## Analog Integrated Systems

3. ULP Digital Circuit Design

#1/4

1. Conventional CMOS Digital Design
   - Combinational Logic
   - Sequential Logic

2. Technology Driven:
   - (1) Reducing $V_{DD}$ (2) dual-voltage CPUs (3) dynamic voltage scaling (4) undervolting
   - (1) Reducing Frequency (2) underclocking (3) dynamic frequency scaling
   - (1) Reducing Capacitance (2) reduce transistor size (3) add functionality in IC (replace PCB) (4) low-k dielectric (5) Dual-Vt

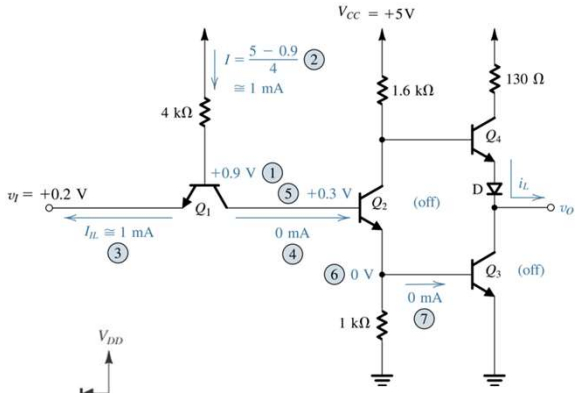
3. Design Driven:
   - (1) Lower logic Swing (2) Reduce the switching activity (3) Optimizing machine code
   - (1) clock gating (2) Power gating (3) globally asynchronous locally synchronous
   - Recycling energy stored in the capacitors (1) adiabatic circuit (2) energy recovery logic



1

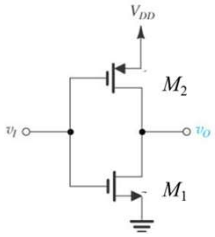
## Evolution: Bipolar→CMOS

Bipolar

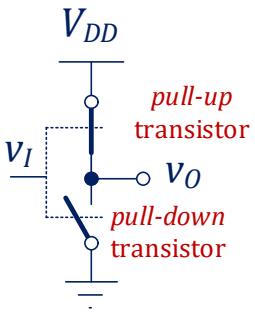



**TTL logic:**
**~10 mW ~10 ns /gate.**

CMOS




**CMOS logic (28nm):**
**~10 fJ ~10 ps**
**/gate /transaction**






2

## Static parameters

- $V_{OH}$ → state 1 voltage
- $V_{OL}$ → state 0 voltage
- $V_{IH}$ → Minimum input voltage interpreted as 1
- $V_{IL}$ → Maximum input voltage interpreted as 0

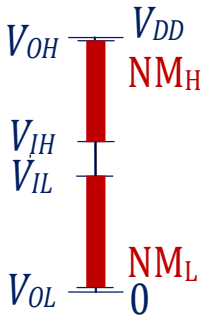
- Noise Margins
  - $NM_H = V_{OH} - V_{IH}$
  - $NM_L = V_{IL} - V_{OL}$

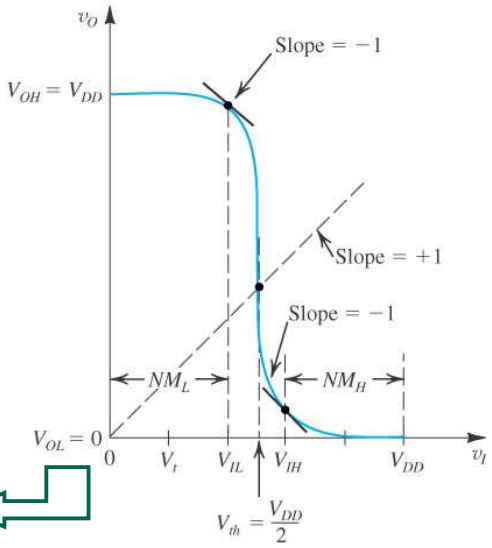
(note: $V_{OH} \leq V_{DD}$ e $V_{OL} \geq 0$)

Static power $P_D = 1/2(P_{DH} + P_{DL})$

for CMOS Inverter

$P_D$~0; $V_{OH} = V_{DD}$; $V_{OL} = 0$






## Power Consumption CMOS inverter

**Static Power** → average from states 0 and 1
CMOS state 1 → $P_{DH} = 0$; state 0 → $P_{DL} = 0$.

**Dynamic Power**

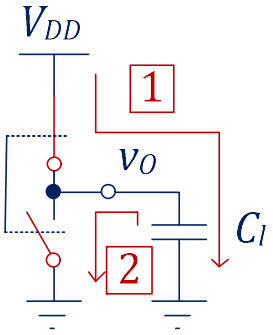
[1] $v_I: 1 \to 0 \Rightarrow v_O: 0 \to 1$

$$W_B = \underbrace{QV_{DD}}_{\text{battery}} = C_l V_{DD}^2 = \underbrace{\frac{1}{2}C_l V_{DD}^2}_{\text{stored in } C_l} + \underbrace{\frac{1}{2}C_l V_{DD}^2}_{\text{dissipated}}$$

[2] $v_I: 0 \to 1 \Rightarrow v_O: 1 \to 0$

$$W_B = \underbrace{\frac{1}{2}C_l V_{DD}^2}_{\text{dissipated}}$$

$$\underbrace{W_B = \frac{1}{2}C_l V_{DD}^2 + \frac{1}{2}C_l V_{DD}^2 = C_l V_{DD}^2}_{\text{dissipated in one period}}$$

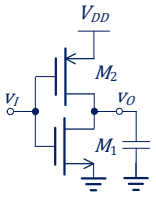



**Dynamic power** → $\boxed{P = f\ C_l\ V_{DD}^2}$

## Dynamic parameters

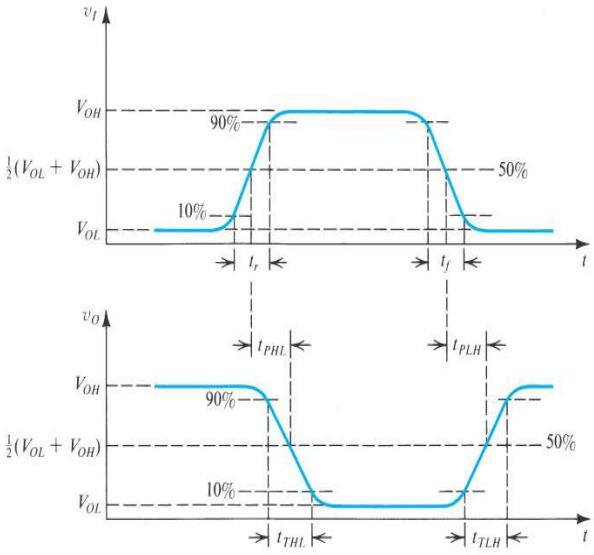
$t_r \rightarrow$ rise time
$t_f \rightarrow fall$ time
(measured 10−90% from $V_{OH}$ to $V_{OL}$)

$t_{PHL} \rightarrow$ delay time from "High" to "Low"
$t_{PLH} \rightarrow$ delay time from"Low" to "High"
(measured from 50% of $v_I$ to $v_O$)
$t_P = \frac{1}{2}(t_{PHL} + t_{PLH}) \rightarrow$ delay time



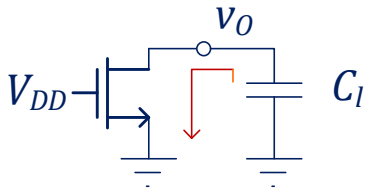

$$i_C = C\frac{dv_C}{dt}$$

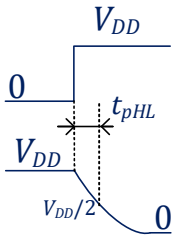
$$I_C = C\frac{\Delta v_C}{\Delta t}$$

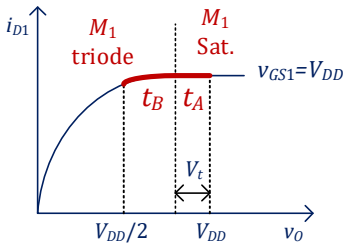





## Delay: simple model



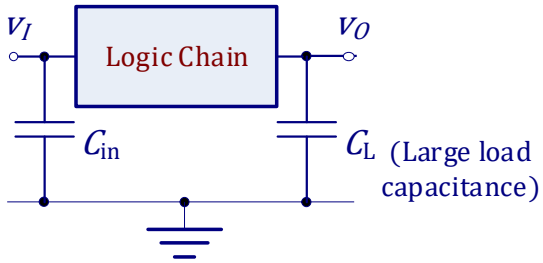



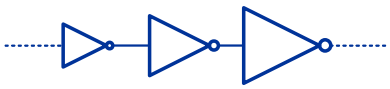

$t_{pHL} = t_{pLH} = t_p(M_1, M_2$ **adapted, $\lambda$=0**$)$

$(i_{D1})_{av} t_{pHL} = C_l \frac{1}{2}(V_{OH} - V_{OL})$; $\quad (i_{D1})_{av} = \frac{1}{2}\left[(i_{D1})_{o=V_{DD}} + (i_{D1})_{v_O=V_{DD}/2}\right]$

## Logic Chain

- Problem: Drive of large capacitances (e.g. Output stages)
  - PCB/package capacitance >> IC internal capacitances.
  - Reduction of propagation times... logical circuit chain sizing

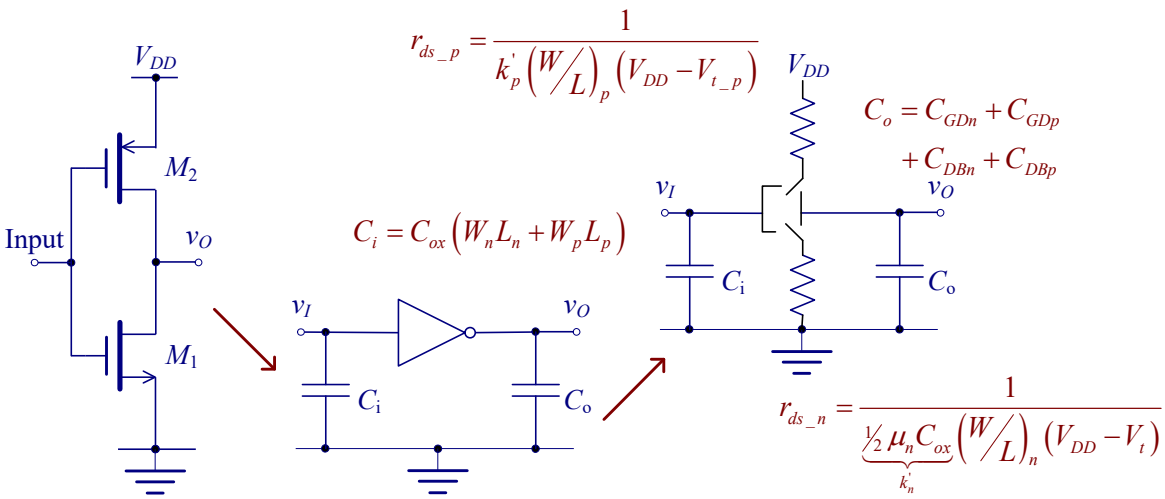




## 1st order model

- Models (1st order approach appropriate to the problem)
-



## Inverters Logic chain analysis #1

- $t_d \approx RC_o$ assuming $R=r_{ds\_p}=r_{ds\_n}$
- $R \propto W$: $W\uparrow \Rightarrow R\downarrow, C_i\uparrow, C_o\uparrow$

**Scaling factor:** $S_j$ with $S_1 = 1$ and $S_j > 1$ for $j > 1$

**1st stage:** $\beta_1 = k'\left(\frac{W}{L}\right)_1$ $\quad$ **$j-$stage:** $\beta_j = S_j\beta_1$

$S_jC_o$, **Output capacitance stage $j$**

$S_{j+1}C_i$, **Input capacitance stage $j+1$**

$S_{j+1}C_w$, **Connection capacitance stage $j+1$**

**Stage $j$ delay:** $t_{d,j} = \left(\frac{R}{S_j}\right)\left[S_jC_o + S_{j+1}(C_i + C_w)\right]$



## Inverters Logic chain analysis #2

Total delay: $t_d = \sum_{j=1}^{N} \frac{R\left(S_jC_o + S_{j+1}(C_i + C_w)\right)}{S_j}$

search minimum: $\frac{\partial t_d}{\partial S_j} = 0 \rightarrow$ recursively $\frac{S_{j+1}}{S_j} = \frac{S_j}{S_{j-1}} = K = \text{constant}$

boundary conditions: $S_1 = 1$ and $S_{N+1} = \frac{C_L}{C_i}$

$\frac{S_2}{S_1}\frac{S_3}{S_2}\frac{S_4}{S_3}\cdots\frac{S_{N+1}}{S_N} = K^N = \frac{C_L}{C_i}$

$\boldsymbol{K = \left(\frac{C_L}{C_i}\right)^{1/N}} \Rightarrow t_{d,min} = \sum_{j=1}^{N} R\left(C_o + K(C_i + C_w)\right) = NR\left(C_o + K(C_i + C_w)\right)$

**Optimal number of stages:**

$\frac{\partial t_{d,\min}}{\partial N} = 0$ (if $C_0$ negligible) $\Rightarrow \boldsymbol{N = \ln\left(\frac{C_L}{C_i}\right)}$ (nearest integer)

## Inverters Logic chain analysis #3

- To finish scaling we need to calculate the normalization value.
- It is done by scaling the top floor (e.g. off-chip driver) and then scaling the previous floors.
- Use the expressions that relate the rise and fall time of the inverter depending on sizing and output capacitance

$$\left(\frac{W}{L}\right)_n = \frac{C_{out}}{\tau_n k'_n (V_{DD} - V_{tn})}$$

$$\left(\frac{W}{L}\right)_p = \frac{C_{out}}{\tau_p k'_p (V_{DD} - V_{t\_p})}$$

$$t_{HL} = \tau_n \left[\frac{2V_{tn}}{V_{DD} - V_{tn}} + \ln\left(\frac{2(V_{DD} - V_{tn})}{V_o} - 1\right)\right]$$

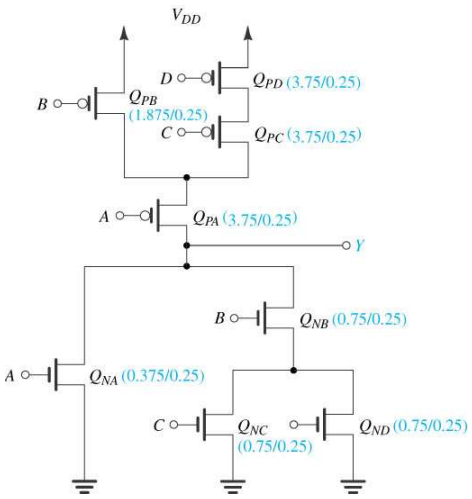
$$t_{LH} = \tau_p \left[\frac{2V_{tp}}{V_{DD} - V_{tp}} + \ln\left(\frac{2(V_{DD} - V_{tp})}{V_o} - 1\right)\right]$$

Note: $V_o$ corresponds to 10% of $V_{DD}$

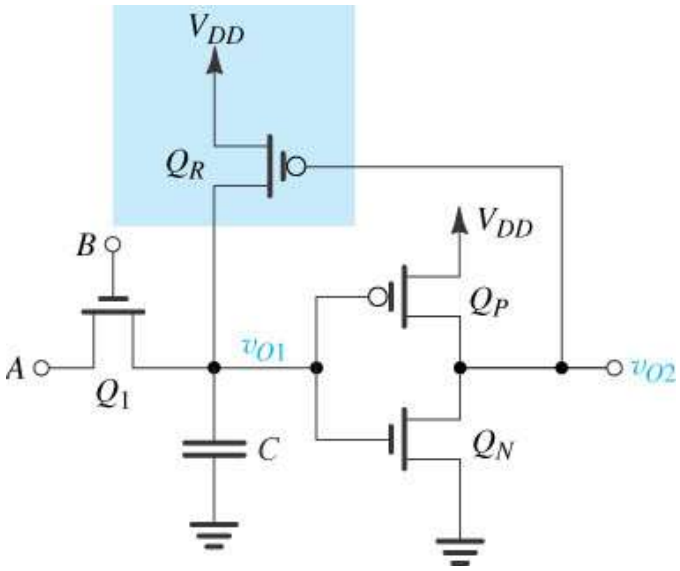


## Custom logic functions with CMOS combinational circuits

- Adapted transistors
  - *L*=0.25um
  - *Wn*=0.375um; *Wp*=1.25um
- $\bar{y} = (C + D) \cdot B + A$



## Domino Logic




$v_A = V_{DD} \rightarrow v_{O1} = V_{DD} - V_t \rightarrow$
$\rightarrow Q_P$ **On** $\rightarrow P_d$ **Static** $\neq 0$

**Add $Q_R$ to restore $v_{O1} = V_{DD}$ if possible zero $-$ $V_t$ transistors $V_t \approx 0$**



## Analog Integrated Systems

3. ULP Digital Circuit Design

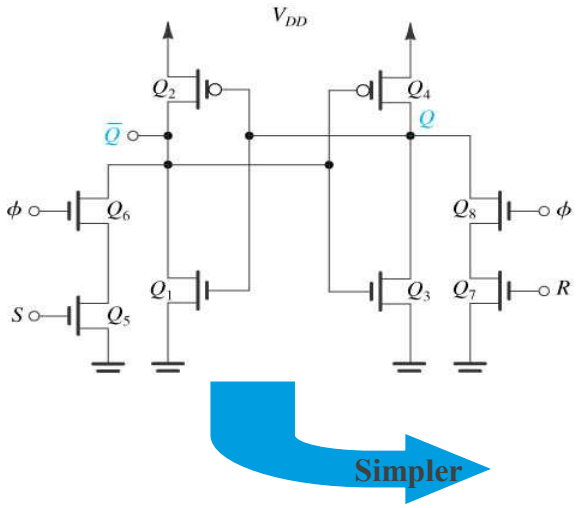
#2/4

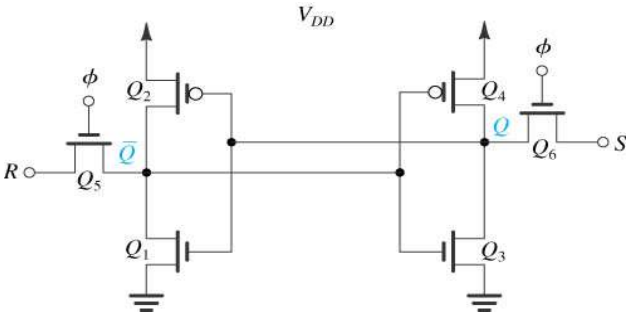
1. Conventional CMOS Digital Design
   - Combinational Logic
   - Sequential Logic

2. Technology Driven:
   - (1) Reducing $V_{DD}$ (2) dual-voltage CPUs (3) dynamic voltage scaling (4) undervolting
   - (1) Reducing Frequency (2) underclocking (3) dynamic frequency scaling
   - (1) Reducing Capacitance (2) reduce transistor size (3) add functionality in IC (replace PCB) (4) low-k dielectric (5) Dual-Vt

3. Design Driven:
   - (1) Lower logic Swing (2) Reduce the switching activity (3) Optimizing machine code
   - (1) clock gating (2) Power gating (3) globally asynchronous locally synchronous
   - Recycling energy stored in the capacitors (1) adiabatic circuit (2) energy recovery logic

## Gated SR Latch



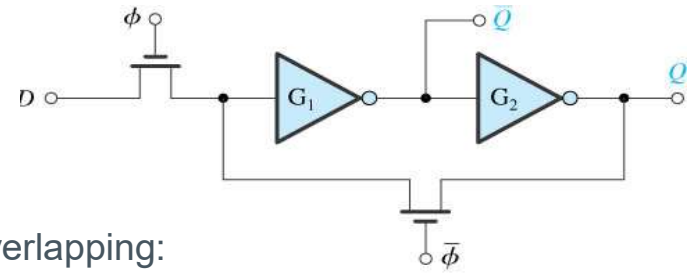



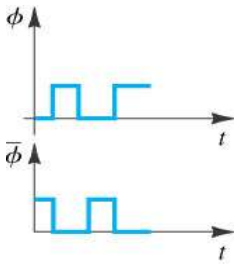

- Used in SRAM



## Latch D

- Inverters and Switches

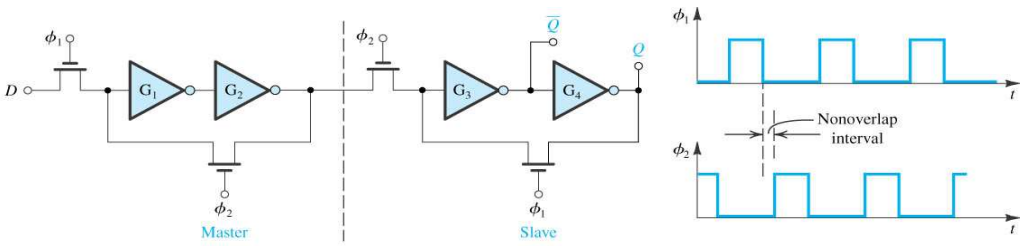



- $\phi$ non-overlapping:
- $\phi$=1 -> Q=D
- Transition: both switches open: Q value remains due to parasitic capabilities
- $\phi$=0 -> Q stored in the latch “edge-triggered”

## Latch D

- “master-slave” configuration

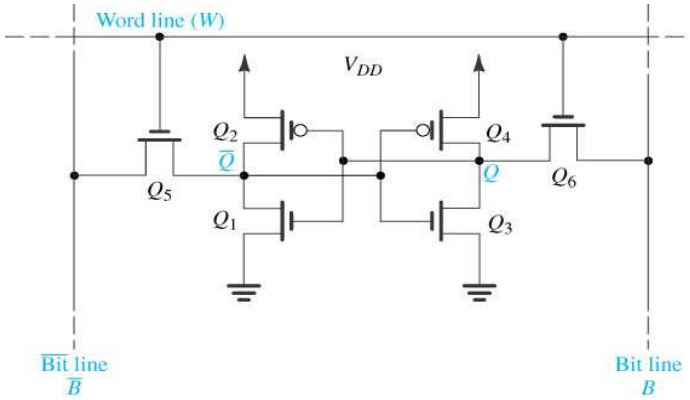



- D on the fall flank $\phi_1$ is stored in the "master"
- On the rise flank of $\phi_2$ passes to the "slave"
- stays in Q for a clock period
- $\phi_1$=$\phi_2$=0 should be short or parasitic capacities discharge



## Static RAM (SRAM)

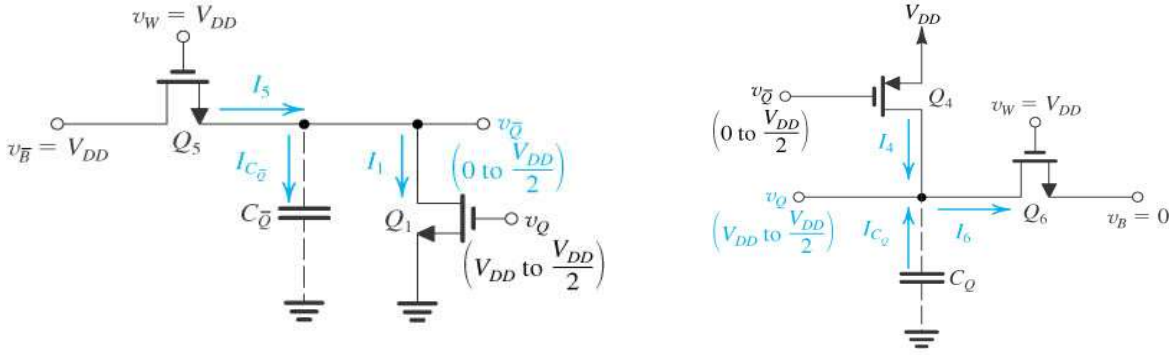
- Q5 and Q6 access transistors: controled by the word line selector, connect the cell to the data lines




- CMOS latch
- Two complementary data lines, B and ~B.

## Static RAM (SRAM): write

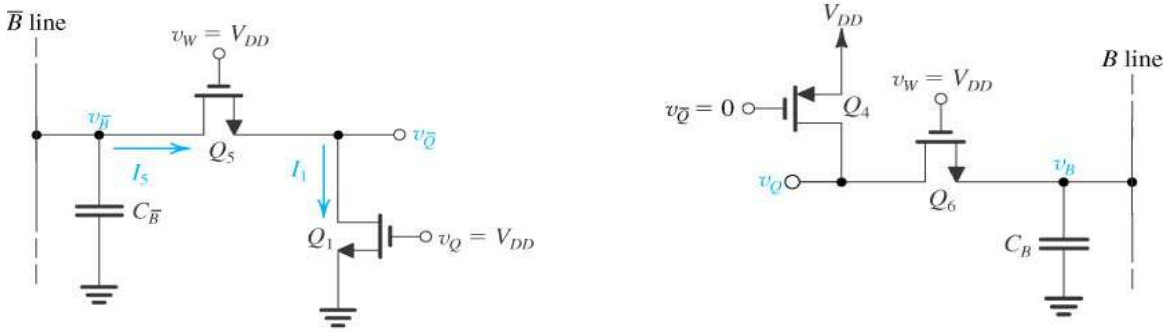
- B e ~B imposed
- Rise and fall times can be calculated similarly to pass-through transistors with capacitive load.





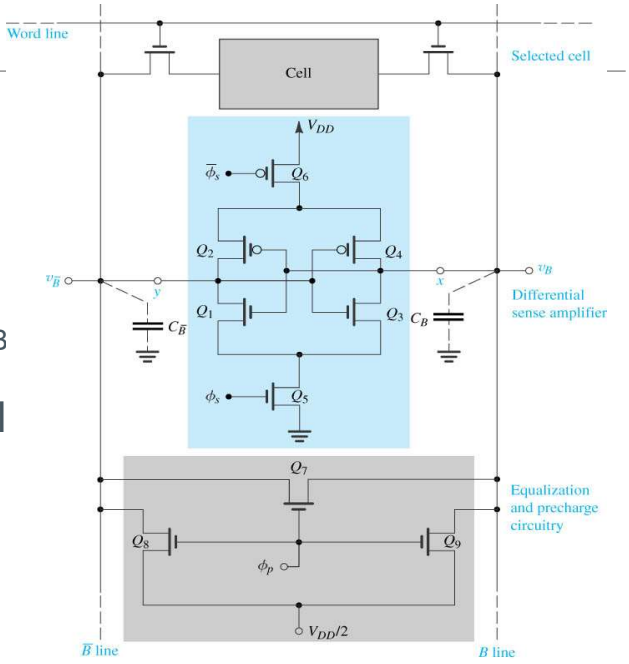

## Static RAM (SRAM): read

- Preload of data lines (latch transition voltage)
- Line enable: |vB-v~B|>0.2V
- Reading amplifier restores levels $V_{DD}$, 0



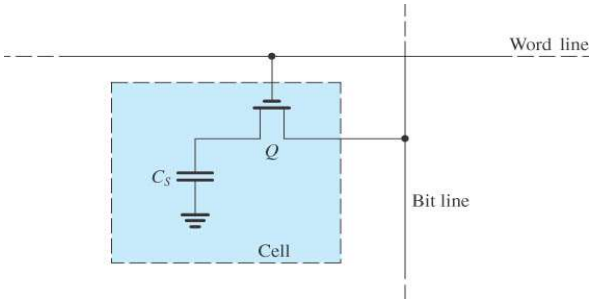
## Static RAM (SRAM): read amplifier

- CMOS latch activated by $ø_S$
- Read operation
- Preload+equalization of data line ($V_{DD}/2$) on $ø_P$
- Cell connected to data line: $|V_B V_{\sim B}|$~30 to 500mV, polarity depends on the state of the cel
- Read amplifier activated: restores levels $V_{DD}$,0





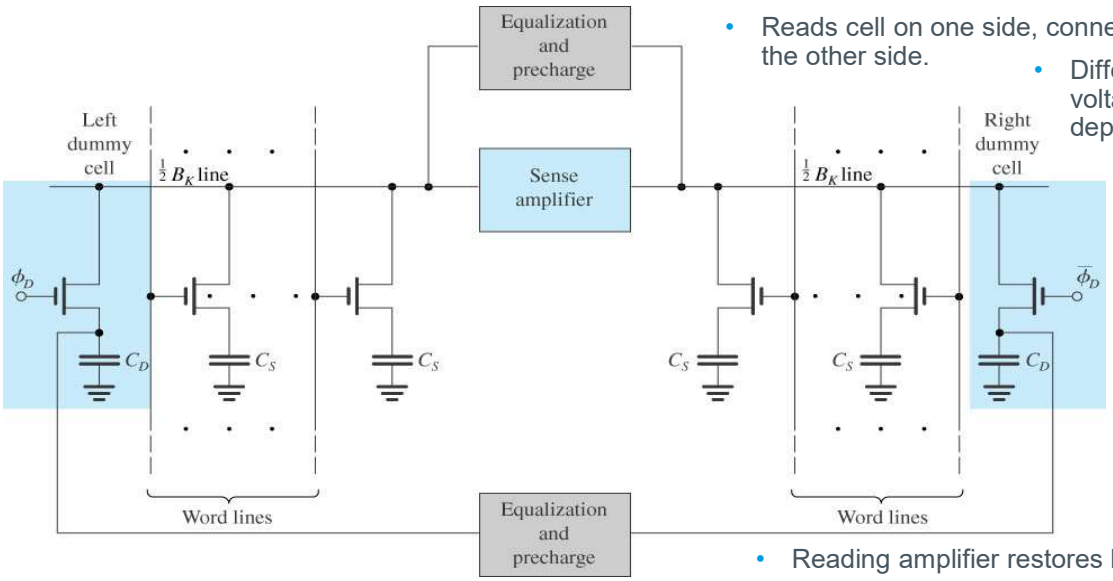

## Dynamic RAM (DRAM)




- Capacitor + access transistor
- DRAM has greater density (4x) than SRAM
- Periodic refresh (requires clock) of cells with logical 1

- 1 data line: $C_{BitLine}$ (~1pF)>>$C_S$(<50fF)
- Read: Δv tens of mV, amplifier reading reestablishes levels $V_{DD}$, 0
- Writing: $C_S$ gets 0 or $V_{DD}$ -$V_t$ (NMOS switch)
- Restore: Enabling rows successively, activated row cells are restored.

## Dynamic RAM (DRAM): Amplifier

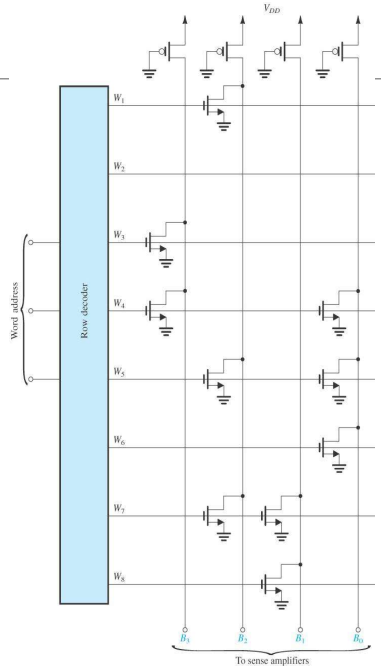



- Reads cell on one side, connected cell on the other side.
  - Differential voltage: polarity depends on state
- Reading amplifier restores levels $V_{DD}$, 0



## Read Only Memory (ROM)

- ROM with 8 words of 4-bit
- Combinational logic system
  - entries=addresses; outputs=contents
- Intersection Lines/Columns
  - Transistor if bit=0, nothing if it is 1
- Pull-up: any load as on NMOS logic
- Mask programming (metallization): Connected only the desired transistors



## 2D to 3D memories…

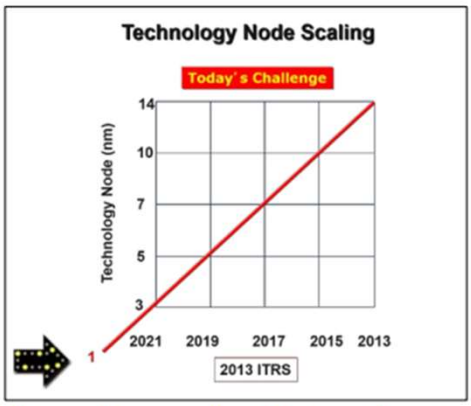



*Figure ES43* *2D scaling will reach fundamental limits beyond 2020*

IEEE IRDS INTERNATIONAL ROADMAP FOR DEVICES AND SYSTEMS™ 2020 EDITION EXECUTIVE SUMMARY

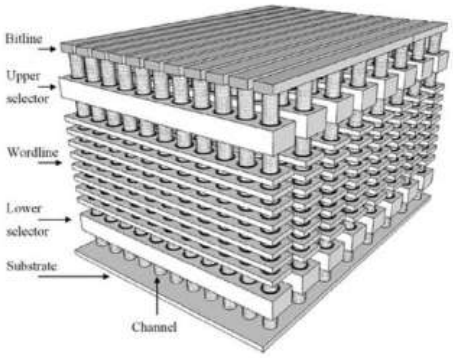



*Figure ES44* *Flash memory aggressively adopts 3D scaling in 2014*



## Analog Integrated Systems

3. ULP Digital Circuit Design

#3/4

1. Conventional CMOS Digital Design
   - Combinational Logic
   - Sequential Logic

2. Technology Driven:
   - (1) Reducing $V_{DD}$ (2) dual-voltage CPUs (3) dynamic voltage scaling (4) undervolting
   - (1) Reducing Frequency (2) underclocking (3) dynamic frequency scaling
   - (1) Reducing Capacitance (2) reduce transistor size (3) add functionality in IC (replace PCB) (4) low-k dielectric (5) Dual-Vt

3. Design Driven:
   - (1) Lower logic Swing (2) Reduce the switching activity (3) Optimizing machine code
   - (1) clock gating (2) Power gating (3) globally asynchronous locally synchronous
   - Recycling energy stored in the capacitors (1) adiabatic circuit (2) energy recovery logic

## Where to save power?

- **Total Power →** $\boxed{P_{dyn} + P_{sc} + P_{leak}}$
  - **Dynamic power →** $\boxed{P_{dyn} = f\ C_l\ V_{DD}^2}$
  - **Short circuit power →** $\boxed{P_{sc}}$
  - **Leakage power →** $\boxed{P_{leak}}$



## Power consumption reduction techniques

$$P = f\ C_l\ V_{DD}^2$$

Technology Driven:

- (1) Reducing $V_{DD}$ (2) dual-voltage CPUs (3) dynamic voltage scaling (4) undervolting
- (1) Reducing Frequency (2) underclocking (3) dynamic frequency scaling
- (1) Reducing Capacitance (2) reduce transistor size (3) add functionality in IC (replace PCB) (4) low-k dielectric (5) Dual-Vt

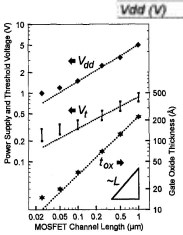
Design Driven:

- (1) Lower logic Swing (2) Reduce the switching activity (3) Optimizing machine code
- (1) Clock gating (2) Power gating (3) Globally asynchronous locally synchronous
- Recycling energy stored in the capacitors (1) Adiabatic circuit (2) Energy recovery logic

## Technology Driven ($V_{DD}$): (1) Reducing $V_{DD}$

*Table ES2 Overall Roadmap Technology Characteristics*

| 2020 IRDS Executive Summary Drivers-ORTC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *YEAR OF PRODUCTION* | *2019* | *2020* | *2022* | *2025* | *2028* | *2031* | *2032* | *2034* |
| *Logic device technology naming [4] NEW node definition* | G54M38 | G48M36 | G45M24 | G45M20 | G40M16 | G38M16T2 | G38M16T3 | G38M16T4 |
| *Logic industry "Node Range" Labeling (nm)* | "7" | "5" | "3" | "2.1" | "1.5" | "1.0nm-eq" | "1.0nm-eq" | "0.7nm-eq" |
| *Logic device structure options* | FinFET | FinFET | FinFET LGAA | LGAA | LGAA VGAA | LGAA-3D VGAA | LGAA-3D VGAA | LGAA-3D VGAA |
| LOGIC CELL AND FUNCTIONAL FABRIC TARGETS | | | | | | | | |
| *Average Cell Width Scaling Factor Multiplier* | 1 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 |
| LOGIC DEVICE GROUND RULES | | | | | | | | |
| *MPU/SoC M0 1/2 Pitch (nm) [1,2]* | 18 | 15 | 12 | 10.5 | 8 | 8 | 8 | 8 |
| *Physical Gate Length for HP Logic (nm) [3]* | 20 | 18 | 16 | 14 | 12 | 12 | 12 | 12 |
| *Lateral GAA (nanosheet) Minimum Thickness (nm)* | | | | 7 | 6 | 5 | 5 | 5 |
| *Minimum Device Width (FinFET fin, nanosheet, SRAM) or Diameter (nm)* | 9 | 7 | 6 | 7 | 6 | 6 | 6 | 6 |
| LOGIC DEVICE Electrical | | | | | | | | |
| *Vdd (V)* | 0.75 | 0.7 | 0.7 | 0.65 | 0.65 | 0.6 | 0.6 | 0.6 |




Dynamic power → $P = f\ C_l\ V_{DD}^2$

IEEE IRDS International Roadmap for Devices and Systems™ 2020 Edition Executive Summary



## Technology Driven ($V_{DD}$): (2) dual-voltage CPUs

Dynamic power → $P = f\ C_l\ V_{DD}^2$

- **Block-level supply assignment**
  - Higher-performance functions use higher $V_{DD}$
  - Lower-performance functions use lower $V_{DD}$
  - Level conversion at block boundaries
- **Multiple supplies inside a block**
  - Non-critical paths use lower $V_{DD}$
  - Level conversion within the block

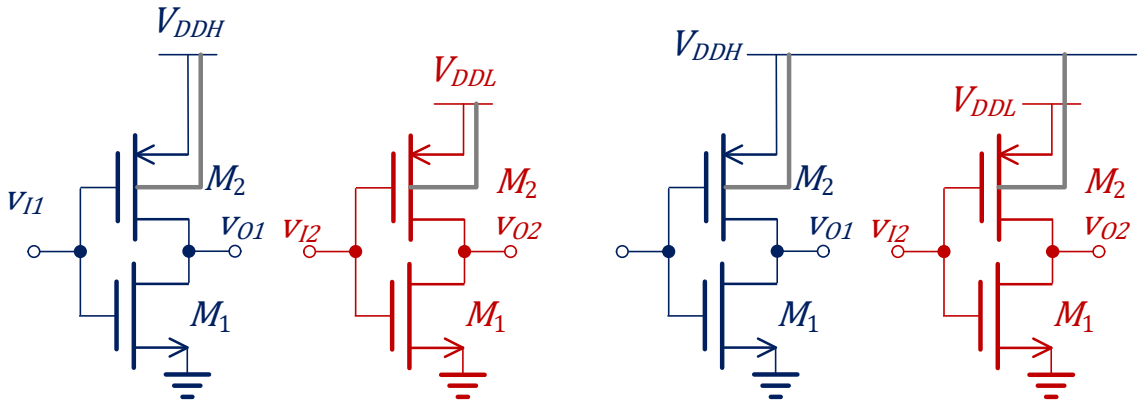
[Rabey, Low Power Design Essentials, Springer 2009]

- Two supply voltages per block are optimal
- Optimal ratio between the supply voltages is 0.7
- Best level conversion is performed by a level-converting flip-flop (LCFF)

## Technology Driven ($V_{DD}$): (2) dual-voltage CPUs

Dynamic power → $P = f\ C_l\ V_{DD}^2$

- Conventional with separate wells
- Only one shared well; less area

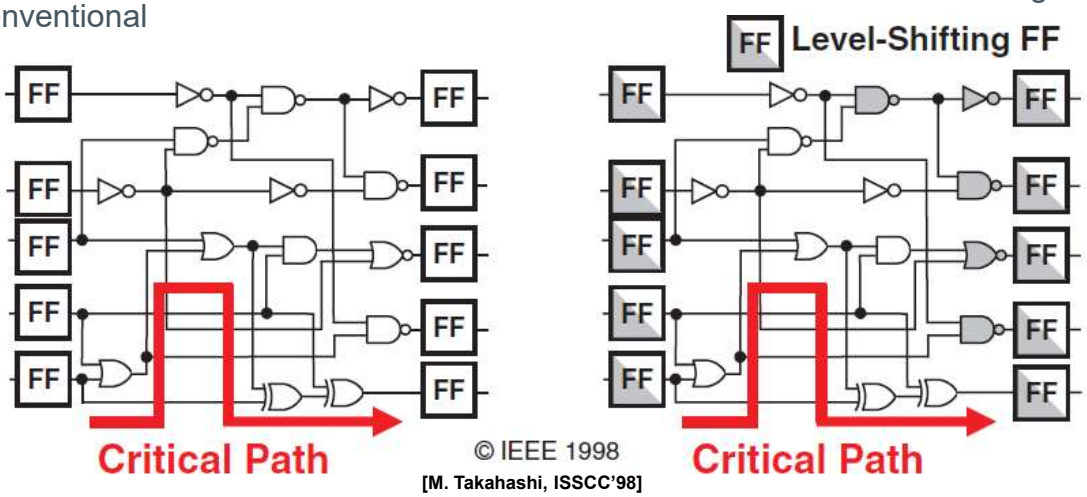





## Technology Driven ($V_{DD}$): (2) dual-voltage CPUs

Dynamic power → $P = f\ C_l\ V_{DD}^2$

- Conventional
- "Clustered voltage scaling"





[M. Takahashi, ISSCC'98]

## Technology Driven ($V_{DD}$): (2) dual-voltage CPUs

IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS, VOL. 12, NO. 2, FEBRUARY 2004 185

Level Conversion for Dual-Supply Systems

Fujio Ishihara, Farhana Sheikh, *Member, IEEE*, and Borivoje Nikolić, *Member, IEEE*

- Level-Converting Flip-Flops

IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS, VOL. 12, NO. 2, FEBRUARY 2004

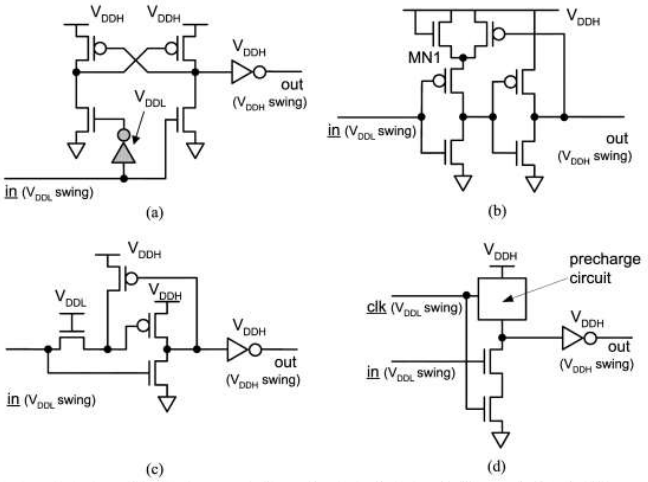



Fig. 1. Basic level converter structures. A shaded gate represents a $V_{DDL}$ gate and underlined nodes show $V_{DDL}$-swing signals. (a) Cross-coupled pMOS pair (CCLC) [11]. (b) Single-supply diode-voltage-limited buffer (SSLC) [12]. (c) Pass-transistor half latch. (d) Precharged circuit.



## Technology Driven: (3) dynamic voltage scaling and/or (4) undervolting

Dynamic power → $\boxed{P = f\ C_l\ V_{DD}^2}$

**Dynamic voltage scaling:** power management technique in computers, to increase or decrease the voltage in a component, depending on load.

It can be Overvolting or **Undervolting.**

**Undervolting:** lower voltage supplying the circuit core.

- **reduce power consumption**
- **increase battery life**
- **reduce heat emission**

## Technology Driven (*f*): (1) Reducing Frequency (2) underclocking (3) dynamic frequency scaling

Dynamic power → $P = f\ C_l\ V_{DD}^2$

**(1) Reduce Frequency:** work at lower frequency

**(2) Underclocking:** modify a computer or electronic circuits' time settings to run at a lower clock rate than specified.

**(3) Dynamic frequency scaling:** microprocessor frequency is automatically adjusted "on the fly" depending on the actual load. Can be combined as dynamic voltage and frequency scaling (DVFS).

- **reduce power consumption**
- **increase battery life**
- **reduce heat emission**



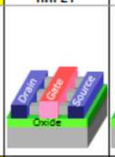
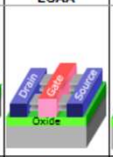
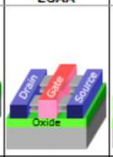
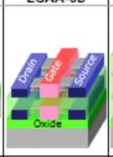
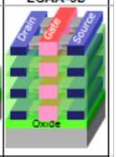
## Technology Driven: (1) Reducing Capacitance (2) Reduce transistor size

Dynamic power → $P = f\ C_l\ V_{DD}^2$

| YEAR OF PRODUCTION | 2019 | 2020 | 2022 | 2025 | 2028 | 2031 | 2032 | 2034 |
|---|---|---|---|---|---|---|---|---|
| Logic device technology naming [4] NEW node definition | G54M38 | G48M36 | G45M24 | G45M20 | G40M16 | G38M16T2 | G38M16T3 | G38M16T4 |
| Logic industry "Node Range" Labeling (nm) | "7" | "5" | "3" | "2.1" | "1.5" | "1.0nm-eq" | "1.0nm-eq" | "0.7nm-eq" |
| Logic device structure options | FinFET | FinFET | FinFET LGAA | LGAA | LGAA VGAA | LGAA-3D VGAA | LGAA-3D VGAA | LGAA-3D VGAA |
| LOGIC CELL AND FUNCTIONAL FABRIC TARGETS | | | | | | | | |
| Average Cell Width Scaling Factor Multiplier | 1 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 |
| LOGIC DEVICE GROUND RULES | | | | | | | | |
| MPU/SoC M0 1/2 Pitch (nm) [1,2] | 18 | 15 | 12 | 10.5 | 8 | 8 | 8 | 8 |
| Physical Gate Length for HP Logic (nm) [3] | 20 | 18 | 16 | 14 | 12 | 12 | 12 | 12 |
| Lateral GAA (nanosheet) Minimum Thickness (nm) | | | | 7 | 6 | 5 | 5 | 5 |
| Minimum Device Width (FinFET fin, nanosheet, SRAM) or Diameter (nm) | 9 | 7 | 6 | 7 | 6 | 6 | 6 | 6 |
| LOGIC DEVICE Electrical | | | | | | | | |
| Vdd (V) | 0.75 | 0.7 | 0.7 | 0.65 | 0.65 | 0.6 | 0.6 | 0.6 |
| DRAM TECHNOLOGY | | | | | | | | |
| DRAM Min half pitch (nm) [1] | 18 | 17.5 | 17 | 14 | 11 | 8.4 | 8.4 | 7.7 |
| DRAM Min Half Pitch (Calculated Half pitch) (nm) [1] | 20.5 | 17.5 | 18.5 | 15 | 12 | 10 | 10 | 8.5 |
| DRAM Cell Size Factor: aF^2 [4] | 6 | 6 | 4 | 4 | 4 | 4 | 4 | 4 |
| DRAM Gb/1chip target | 8 | 8 | 16 | 16 | 32 | 32 | 32 | 32 |
| NAND Flash | | | | | | | | |
| Flash 2D NAND Flash uncontacted poly 1/2 pitch – F (nm) 2D [1][2] | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 |
| Flash Product highest density (independent of 2D or 3D) | 512G | 1T | 1T | 1.5T | 3T | 4T | 4T | 4T+ |
| Flash Product Maximum bit/cell (2D_3D) [6] | 2_4 | 2_4 | 2_4 | 2_4 | 2_4 | 2_4 | 2_4 | 2_4 |
| Flash 3D NAND Maximum Number of Memory Layers [6] | 48-65 | 64-96 | 96-128 | 128-192 | 256-384 | 384-512 | 384-512 | 512+ |

## Technology Driven: (3) add functionality in IC (replace PCB)

Dynamic power → $P = f\ C_l\ V_{DD}^2$

| YEAR OF PRODUCTION | 2020 | 2022 | 2025 | 2028 | 2031 | 2034 |
|---|---|---|---|---|---|---|
| | G48M36 | G45M24 | G42M20 | G40M16 | G38M16T2 | G38M16T4 |
| Logic industry "Node Range" Labeling (nm) | "5" | "3" | "2.1" | "1.5" | "1.0 eq" | "0.7 eq" |
| IDM-Foundry node labeling | i7-f5 | i5-f3 | i3-f2.1 | i2.1-f1.5 | i1.5e-f1.0e | i1.0e-f0.7e |
| Logic device structure options | FinFET | finFET LGAA | LGAA | LGAA | LGAA-3D | LGAA-3D |
| Mainstream device for logic | finFET | finFET | LGAA | LGAA | LGAA-3D | LGAA-3D |
| |  |  |  |  |  |  |
| Vdd (V) | 0.70 | 0.70 | 0.65 | 0.65 | 0.60 | 0.60 |
| Gate length (nm) | 18 | 16 | 14 | 12 | 12 | 12 |
| Number of stacked tiers | 1 | 1 | 1 | 1 | 2 | 4 |
| Number of stacked devices | 1 | 1 | 3 | 3 | 4 | 4 |
| Digital block area scaling - node-to-node | - | 0.75 | 0.78 | 0.79 | 0.52 | 0.50 |
| Cell height limitation - HD | M0 | M0 | M0 | M0 | M0 | M0 |
| SoC area scalling (stacked) - node-to-node | - | 0.78 | 0.82 | 0.83 | 0.56 | 0.59 |
| CPU frequency (GHz) | 3.13 | 3.27 | 3.51 | 3.47 | 3.25 | 2.93 |
| Frequency scaling - node-to-node | - | 0.04 | 0.08 | -0.01 | -0.07 | -0.10 |
| CPU frequency at constant power density (GHz) | 3.13 | 2.58 | 2.92 | 2.19 | 1.29 | 0.78 |
| Power at iso frequency - node-to-node | - | -0.08 | -0.26 | -0.05 | -0.07 | -0.09 |
| Power density - relative | 1.00 | 1.27 | 1.20 | 1.58 | 2.52 | 3.73 |



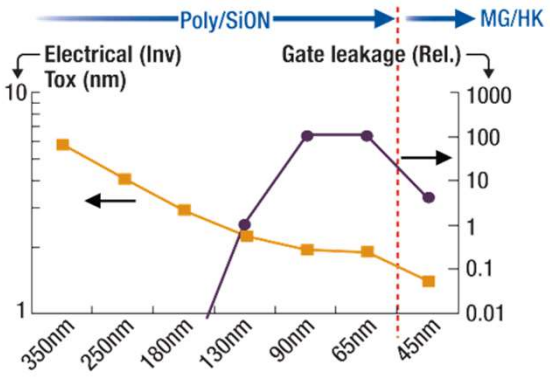
## Technology Driven: (3) add functionality in IC (replace PCB)

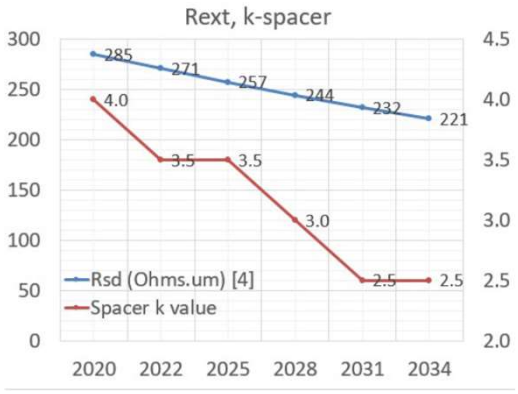
Dynamic power → $P = f\ C_l\ V_{DD}^2$

| YEAR OF PRODUCTION | 2020 | 2022 | 2025 | 2028 | 2031 | 2034 |
|---|---|---|---|---|---|---|
| | G48M36 | G45M24 | G42M20 | G40M16 | G38M16T2 | G38M16T4 |
| Logic industry "Node Range" Labeling (nm) | "5" | "3" | "2.1" | "1.5" | "1.0 eq" | "0.7 eq" |
| IDM-Foundry node labeling | i7-f5 | i5-f3 | i3-f2.1 | i2.1-f1.5 | i1.5e-f1.0e | i1.0e-f0.7e |
| Logic device structure options | FinFET | finFET LGAA | LGAA | LGAA | LGAA-3D | LGAA-3D |
| Mainstream device for logic | finFET | finFET | LGAA | LGAA | LGAA-3D | LGAA-3D |
| INTERCONNECT TECHNOLOGY | | | | | | |
| Number of Mx layers | 3 | 3 | 3 | 3 | 2 | 2 |
| Number of P80 layers | 12 | 14 | 14 | 15 | 17 | 17 |
| Number of P720 layers | 2 | 2 | 2 | 2 | 2 | 2 |
| Routing resources - Mx+P80+P720 - relative | 1.00 | 1.00 | 0.98 | 0.98 | 1.00 | 1.00 |
| Number of wiring layers - M1+Mx+P80+P720 | 18 | 20 | 20 | 21 | 22 | 22 |
| Mx - tight-pitch interconnect resistance (Ohms/um) | 133 | 301 | 474 | 920 | 1447 | 1447 |
| Mx - tight-pitch interconnect capacitance (aF/um) | 208 | 208 | 208 | 208 | 208 | 208 |
| Vx - tight-pitch interconnect via resistance (Ohms/via) | 28.4 | 50.0 | 52.8 | 38.3 | 63.9 | 63.9 |
| MP80 - 80nm pitch interconnect resistance (Ohms/um) | 13.3 | 13.3 | 13.3 | 13.3 | 13.3 | 13.3 |
| MP80 - 80nm pitch interconnect capacitance (aF/um) | 198 | 198 | 198 | 198 | 198 | 198 |
| VP80 - 80nm pitch interconnect via resistance (Ohms/via) | 5.0 | 5.0 | 5.0 | 5.0 | 5.0 | 5.0 |
| Aspect ratio - M0, M1, Mx, MP80, MP720 | 1.5-2.5 | 1.5-2.5 | 1.5-2.5 | 1.5-2.5 | 1.5-2.5 | 1.5-2.5 |
| Metallization - M0 | Co, Cu | Co, Ru | Co, Ru | Co, Ru | Co, Ru | Co, Ru |
| Barrier - Cu M0 | 2.0nm TaNRuCo, TaNCo | | | | | |
| Barrier - Non-Cu M0 | 1.0nm TiN+WC | 0.5nm TiN+WC | 0.5nm TiN+WC | 0.5nm TiN+WC | 0.5nm TiN+WC | 0.5nm TiN+WC |
| Di-electrics k value - M0,M1,Mx | SiCOH (2.70-3.20) | SiCOH (2.70-3.20) | SiCOH (2.70-3.20) | SiCOH (2.70-3.20) | SiCOH (2.70-3.20) | SiCOH (2.70-3.20) |
| Metallization - M1, Mx | Cu | Cu | Cu, Co, Ru | Cu, Co, Ru | Cu, Co, Ru | Cu, Co, Ru |
| Barrier metal - M1,Mx | 2.0nm TaNRuCo, TaNCo | 1.5nm TaNRuCo, TaNCo | 0.5nm TiN+WC | 0.5nm TiN+WC | 0.5nm TiN+WC | 0.5nm TiN+WC |
| Di-electrics k value - M0,M1,Mx | SiCOH (2.70-3.20) | SiCOH (2.70-3.20) | SiCOH (2.70-3.20) | SiCOH (2.70-3.20) | SiCOH (2.70-3.20) | SiCOH (2.70-3.20) |
| Metallization - MP80, MP720 | Cu | Cu | Cu | Cu | Cu | Cu |
| Barrier metal - M1,Mx | 2.5nm TaNRuCo, TaNCo | 2.5nm TaNRuCo, TaNCo | 2.5nm TaNRuCo, TaNCo | 2.5nm TaNRuCo, TaNCo | 2.5nm TaNRuCo, TaNCo | 2.5nm TaNRuCo, TaNCo |
| Di-electrics k value - MP80 | SiCOH (2.40-2.55) Airgap (1.0) | SiCOH (2.40-2.55) Airgap (1.0) | SiCOH (2.20-2.55) Airgap (1.0) | SiCOH (2.20-2.55) Airgap (1.0) | SiCOH (2.20-2.55) Airgap (1.0) | SiCOH (2.20-2.55) Airgap (1.0) |
| TDDB (MV/cm) - M0,M1,Mx | | | | | | |
| Jmax (MA/cm2 at 105C) - M0,M1,Mx | | | | | | |

## Technology Driven: (4) low-k dielectric

- Static power → Leakage

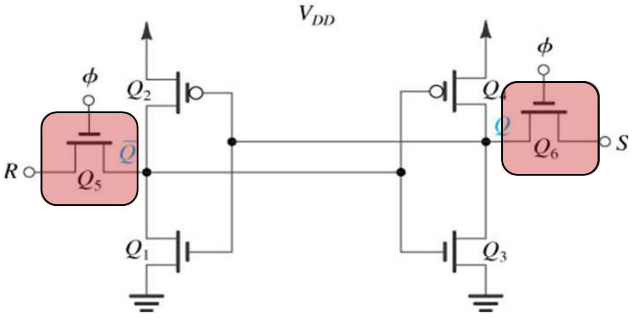





## Technology Driven: (5) Dual-Vt

- Static power → Leakage

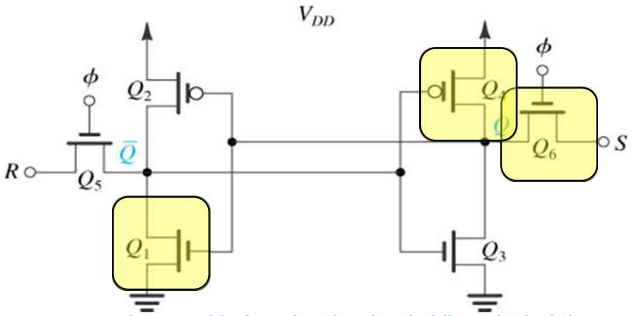



- low-Vt access transistors

Low-Vt

- High-Vt access transistors

High-Vt




decreased leakage by 70x when holding a logical '0.

3. ULP Digital Circuit Design

#4/4

1. Conventional CMOS Digital Design
   - Combinational Logic
   - Sequential Logic

2. Technology Driven:
   - (1) Reducing $V_{DD}$ (2) dual-voltage CPUs (3) dynamic voltage scaling (4) undervolting
   - (1) Reducing Frequency (2) underclocking (3) dynamic frequency scaling
   - (1) Reducing Capacitance (2) reduce transistor size (3) add functionality in IC (replace PCB) (4) low-k dielectric (5) Dual-Vt

3. Design Driven:
   - (1) Lower logic Swing (2) Reduce the switching activity (3) Optimizing machine code
   - (1) clock gating (2) Power gating (3) globally asynchronous locally synchronous
   - Recycling energy stored in the capacitors (1) adiabatic circuit (2) energy recovery logic



## Power consumption reduction techniques

$$P = f\, C_l\, V_{DD}^2$$

Technology Driven:

- (1) Reducing $V_{DD}$ (2) dual-voltage CPUs (3) dynamic voltage scaling (4) undervolting
- (1) Reducing Frequency (2) underclocking (3) dynamic frequency scaling
- (1) Reducing Capacitance (2) reduce transistor size (3) add functionality in IC (replace PCB) (4) low-k dielectric (5) Dual-Vt

Design Driven:

- (1) Lower logic Swing (2) Reduce the switching activity (3) Optimizing machine code
- (1) Clock gating (2) Power gating (3) Globally asynchronous locally synchronous
- Recycling energy stored in the capacitors (1) Adiabatic circuit (2) Energy recovery logic

## Design Driven: (1) Lower logic Swing

Dynamic power → $P = \alpha\ f\ C_l\ V_{\text{Swing}} V_{DD}$

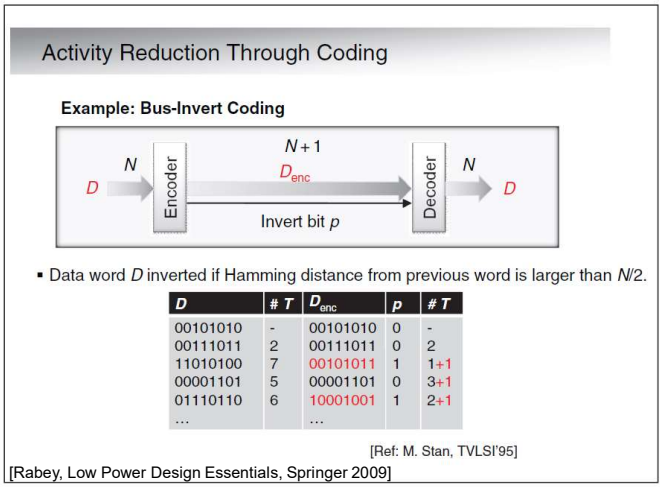
- (1) $V_{\text{Swing}}$: Similar to dual-Voltage approach
  - Reduces noise margins
  - Area and delay overhead
  - May require level regeneration
- Good technique to apply on clock distribution
  - Reduced swing
  - Alternative clock distribution schemes
  - Avoid global clock



## Design Driven: (2) Reduce the switching activity (3) Optimizing machine code

Dynamic power → $P = \alpha\ f\ C_l\ V_{\text{Swing}} V_{DD}$

- (2) and (3) $\alpha$:
  - @coding level: seek sequences with less 0↔1 transitions
    - Bus-Invert Coding; Advanced bus-invert coding; Coding for address busses; Full-fledged channel coding

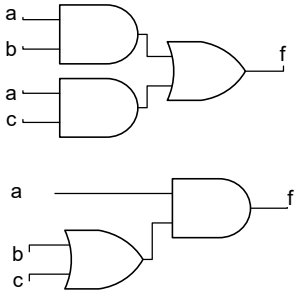


[Rabey, Low Power Design Essentials, Springer 2009]

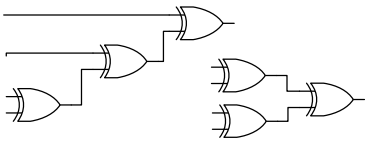
## Design Driven: (2) Reduce the switching activity (3) Optimizing machine code

Dynamic power → $P = \alpha\, f\, C_l\, V_{\text{Swing}} V_{DD}$

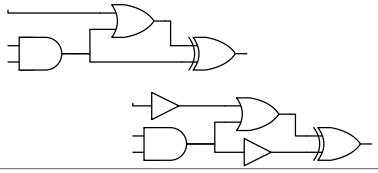
- (2) and (3) $\alpha$:
  - Find and optimize repetitive regular structures (Reduce $C$ may increase $\alpha$)




- Reduce glitching
  - restructuring



  - by balancing logic

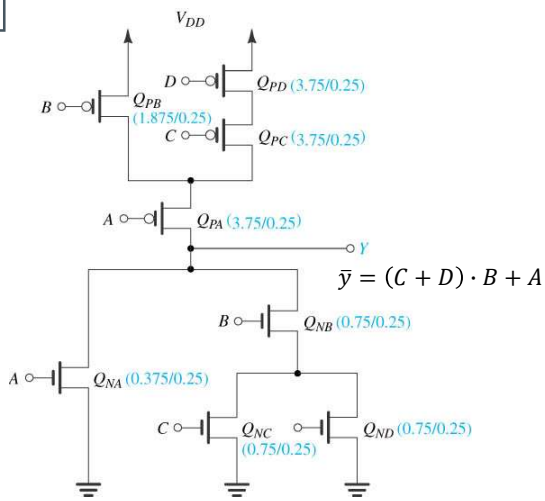




## Design Driven: (2) Reduce the switching activity (3) Optimizing machine code

Dynamic power → $P = \alpha\, f\, C_l\, V_{\text{Swing}} V_{DD}$

- (2) and (3) $\alpha$:
  - @gate level: Custom blocks with more complex custom gates;



## Design Driven: (1) clock gating

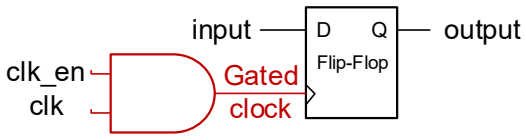
- **Clock Gating:** simple concept, **turn off** the design when not needed – without affecting the functionality.
  - introduced in the design as part of functionality
  - automatically by the tool

- **reduce switching**
- **reduce glitches**

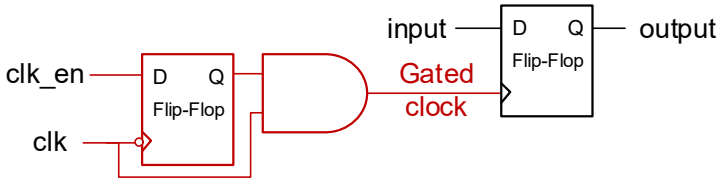


## Design Driven: (1) clock gating

- AND-based
  - has glitches

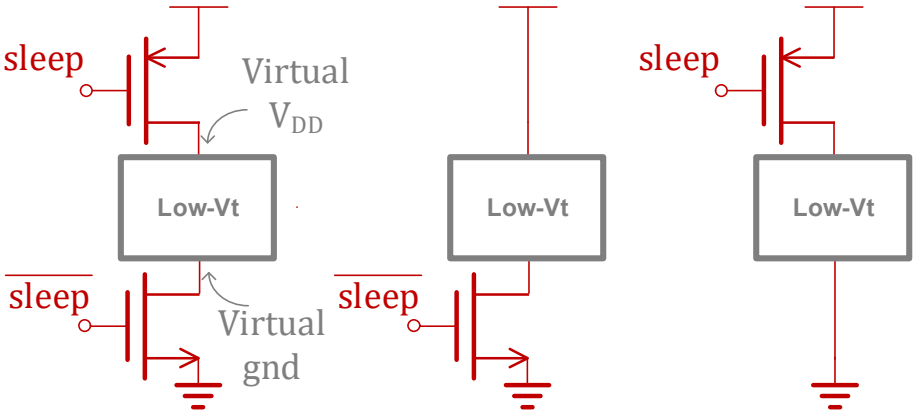



- Latch-based



## Design Driven: (2) Power gating

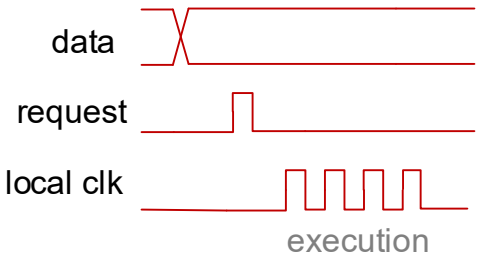
- Disconnect module from supply rail(s) during standby
- Most effective with high-*Vt* transistors
- footer+header  footer  header




- **reduce leakage**
- May use |higher| voltages at sleep signals
- Noisier supplies may require decoupling capacitor



## Design Driven: (3) globally asynchronous locally synchronous




- **reduce activity**
- reduce clock load driving

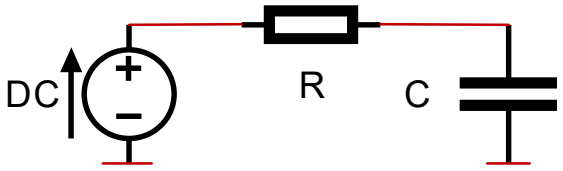
## Design Driven: (1) adiabatic circuit

Energy charging capacitors

Voltage step: [1] $v_{DC}: \mathbf{0} \rightarrow \mathbf{1}$ $\quad E_B = \underbrace{QV_{DC}}_{battery} = C_l V_{DC}^2 = \underbrace{\frac{1}{2} C_l V_{DC}^2}_{\text{stored in } C_l} + \underbrace{\frac{1}{2} C_l V_{DC}^2}_{\text{dissipated in R}}$

Current step: [1] $I_{DC}$: constant $\quad E_B = E_C + E_R = \underbrace{\frac{1}{2} C_l V_{DC}^2}_{\text{stored in } C_l} + \underbrace{\frac{RC}{T} C_l V_{DC}^2}_{\text{dissipated in R}}$

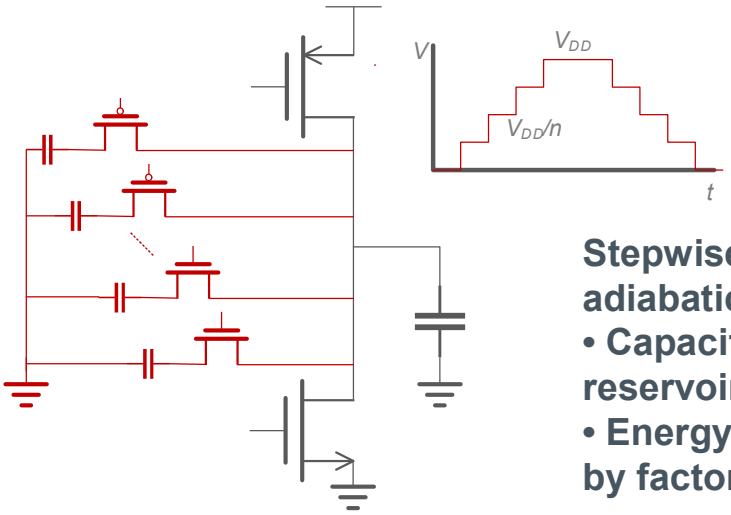
Current more eficiente if $T > 2RC$






## Design Driven: (1) adiabatic circuit

Quasi-adiabatic charging capacitors




**Stepwise approximation of adiabatic (dis)charging**
- **Capacitors acting as "charge reservoir"**
- **Energy drawn from supply reduced by factor *n***

# Design Driven: (2) energy recovery logic

IEEE JOURNAL OF SOLID-STATE CIRCUITS, VOL. 42, NO. 1, JANUARY 2007

Energy-Efficient GHz-Class Charge-Recovery Logic

Visvesh S. Sathe, *Member, IEEE*, Juang-Ying Chueh, *Member, IEEE*, and Marios C. Papaefthymiou, *Senior Member, IEEE*

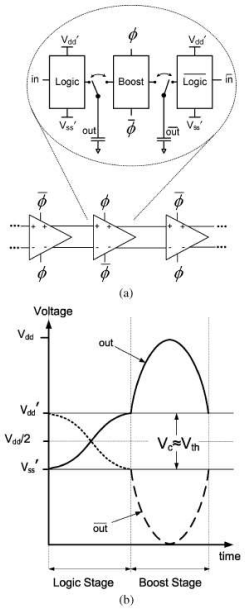



Fig. 1. Boost Logic. (a) Cascade. (b) Simplified waveform.