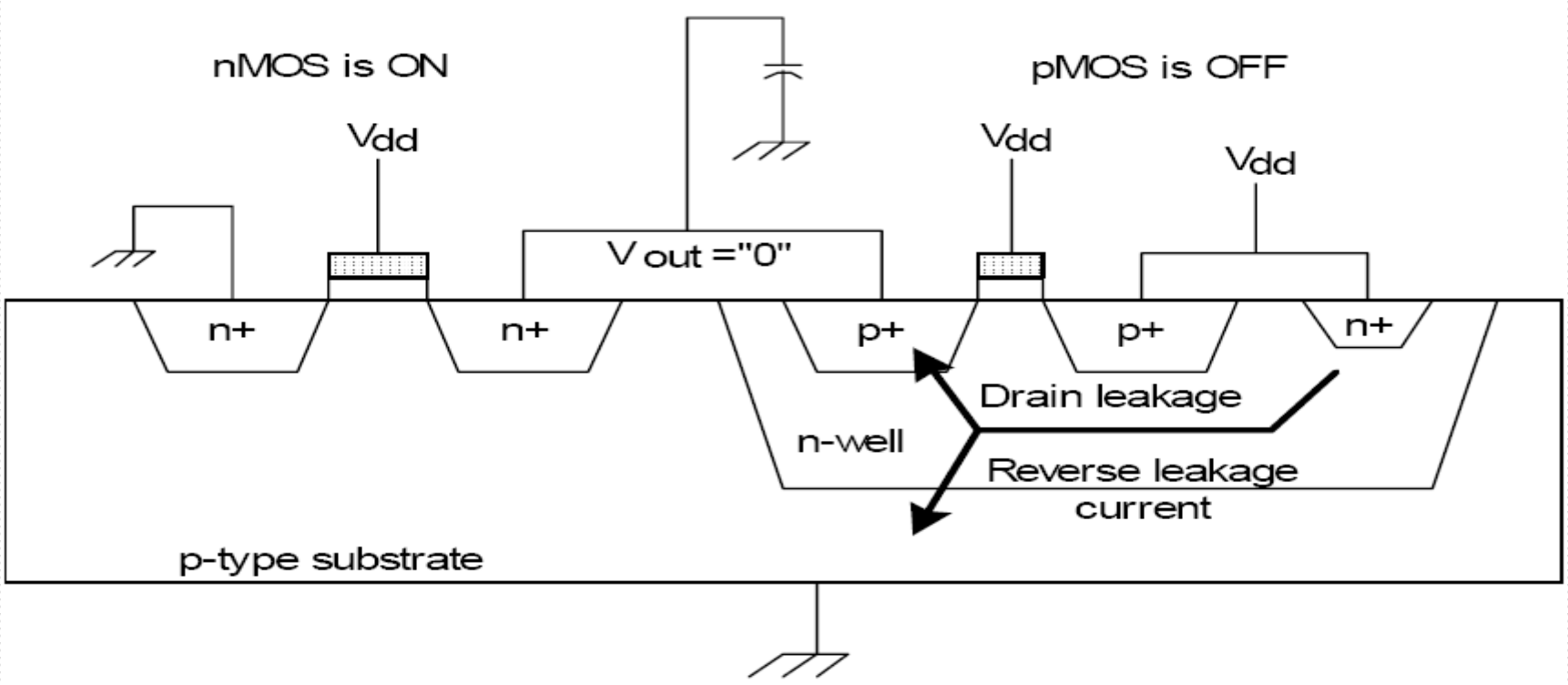


- A : the junction area
- J_s : the *maximum reverse saturation current density* (typically 1-5 pA/μm^2)
- n : the *emission coefficient*, usually set to 1, although can be larger depending on the type of junction
- V_{RB} : the *reverse bias voltage* across the junction, i.e., the voltage of drain diffusion with respect to the bulk or well
- $\upsilon_T = kT/q$ denotes the thermal voltage at absolute junction temperature, T

P-N Junction Reverse-Biased Currents:



- Consider a CMOS inverter with a high input voltage
  - Although pMOS transistor is turned off, there will be a reverse potential difference of V_{DD} between its drain and the n-well



3. Gate Induced Drain Leakage, I_{GIDL}

- I_{GIDL} is due to the high field effect at the drain junction of the MOSFET.
- This avalanche like multiplicative tunneling is called band to band tunneling(BBT or BTBT)

4. Drain Source Punch-Through, I_{PT}

- When the drain is at high enough voltage wrt souce, the depletion region around the drain may reach the source
- Causes the current to flow without the gate control
- It is called punch through condition

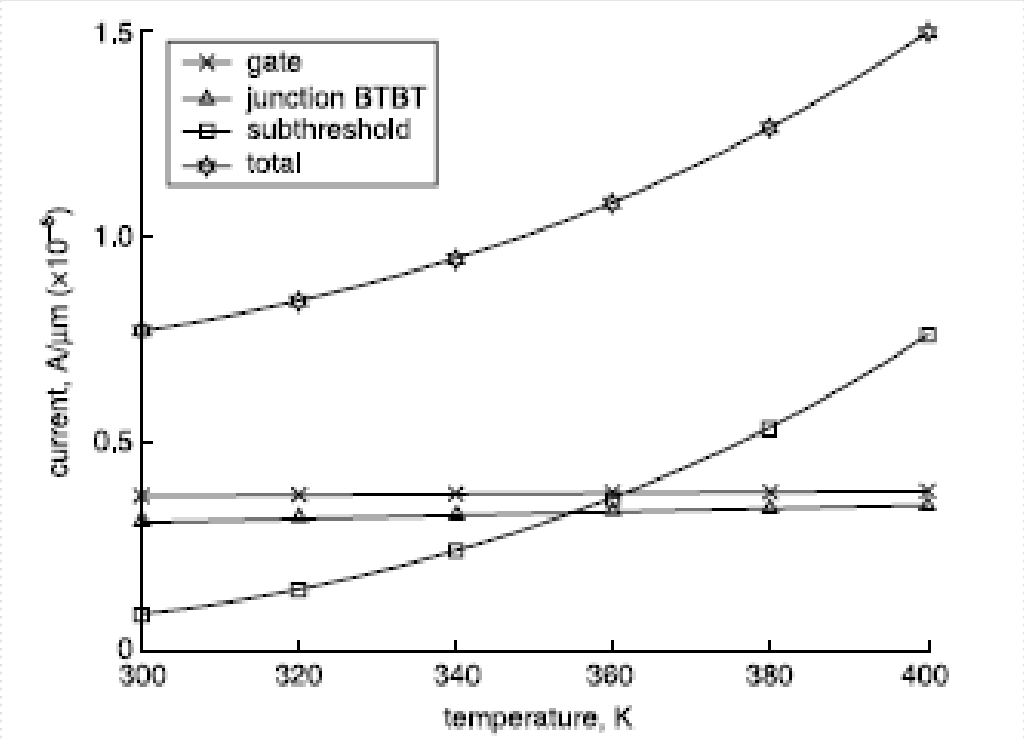
5. Gate Tunneling, I_G through thin oxide

Factors affecting leakage power

- The total leakage current $\mathbf{I_{OFF}}$ is influenced by the following factors
  - Temperature
  - Threshold voltage
  - Gate oxide thickness
  - Channel/surface doping profile
  - Drain/source junction depths
  - Channel physical dimensions
  - V_{DD}

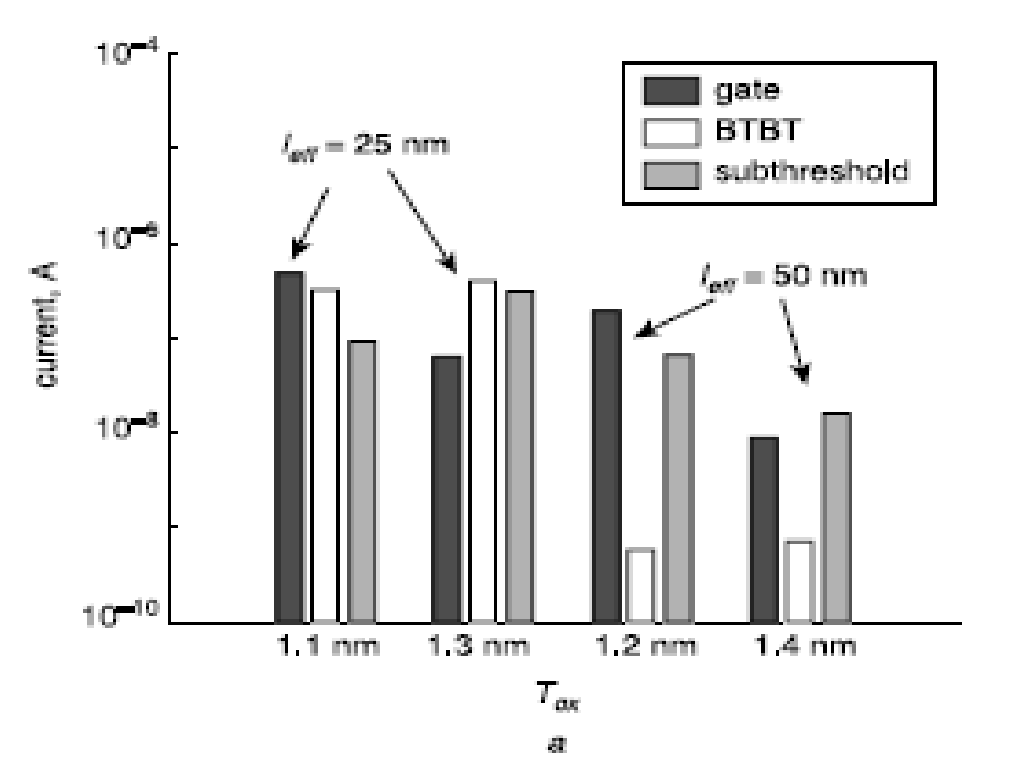
Factors affecting leakage power

- Temperature
  - Sub-threshold current increases exponentially
    - Reduction in V_{TH}
    - Increase in thermal voltage
  - BTBT increases due to band gap narrowing
  - Gate leakage is insensitive to temperature change



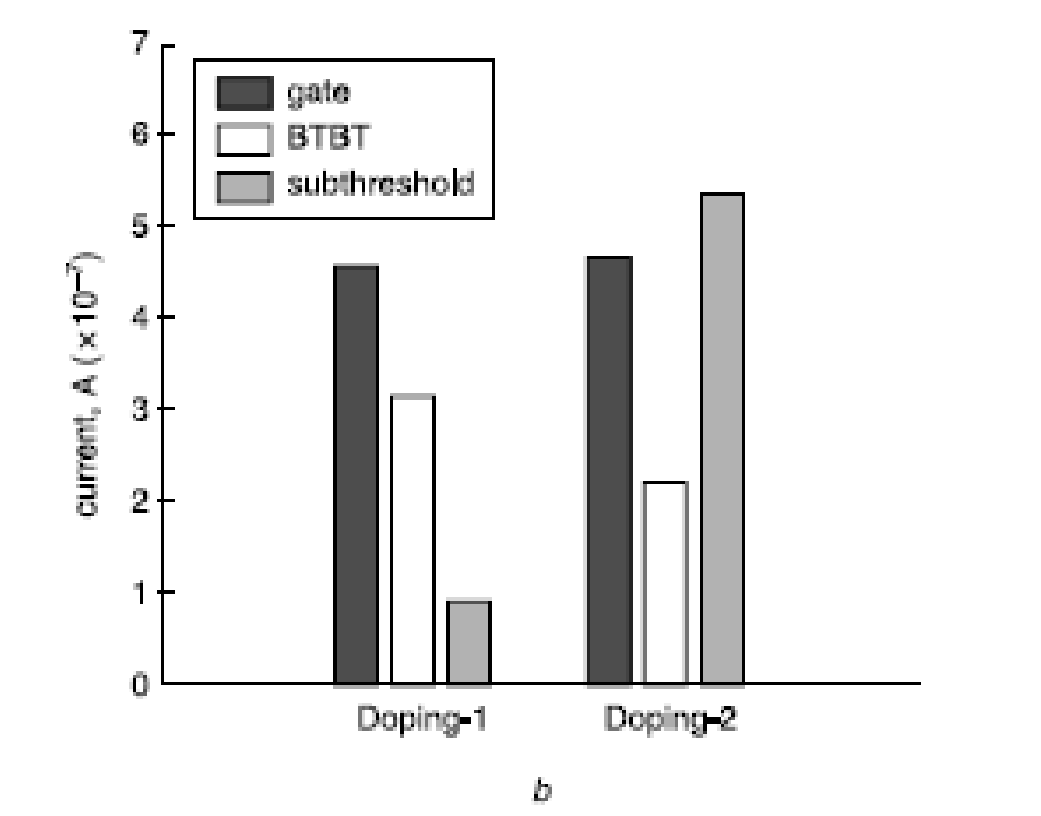
Factors affecting leakage power

- Gate oxide thickness
  - Sub-threshold current decreases in long channel transistors and increases in short channel devices
  - BTBT (Band-To-Band Tunneling) is insensitive
  - Gate leakage increases as thickness reduces



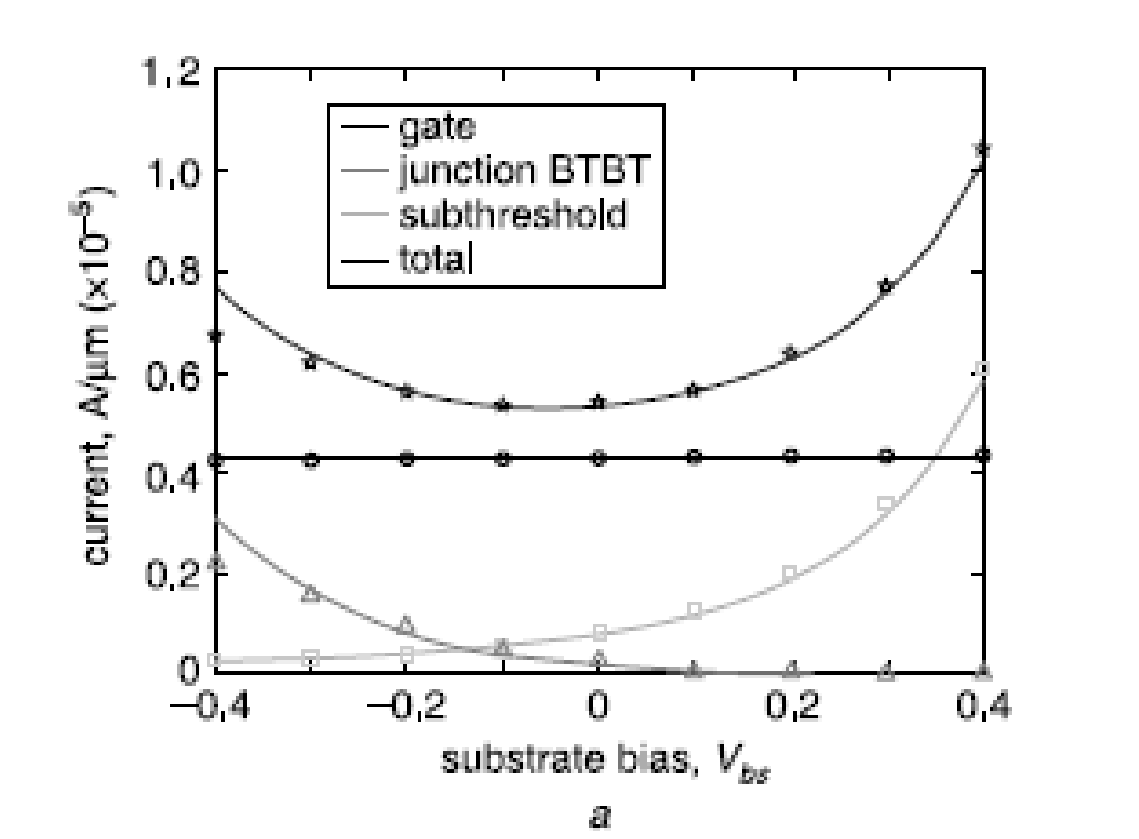
Factors affecting leakage power

- Doping Profile
  - Sub-threshold leakage reduces with halo doping
  - BTBT increases
  - Gate leakage insensitive



Factors affecting leakage power

- Reverse substrate bias
  - Sub-threshold current reduces since Vt increases
  - BTBT increases exponentially
  - Gate leakage almost insensitive



Problem: A Design Example

- A battery-operated 65nm digital CMOS device is found to consume equal amounts of dynamic power and leakage power (P), while the short-circuit power is negligible. The energy consumed by a computing task, that takes T seconds, is $2PT$

- Compare the following two power reduction strategies for extending the battery life:
  A. Clock frequency is reduced to half, keeping all other parameters constant
  B. Supply voltage is reduced to half. This slows the gates down and forces the clock frequency to be lowered to half of its original (full voltage) value. Assume that leakage current is held unchanged by modifying the design of transistors

A. Clock Frequency Reduction

- Reducing the clock frequency will reduce dynamic power to $P / 2$
- Static power consumption remains at P
- But the execution time of the task doubles
- Energy consumption for the task will be,

  Energy = $(P / 2 + P)\ 2T = 3PT$

  which is greater than the original 2PT.

B. Supply Voltage Reduction

- When the supply voltage and clock frequency are reduced to half their values, dynamic power is reduced to $P / 8$
- Static power will be $P / 2$
- The time of task is doubled
- The total energy consumption is,

  Energy = $(P / 8 + P / 2)\ 2T = 5PT / 4 = 1.25PT$
- *The voltage reduction strategy reduces energy consumption while a simple frequency reduction consumes more energy*

Summary: Leakage Power

- Leakage power as a fraction of the total power increases as clock frequency drops. *Turning supply off in unused parts can save power*
- For a gate it is a small fraction of the total power; it can be significant for very large circuits
- Scaling down features requires lowering the threshold voltage, which increases leakage power; roughly doubles with each shrinking
- Multiple-threshold devices are used to reduce leakage power