ECE 124a/256c
MOS Gate Models

Forrest Brewer
Displays: W. Burleson, K. Bernstein, P. Gronoski

Basics (MOS Electrical Model)

$$I_{ds} = \frac{kW}{L}((V_{gs} - V_T)V_F - \frac{V_F^2}{2})(1 + \lambda V_{ds})$$

- Nonlinear model with 3 conduction modes:
  - Linear Mode ($V_{ds} < V_{gs}$-V_T) and ($V_{ds} < V_{sat}$): $V_F = V_{ds}$
  - Saturation ($V_{ds} > V_{gs}$-V_T) and ($V_{ds} < V_{sat}$): $V_F = V_{gs}$-V_T
  - Velocity Saturation ($V_{ds} > V_{sat}$): $V_F = V_{sat}$
- $V_F = \mathrm{Min}(V_{gs}\text{-}V_T, V_{sat}, V_{ds})$

Body Effect

- Threshold is function of back potential
  - Increases difficulty of turn on for junction reverse bias increase

0.18μ	Φ_f	γ
N-type	0.4 (V)	0.32
P-type	0.4 (V)	-0.42

$$V_T = V_{T0} + \gamma\left(\sqrt{2\left|\phi_f\right| + V_{SB}} - \sqrt{2\left|\phi_f\right|}\right)$$

$$\phi_f = \frac{kT}{q}\ln\left(\frac{N_a}{n_i}\right)$$

$$\gamma = \frac{\sqrt{2q\varepsilon_{Si}N_a}}{C_{ox}}$$

Velocity Saturation

- Carrier Velocity Saturates at about $1.7\text{x}10^7$cm/s
- For short channel (small L) this occurs at V_{sat}
- Mobility (μ) is a function of doping and temperature

0.18μ	μ	V_{sat}
N-type	400 (cm^2/Vs)	0.8V
P-type	150 (cm^2/Vs)	2.2V

$$V_{sat} = \frac{(1.7\times10^7 cm/s)L(cm)}{\mu(cm^2/Vs)}$$

$$\mu(T) = \mu(300K)(T/300)^{-3/2}$$

MOS Capacitors

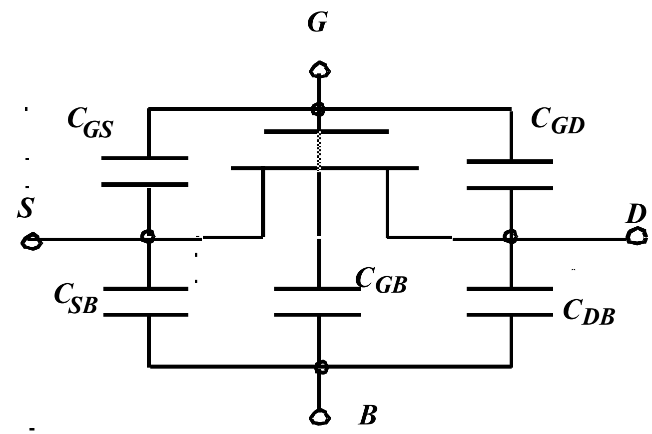
- Gate (assume constant) = ε_{Si} WL/t_{ox}
- Source/Drain
  - Bottom (Area) C_J m_J
  - SideWall (Perimeter) C_{JSW} m_{JSW}
- Equivalent Capacitance (Swing V_{low} to V_{high})

$$C_{eq} = \frac{\Delta Q_j}{\Delta V_D} = \frac{Q_j(V_{high}) - Q_j(V_{low})}{V_{high} - V_{low}} = K_{eq} C_{j0}$$

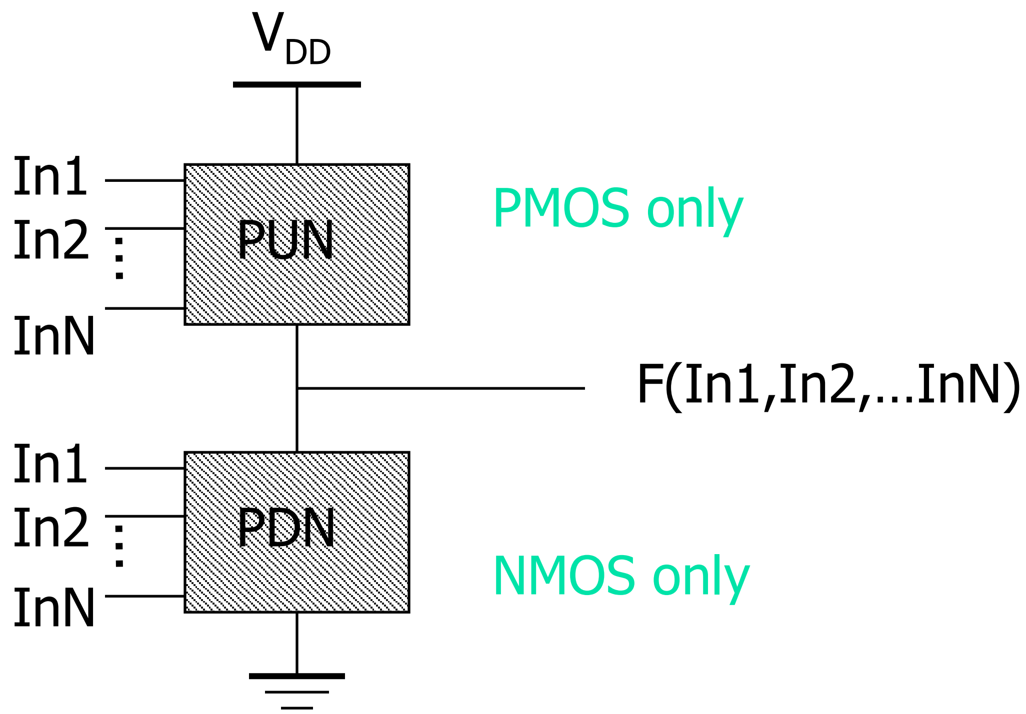
$$K_{eq} = \frac{-\phi_0^m}{(V_{high} - V_{low})(1 - m)}[(\phi_0 - V_{high})^{1-m} - (\phi_0 - V_{low})^{1-m}]$$

Transient Capacitor Parasitics

- Only capacitors which change potential over the swing are included.
- Cgs and Cgd are often modeled as Cg and Cgso, Cgdo. Cgdo models the feed though (input to output) capacitance
- For low swing rates, double Cgdo
- For high swing rates, start the output swing from the offset output voltage
  - Cgdo and Cload produce a capacitive voltage divider.

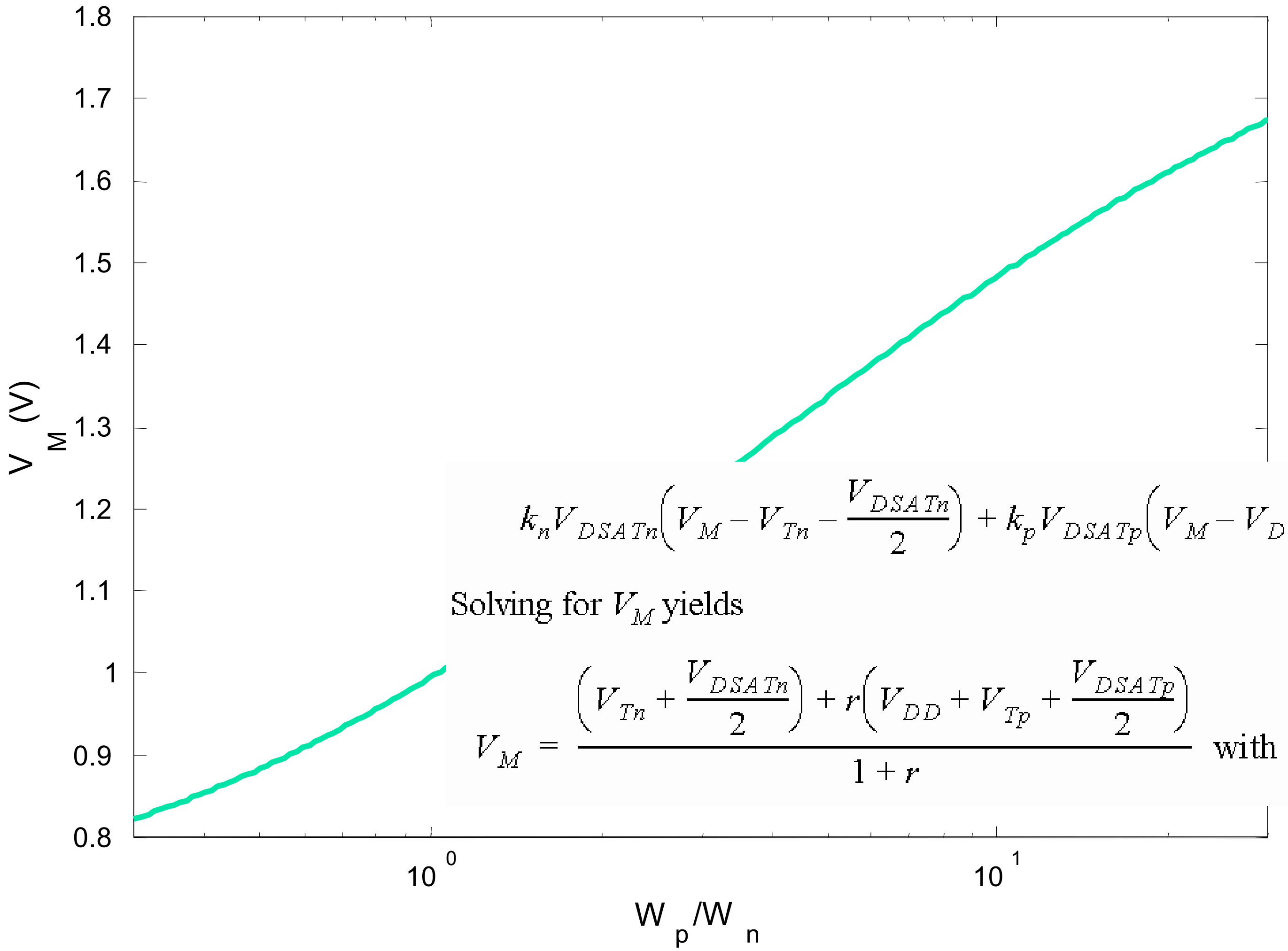


Static Complementary CMOS



PUN and PDN are logically dual logic networks

Inverter Threshold vs. N/P Ratio

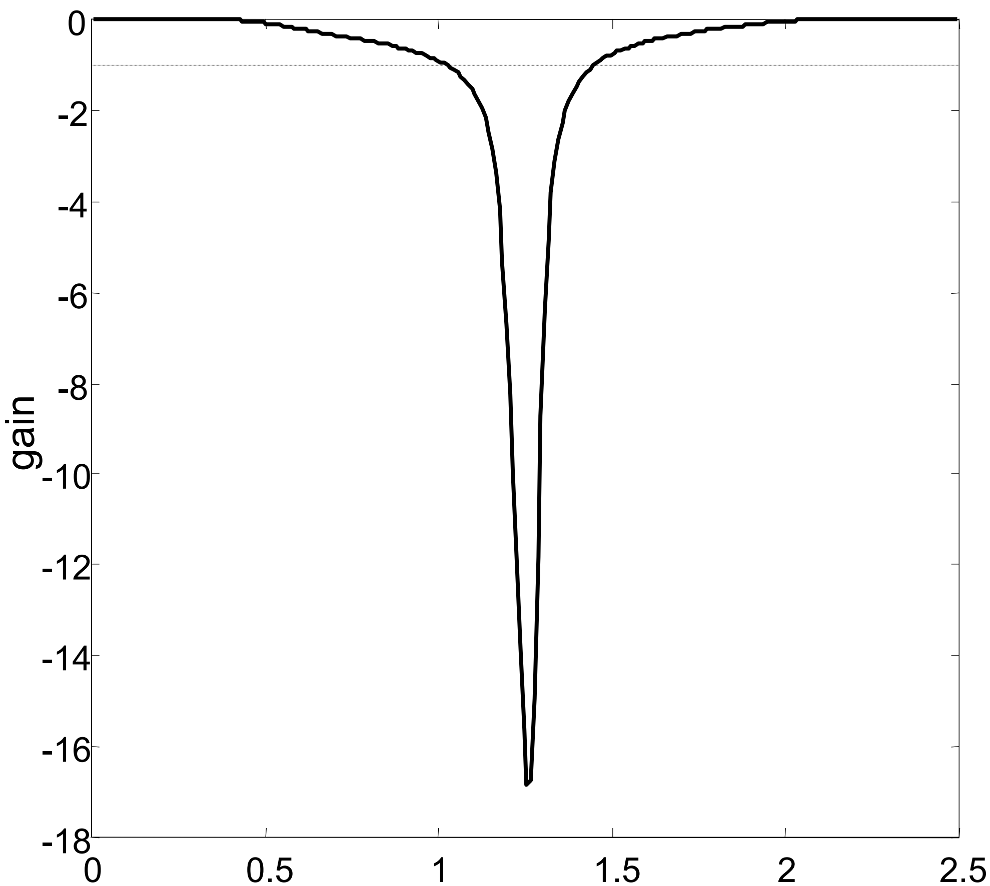


$$k_n V_{DSATn}\left(V_M - V_{Tn} - \frac{V_{DSATn}}{2}\right) + k_p V_{DSATp}\left(V_M - V_{DD} - V_{Tp} - \frac{V_{DSATp}}{2}\right) = 0$$

Solving for V_M yields

$$V_M = \frac{\left(V_{Tn} + \frac{V_{DSATn}}{2}\right) + r\left(V_{DD} + V_{Tp} + \frac{V_{DSATp}}{2}\right)}{1 + r} \quad \text{with} \quad r = \frac{k_p V_{DSATp}}{k_n V_{DSATn}} = \frac{\upsilon_{satp} W_p}{\upsilon_{satn} W_n}$$

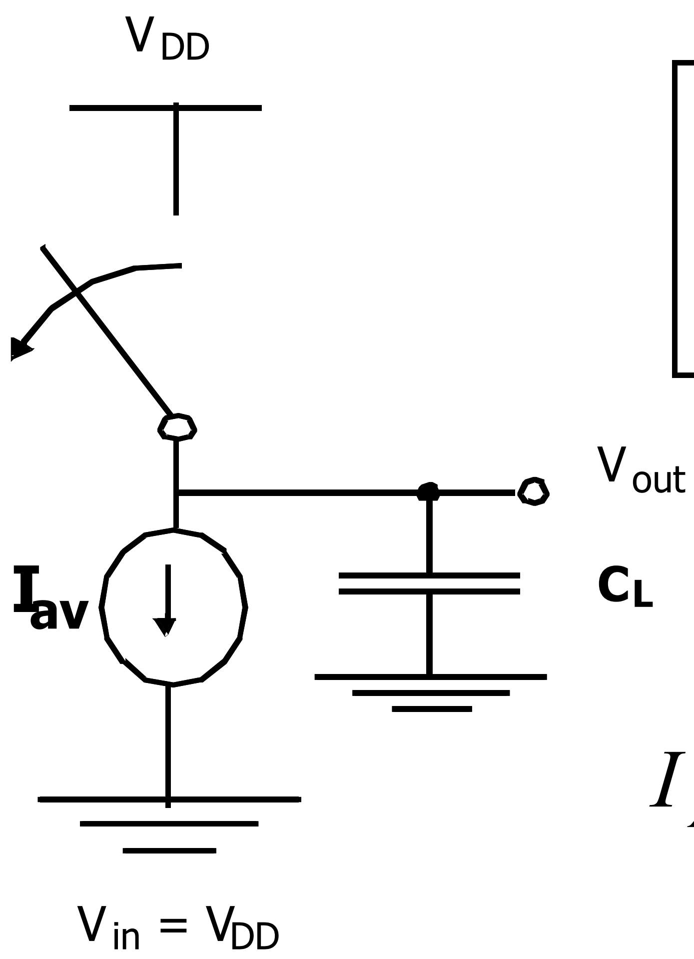
Inverter Gain



$$gain = \frac{-1}{I_D(V_M)} \frac{k_n V_{DSAT(n)} + k_p V_{DSAT(p)}}{\lambda_n - \lambda_p}$$

$$\approx \frac{1+r}{(V_M - V_{Tn} - V_{DSAT(n)}/2)(\lambda_n - \lambda_p)}$$

CMOS Inverter Propagation Delay

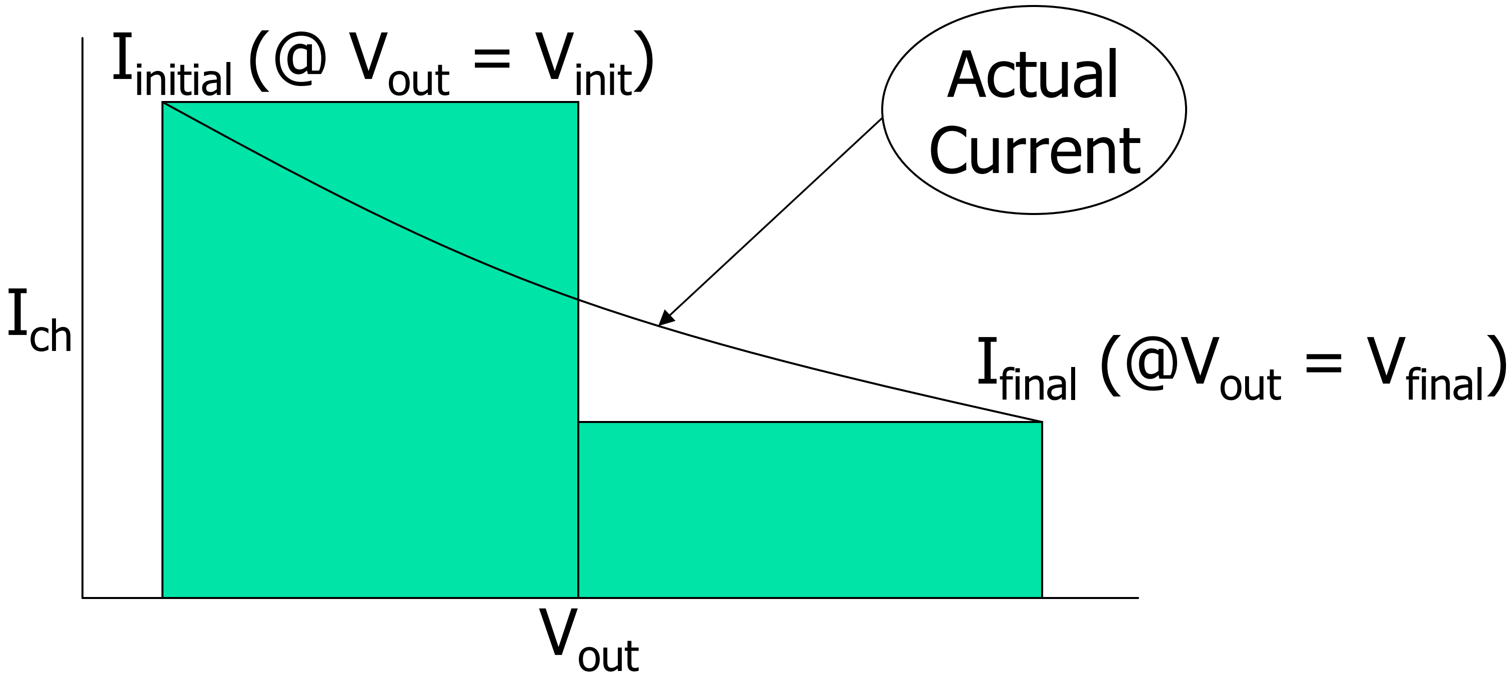


$$t_{pHL} = \frac{C_L \, V_{swing}/2}{I_{av}}$$

$$I_{AV} \approx \frac{I_{ds}(init) + I_{ds}(final)}{2}$$

Approximating I_{avg}

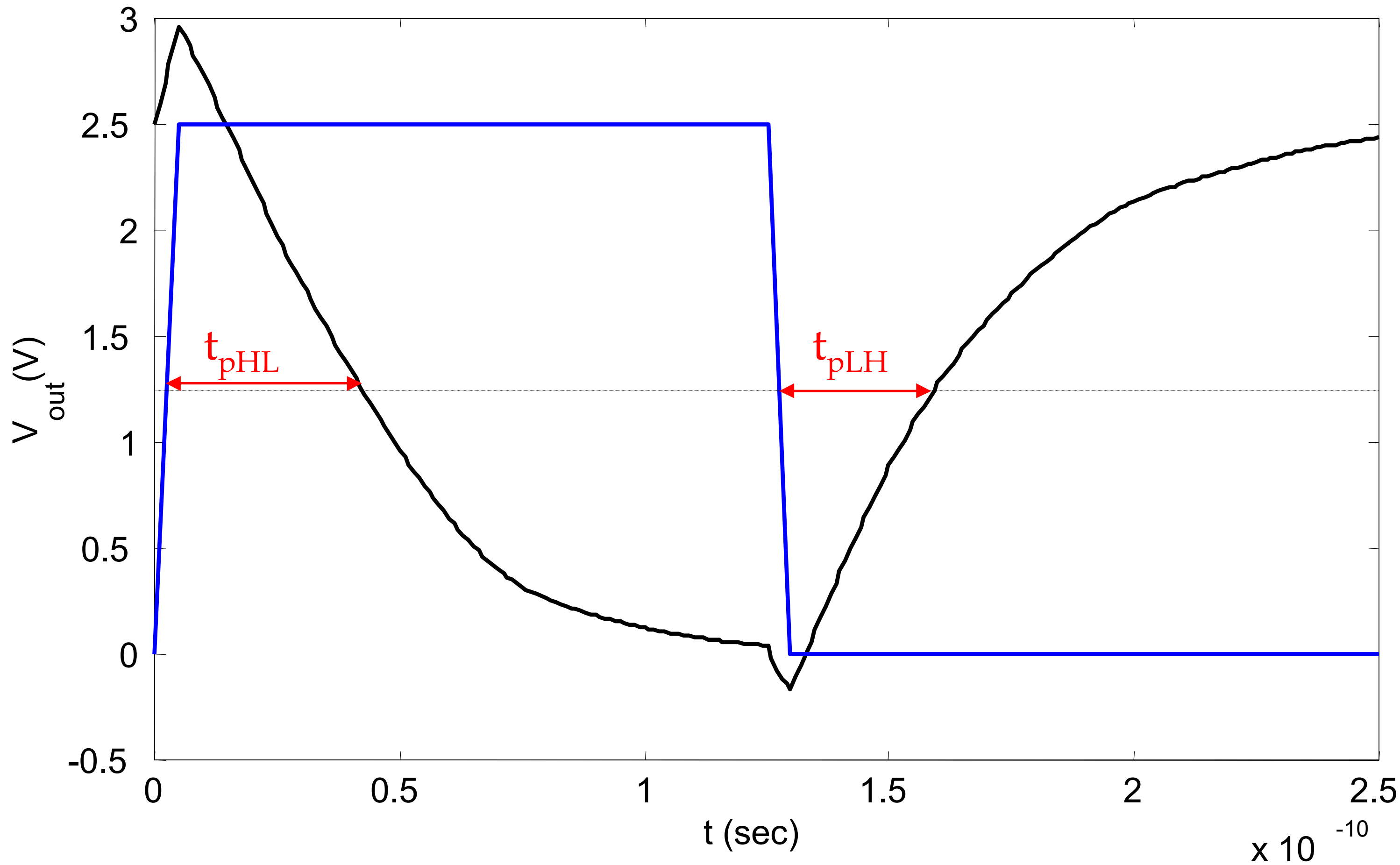
- Prescription from Hodges and Jackson
  - Assume input rise is instantaneous: ignore rise-time effects
- Average charging current at endpoints of swing
  - Initial point is usually a supply rail, final point is threshold of next gate



Hodges-Jackson Current Averaging

- FET's act as a current source
- Simple model for full-swing current:
  - I_1 is initial current at start of swing
  - I_2 is current at threshold of next stage
  - I_{avg} is approximated by $(I_1+I_2)/2$
- Delay $t = \dfrac{C\Delta V}{I_{avg}} = CR_{eff} = 0.69CR_{eq}$

Transient Response



Constructing a Complex Gate

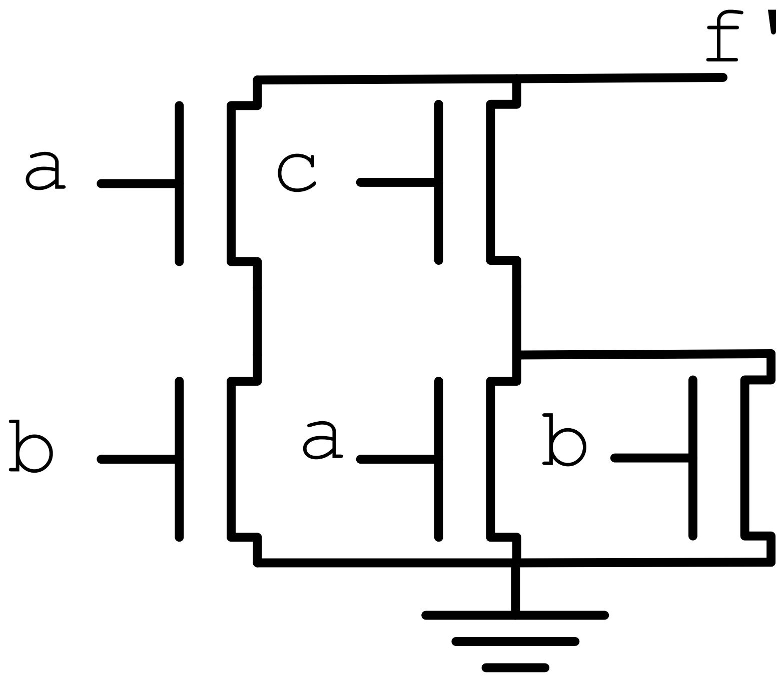
- Logic Dual need **not** be Series/Parallel Dual
- In general, many logical dual exist, need to choose one with best characteristics
- Use Karnaugh-Map to find good duals
  - Goal: find 0-cover and 1-cover with best parasitic or layout properties
  - Maximize connections to power/ground
  - Place critical transistors closest to output node
  - Know the order of arrival of signals! – order the transitions if possible

Example: Carry Gate

	C	C′
AB	0	0
AB′	0	1
A′B′	1	1
A′B	0	1

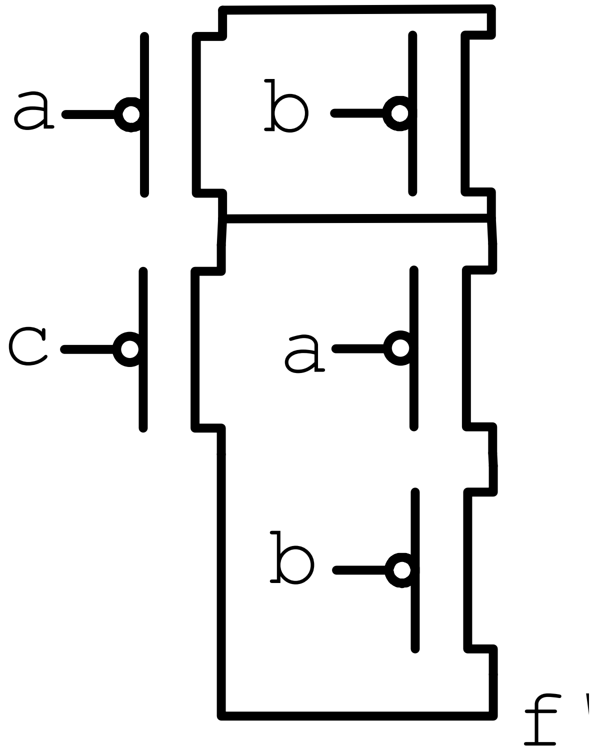
- F = (ab+bc+ac)′
- Carry ‘c’ is critical
- Factor c out: (Why c?)
- F=(ab+c(a+b))′
- 0-cover is n-pull up
- 1-cover is p-pull down

Example: Carry Gate (2)



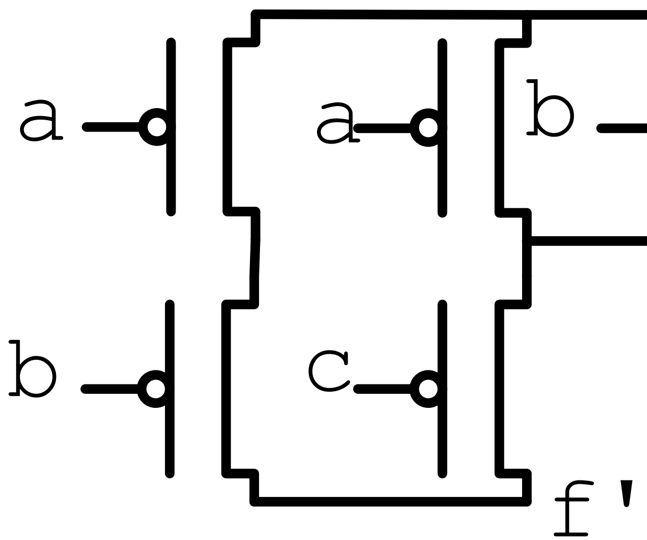
- Pull Down is easy
- Order by maximizing connections to ground and critical transistors
- For pull up – Might guess series parallel graph dual– but would guess wrong

Example: Carry Gate (3)



- Series/Parallel Dual
- 3-series transistors
- 2 connections to Vdd
- 7 floating capacitors

Example: Carry Gate (4)



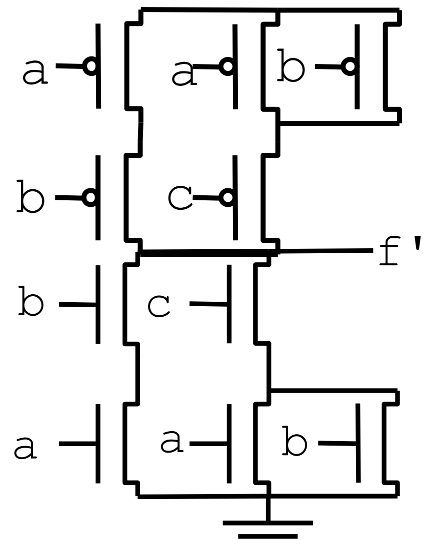
- Pull Up from 1 cover of Kmap
  - Get a'b'+a'c'+b'c'
  - Factor c' out
- 3 connections to Vdd
- 2 series transistors
- Co-Euler path layout
- Moral: Use Kmap!

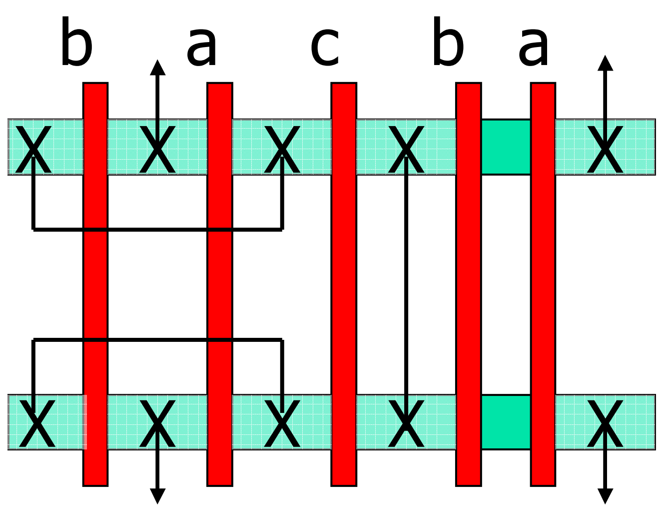
Euler Path

- For CMOS standard cell, and Euler path often helps to organize the transsistor order so that a faster, more dense cell can be constructed.
- Ideally, the p-fet and n-fet sub-circuits can be traversed in identical transistor orders to create a layout without diffusion (thinox) gaps.
- Euler Path:
  - Traversal of entire schematic (every transistor) without traversing any transistor twice.
  - Possible only if 0 or 2 odd nodes in schematics. Node count is the number of transistors incident on a common point.
  - If 0, any point can be start (will also be end) of path, for 2, one of the odd nodes is the start and the other is the end.

Euler Path II

- Eg. Carry Gate
  - Path: b-a-c-b-a or a-b-c-a-b or …
  - Can sometimes also minimize the routing by careful choice of order





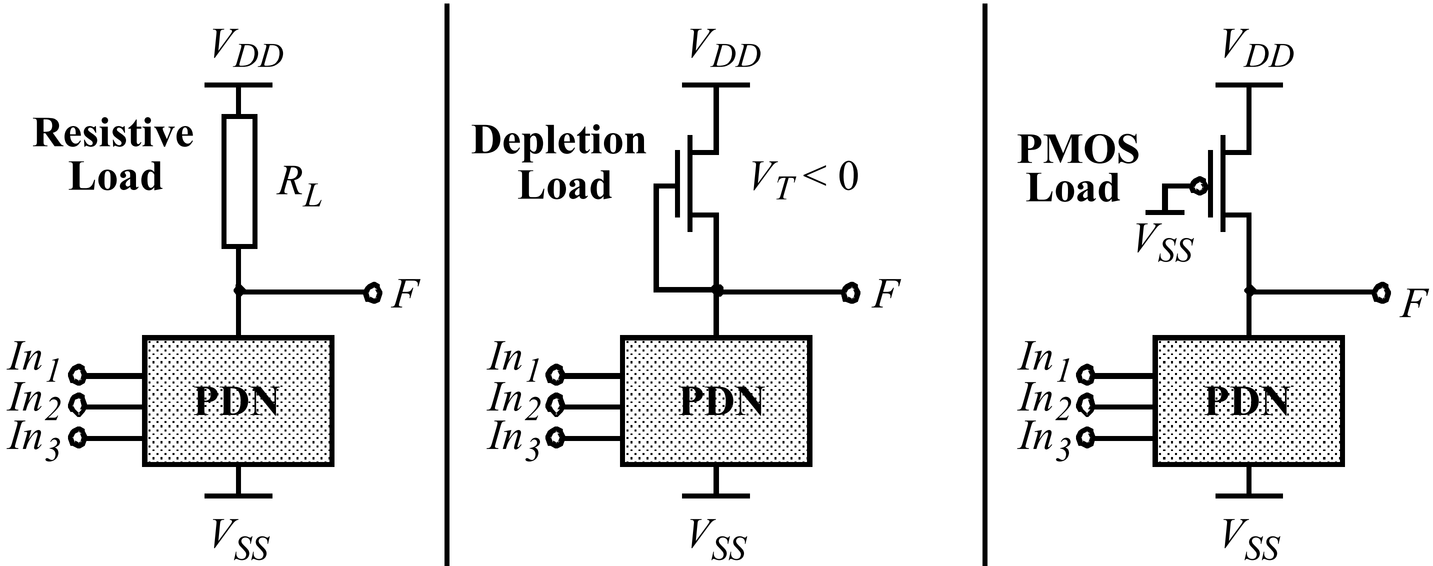
Static Logic: Rules of Thumb

1. Step-up (alpha) ratio of 4 produces minimum power-delay product
2. P vs. N (beta) ratio of 2 balances pull-up and pull-down times and noise margins.
3. Approximately 75% of static logic are NAND stacks (limit stack to 3-4, use ordering and tapering for speed)

More Rules of Thumb

1. Glitches consume approximately 15% of overall chip power.
2. Crossover (short-circuit) current consumes ~ 10% of a static chip's total power (but is a function of input/output slews, ie sizing)

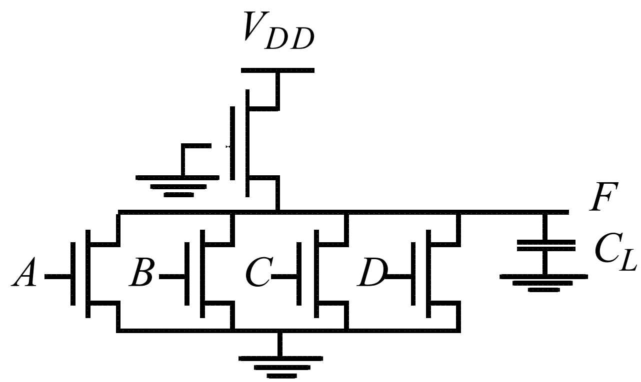
Ratio Logic



(a) resistive load (b) depletion load NMOS (c) pseudo-NMOS

Goal: to reduce the number of devices over complementary CMOS as a means to reduce parasitics (usually for performance).

Pseudo NMOS



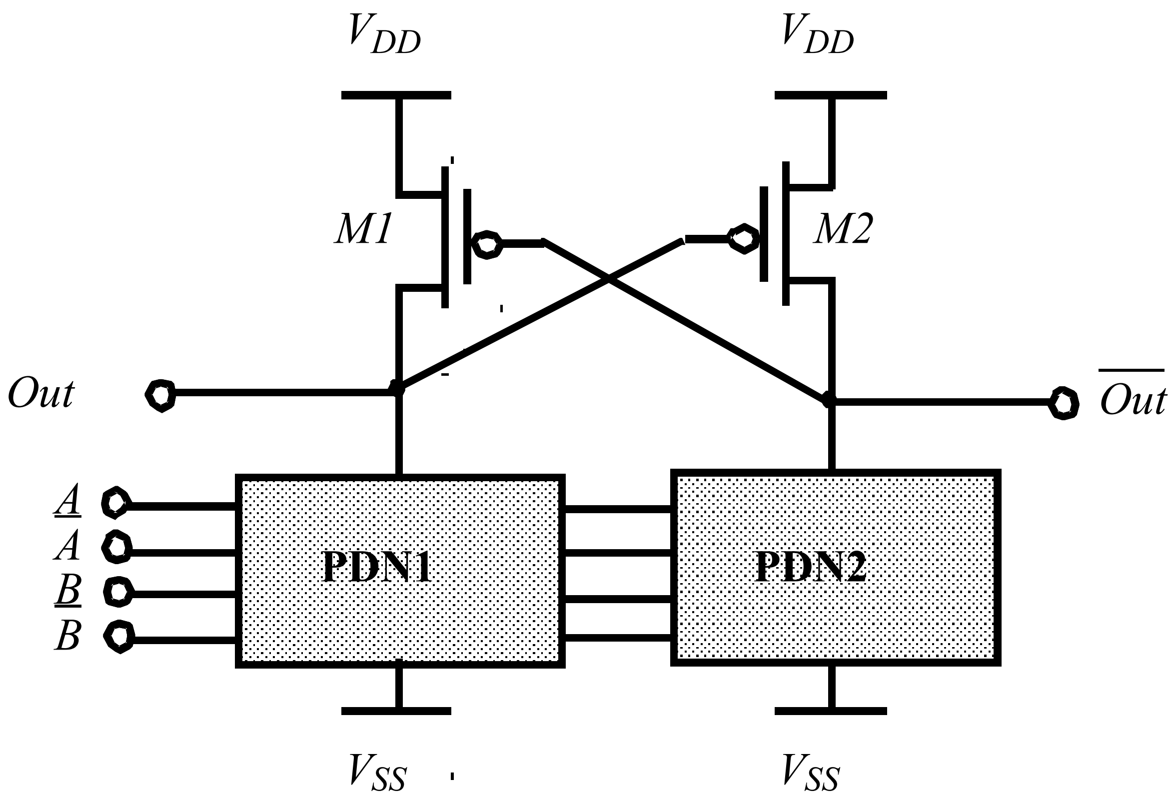
$V_{OH} = V_{DD}$ (similar to complementary CMOS)

$$k_n\left((V_{DD} - V_{Tn})V_{OL} - \frac{V_{OL}^2}{2}\right) = \frac{k_p}{2}(V_{DD} - |V_{Tp}|)^2$$

$$V_{OL} = (V_{DD} - V_T)\left[1 - \sqrt{1 - \frac{k_p}{k_n}}\right] \text{(assuming that } V_T = V_{Tn} = |V_{Tp}|\text{)}$$

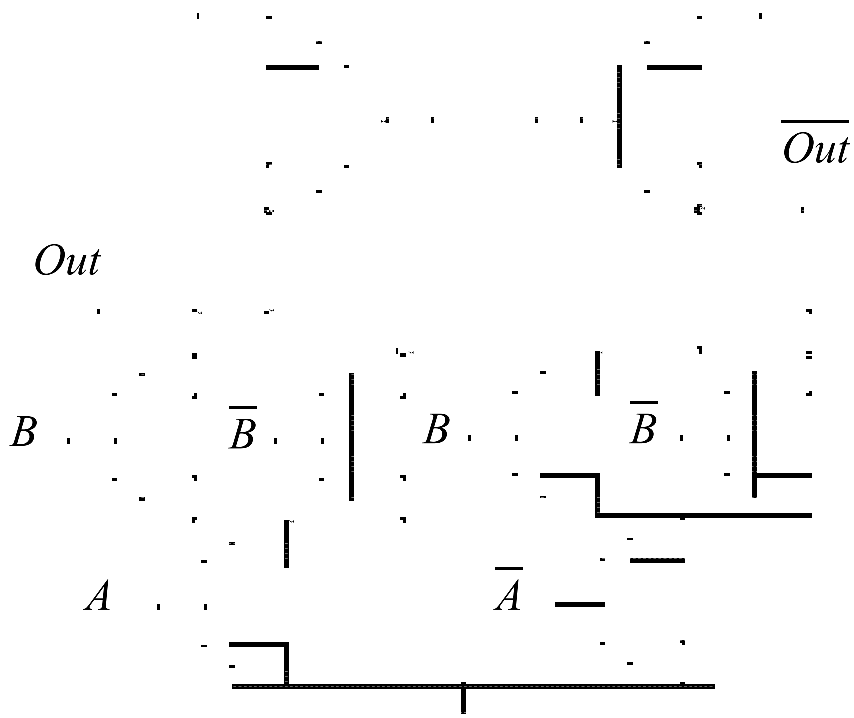
SMALLER AREA & LOAD <u>BUT</u> STATIC POWER DISSIPATION!!!

Even Better Noise Immunity/Density



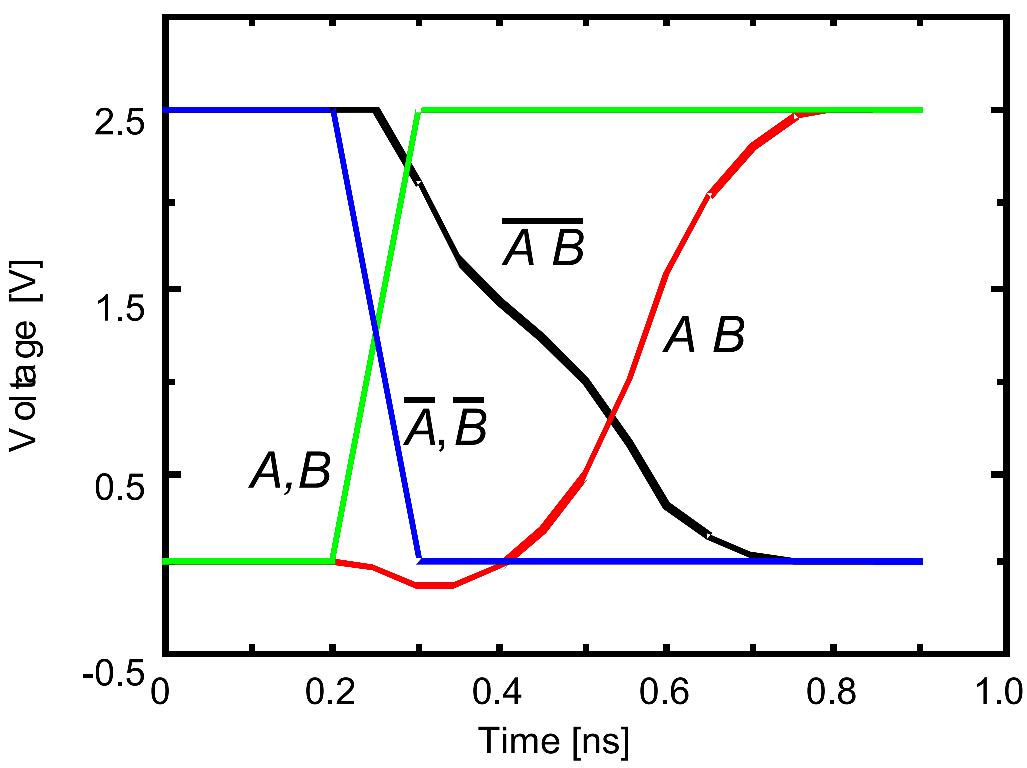
Differential Cascode Voltage Switch Logic (DCVSL)

DCVSL Example

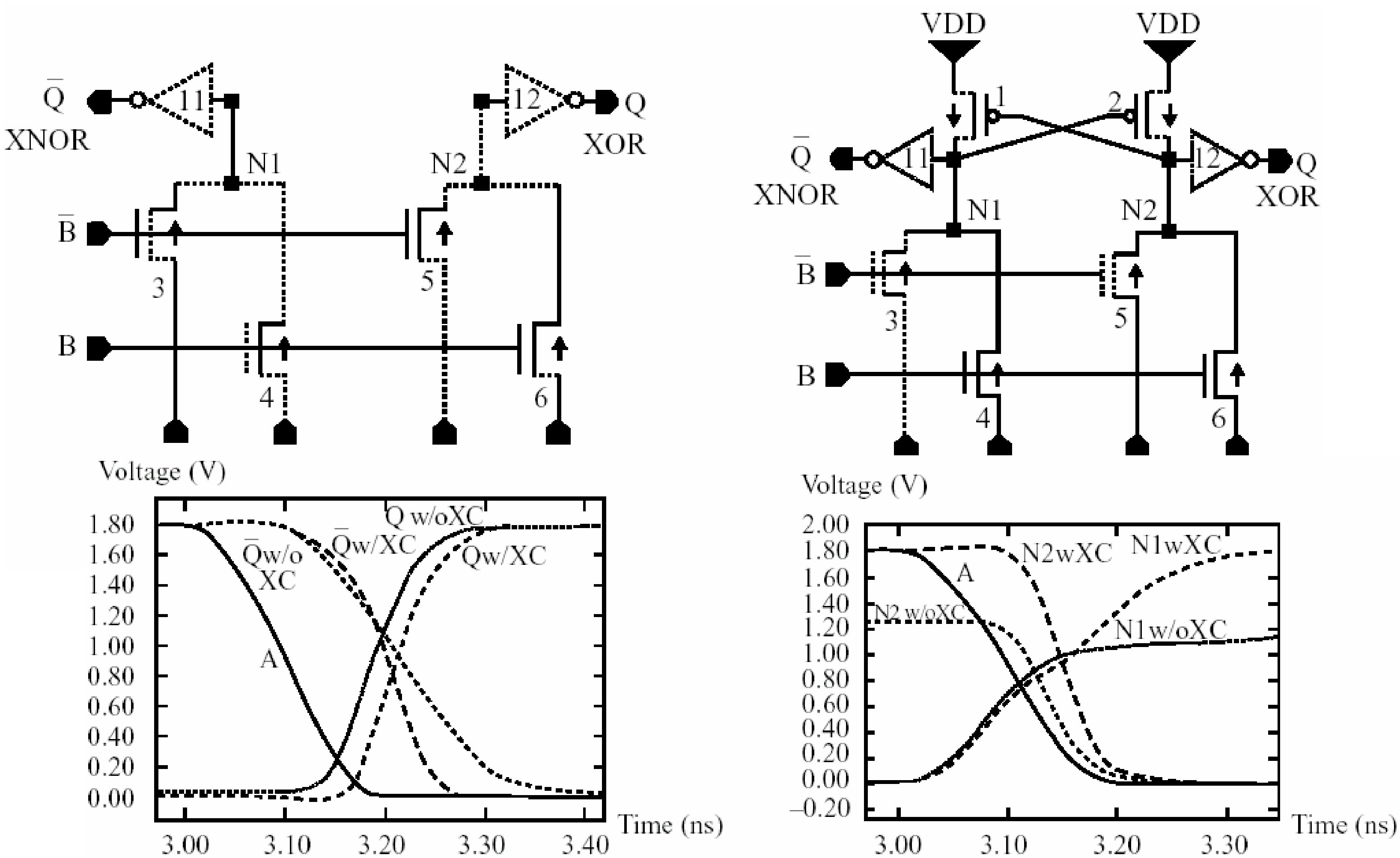


XOR-NXOR gate

DCVSL Transient Response



Complementary Pass Gate Logic (CPL)



Pass-gate Logic issues

- Limited fan-in
- Excessive fan-out
- Noise vulnerability (not restoring)
- Supply voltage offset/bias vulnerability
- Decode exclusivity (else short-circuit!)
- Poor high voltage levels if NMOS-only
- Body effect

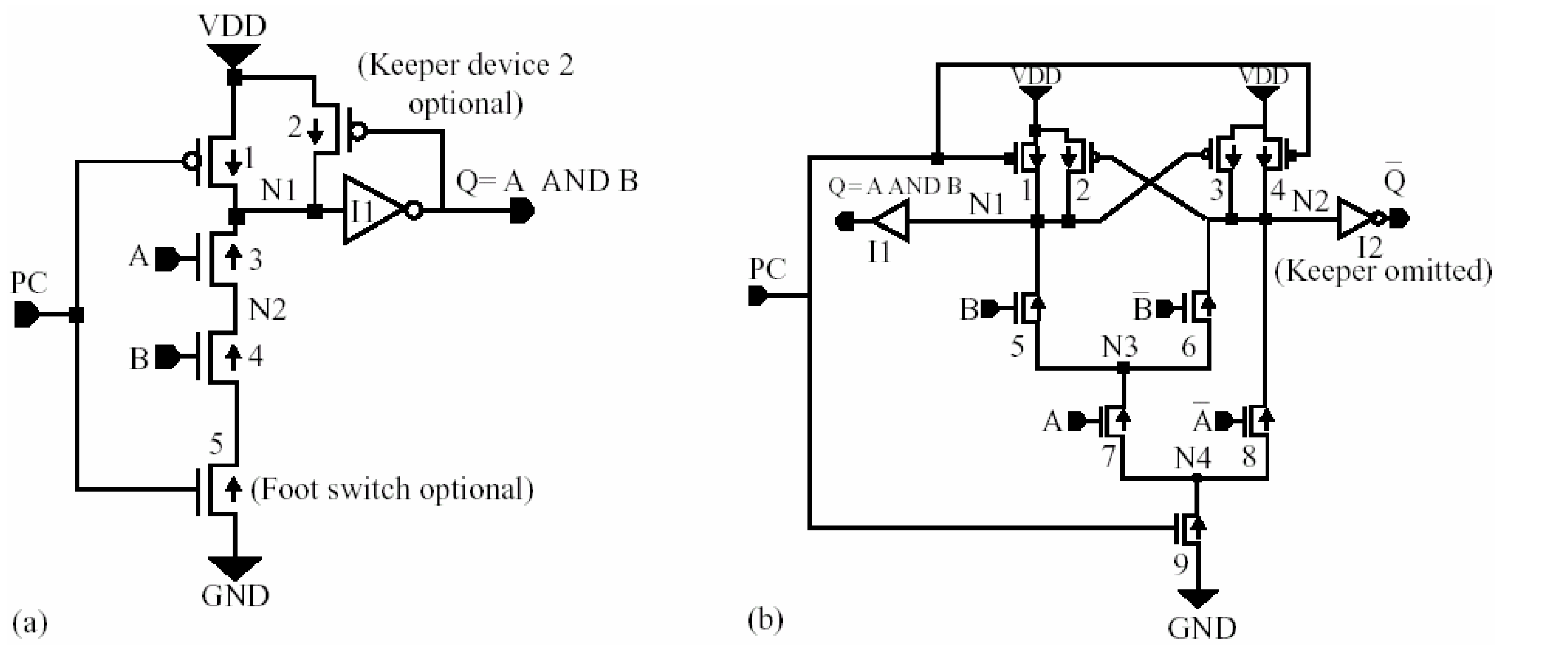
Pass-Gate Logic Rules of Thumb

- Pass-logic may consume half the power of static logic. But be careful of Vt drop resulting in static leakage.
- Pass-gate logic is not appropriate when long interconnects separate logic stages or when circuits have high fan-out load (use buffering).

Dynamic Logic

- Idea – use the low leakage of FETs to store charge instead of moving current. Provides higher density, faster operation at the cost of reduced noise immunity and tricky design…
- Domino is by far the most common style in CMOS

Domino logic (single and dual-rail)



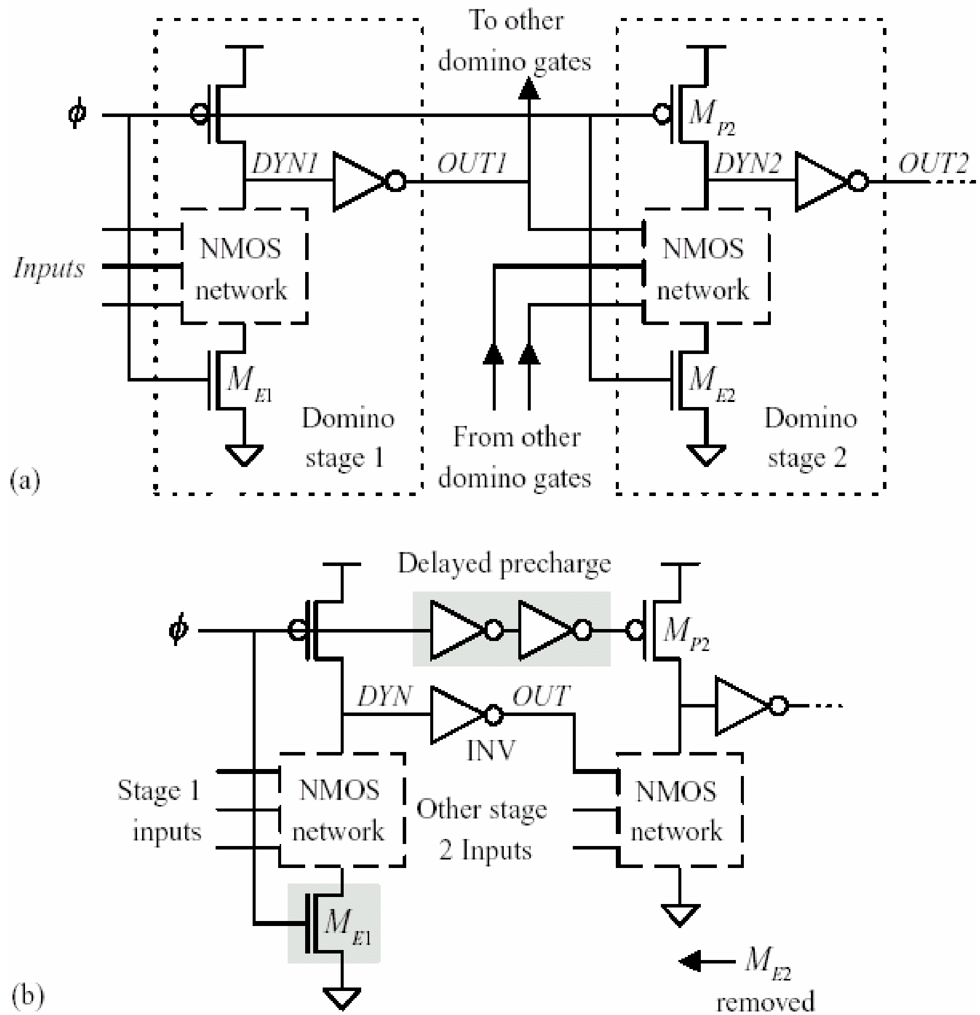
Dynamic Logic Rules of Thumb

- Dynamic logic is best for wide OR/NOR structure (e.g. bit-lines), providing 50% delay improvement over static CMOS.
- Dynamic logic consumes 2x power due to its phase activity (unconditional pre-charging), not counting clock power.

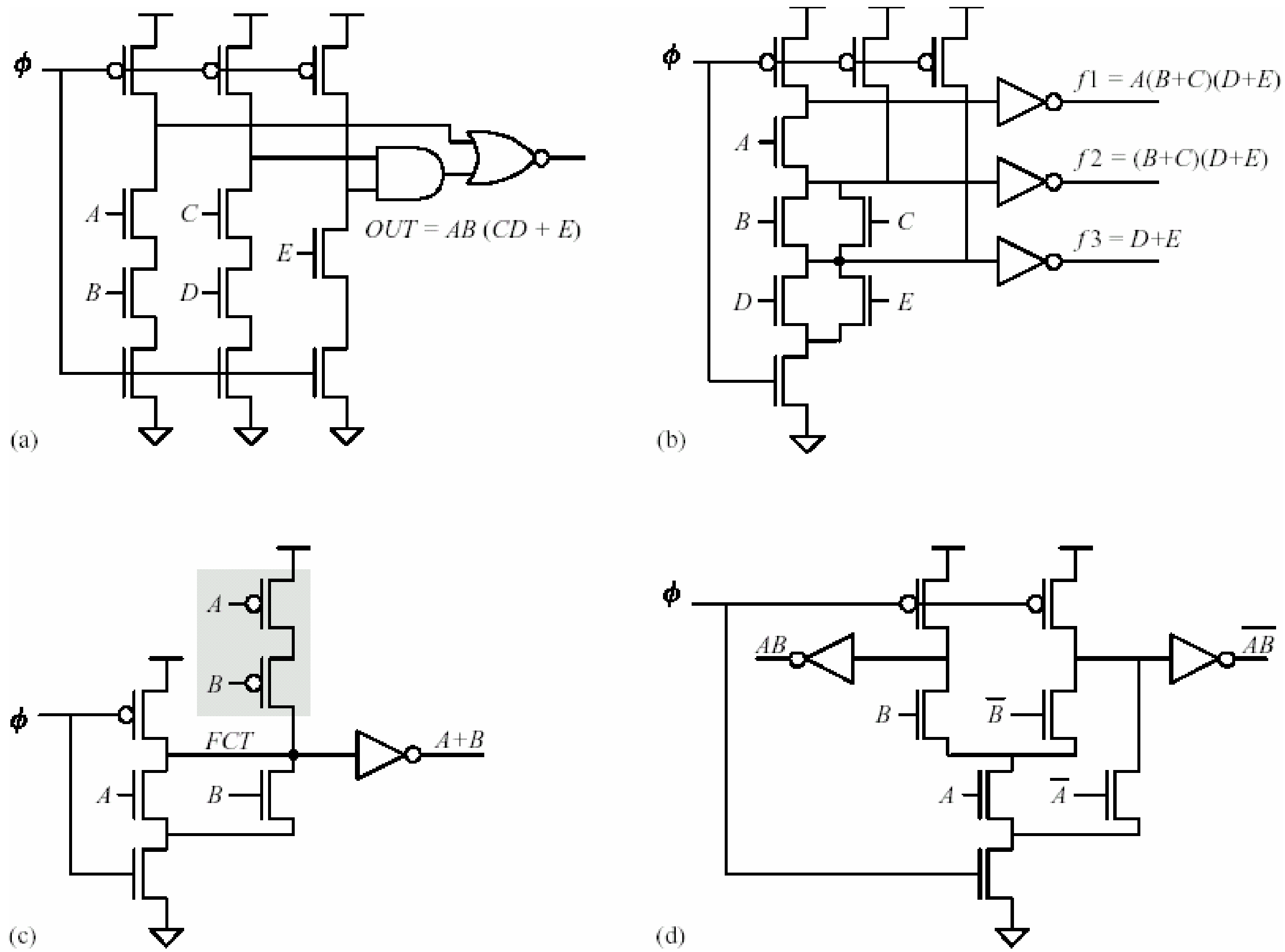
Domino Rules of Thumb

- Typical domino keepers have W/L = 5-20% of effective width of evaluate tree.
- Typical domino output buffers have a beta ratio of ~ 6:1 to push the switch point higher for fast rise-time but reduced noise margin.

Conventional and Delay-precharge domino



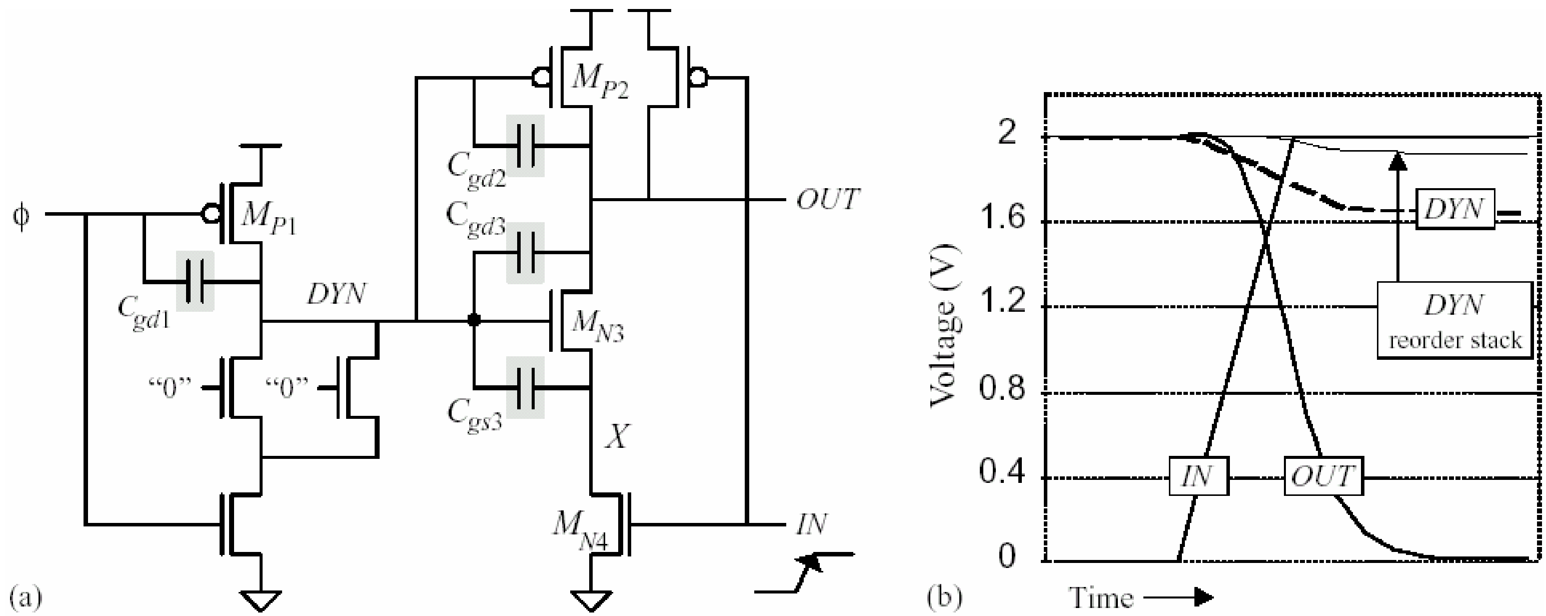
Advanced Domino Logic forms



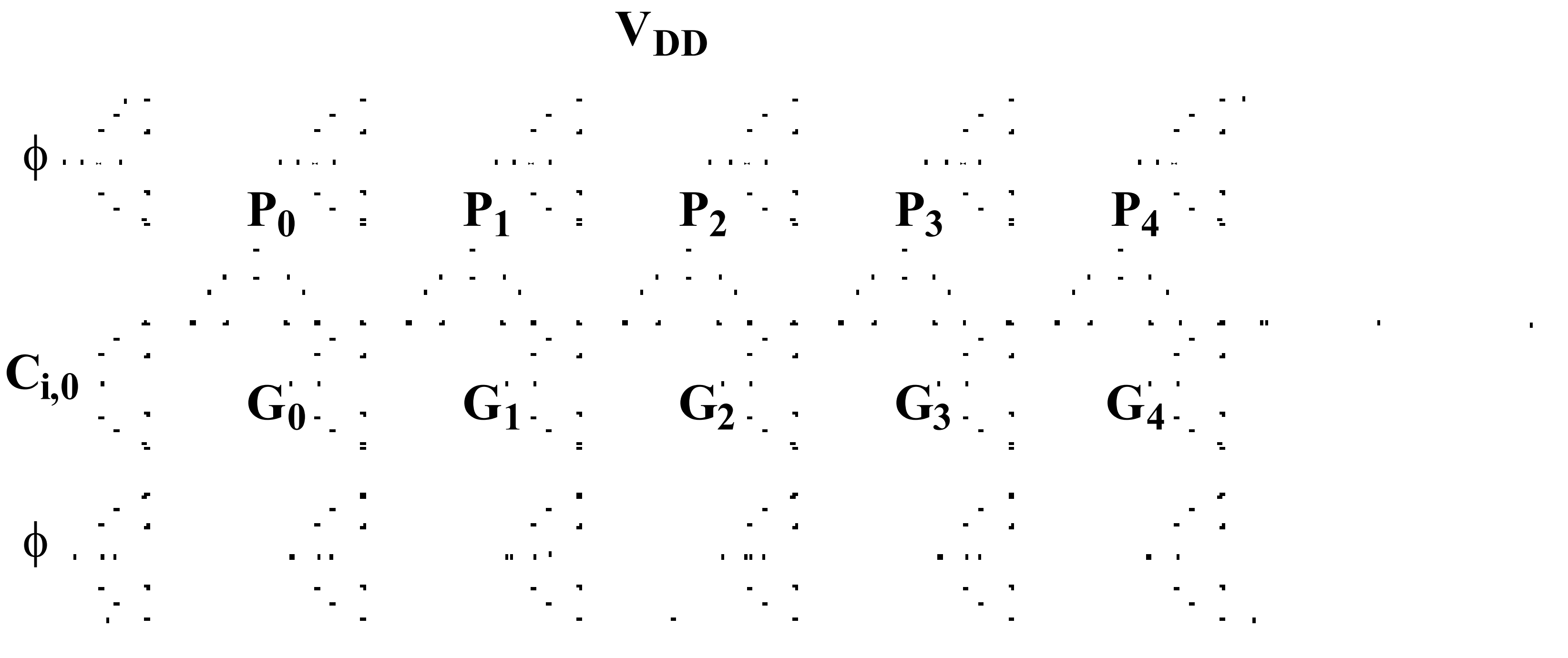
Concerns in Dynamic Logic

- Charge-sharing
- Charge-leakage
- Interconnect coupling
- Back-gate coupling
- Supply noise and variation

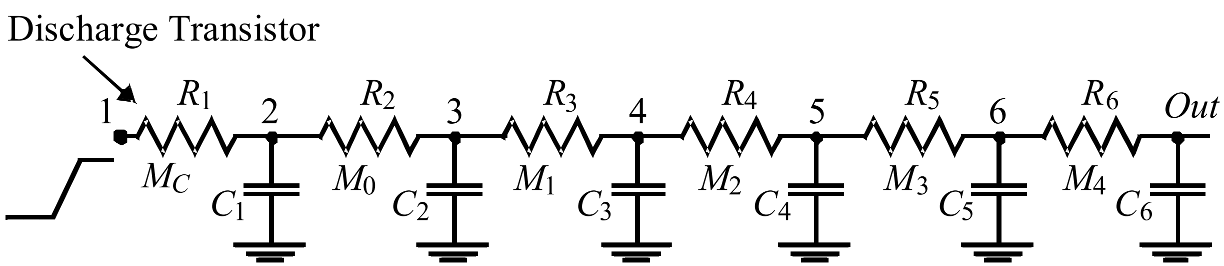
Back-gate coupling



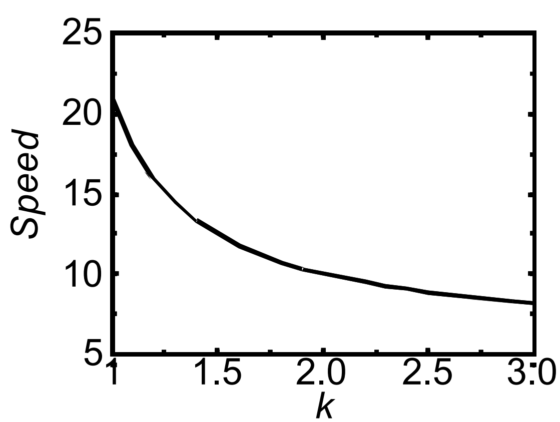
Manchester Carry Chain



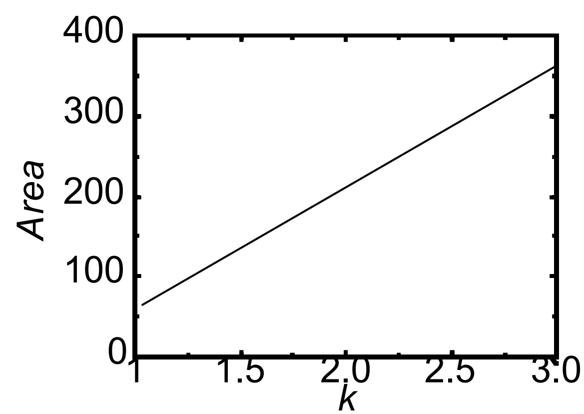
Sizing the Manchester Carry Chain



$$t = 0.69\left(\sum_{i=1}^{N} R_i \left(\sum_{j=i}^{N} C_j\right)\right), R = R_{eq}/k, C = C_p + kC_n$$

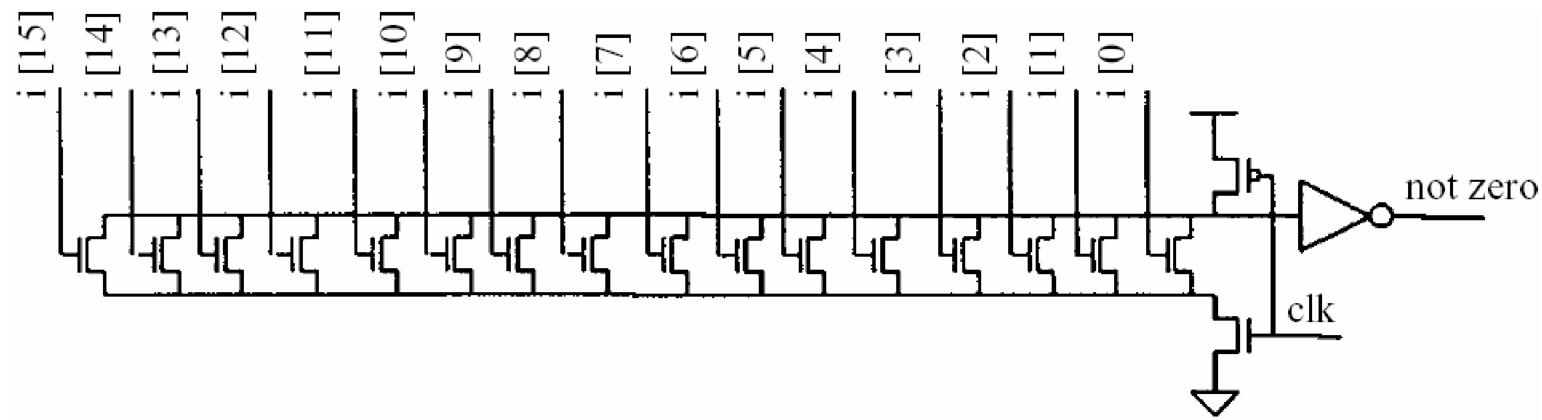


Speed (normalized by 0.69 *RC*)



Area (in minimum size devices)

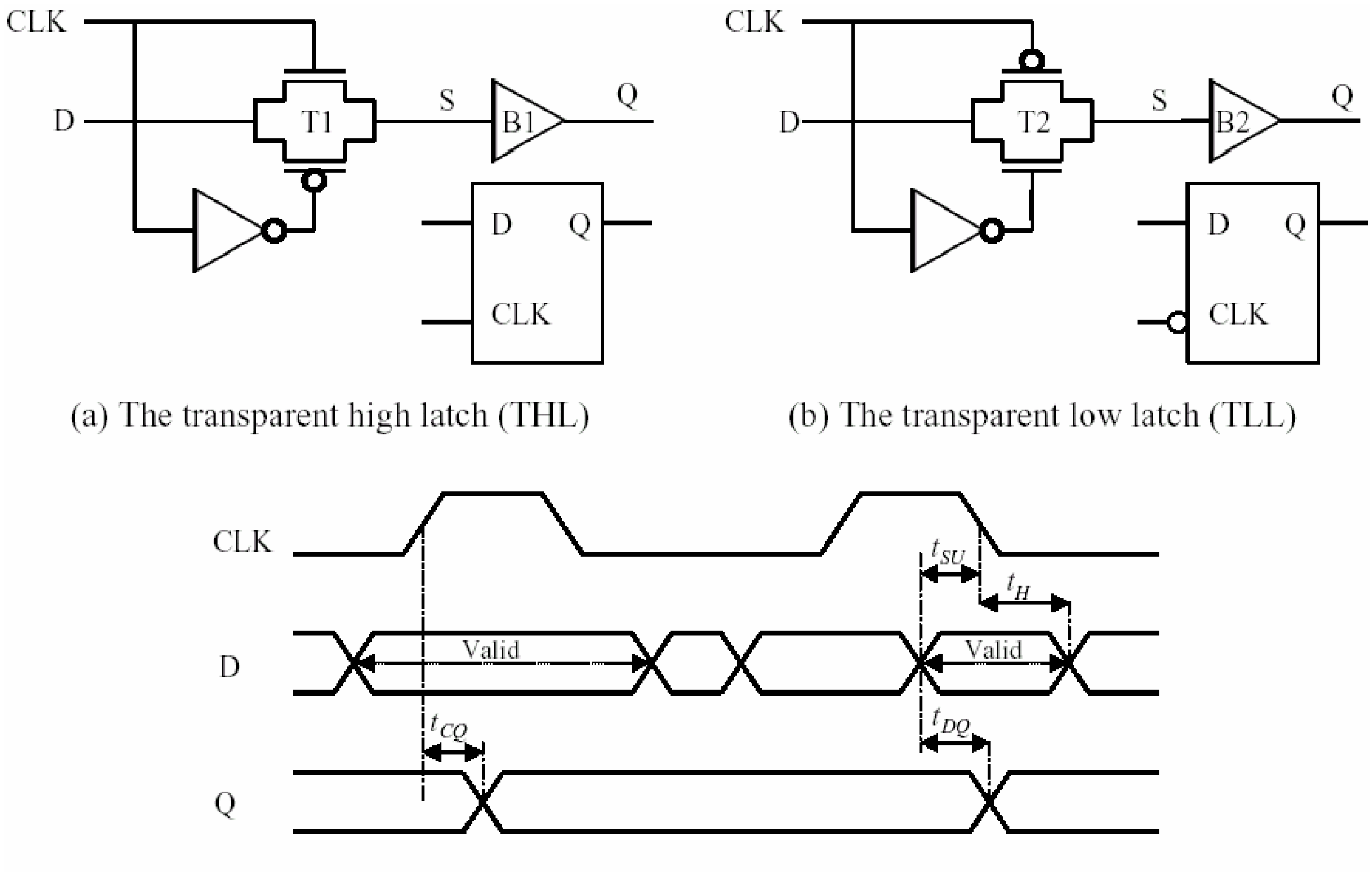
Domino Nor16 (zero-detect)



Flip-flops/latches/state elements

- Flip-flops occupy a special place in conventional digital design
  - Always Dynamic Behavior
  - Allow time coherence across large parts of the circuit
  - Preserve data across synchronization boundaries
  - --Inherently asynchrnous design

Level-sensitive latch pair

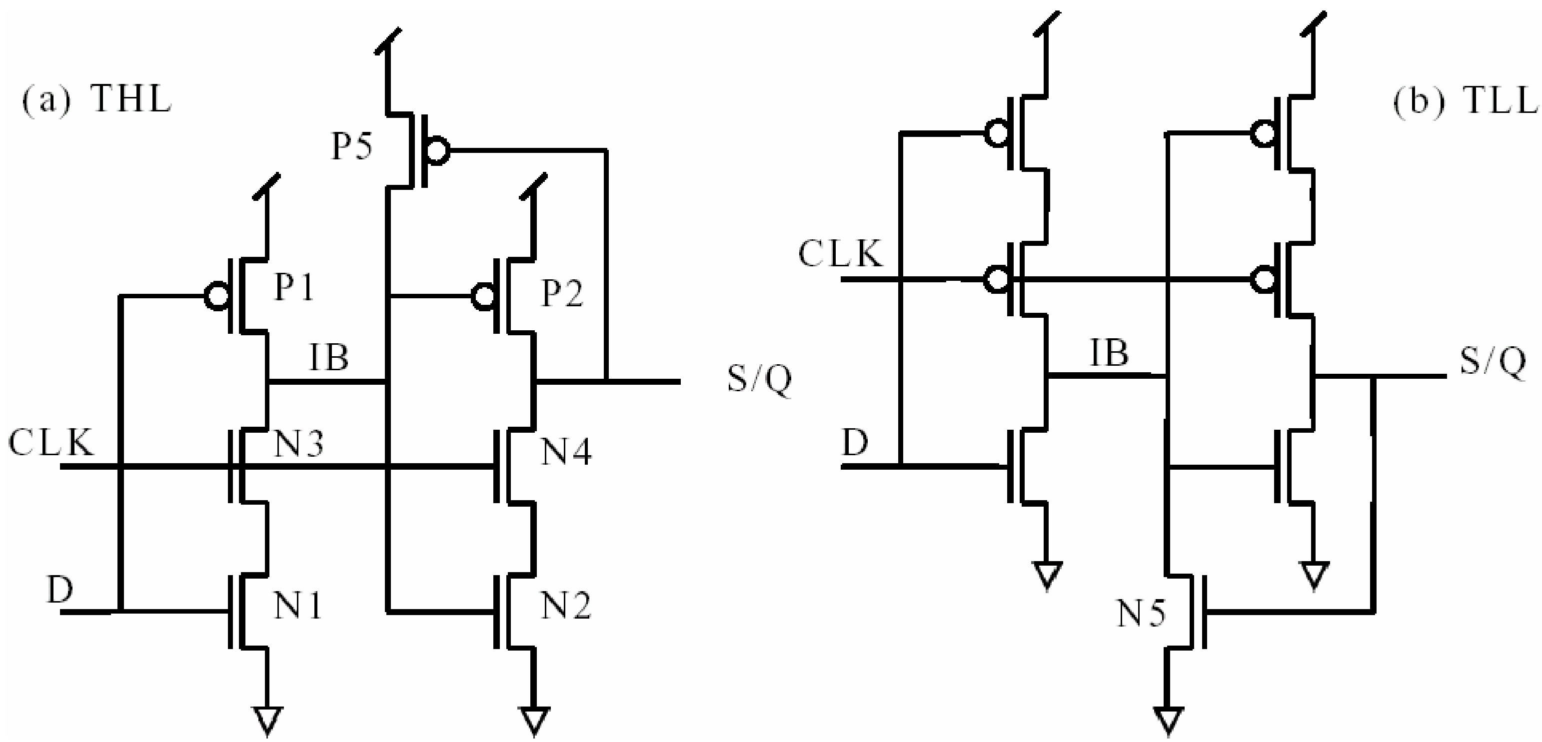


(a) The transparent high latch (THL)

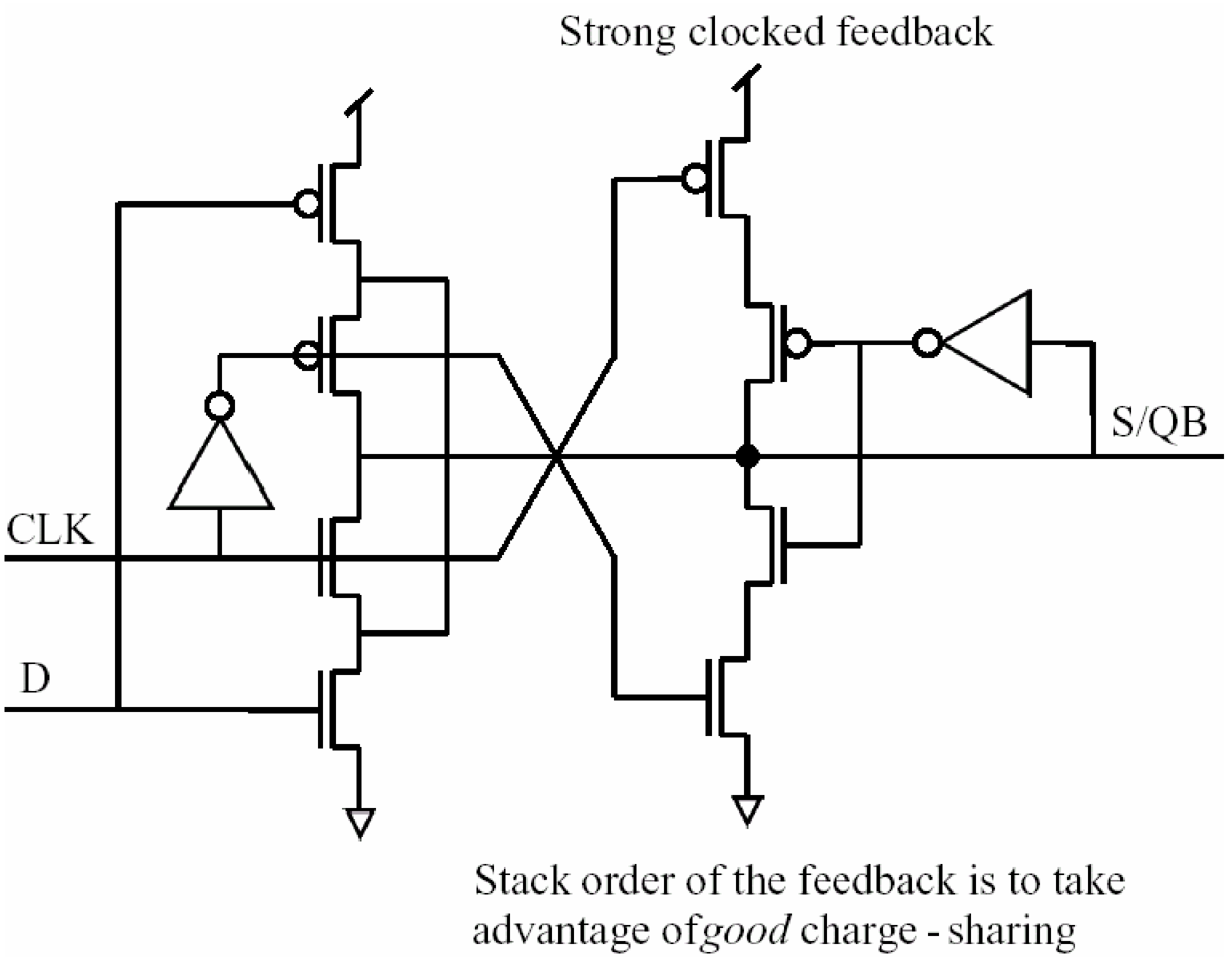
(b) The transparent low latch (TLL)

(c) Timing waveforms for the THL

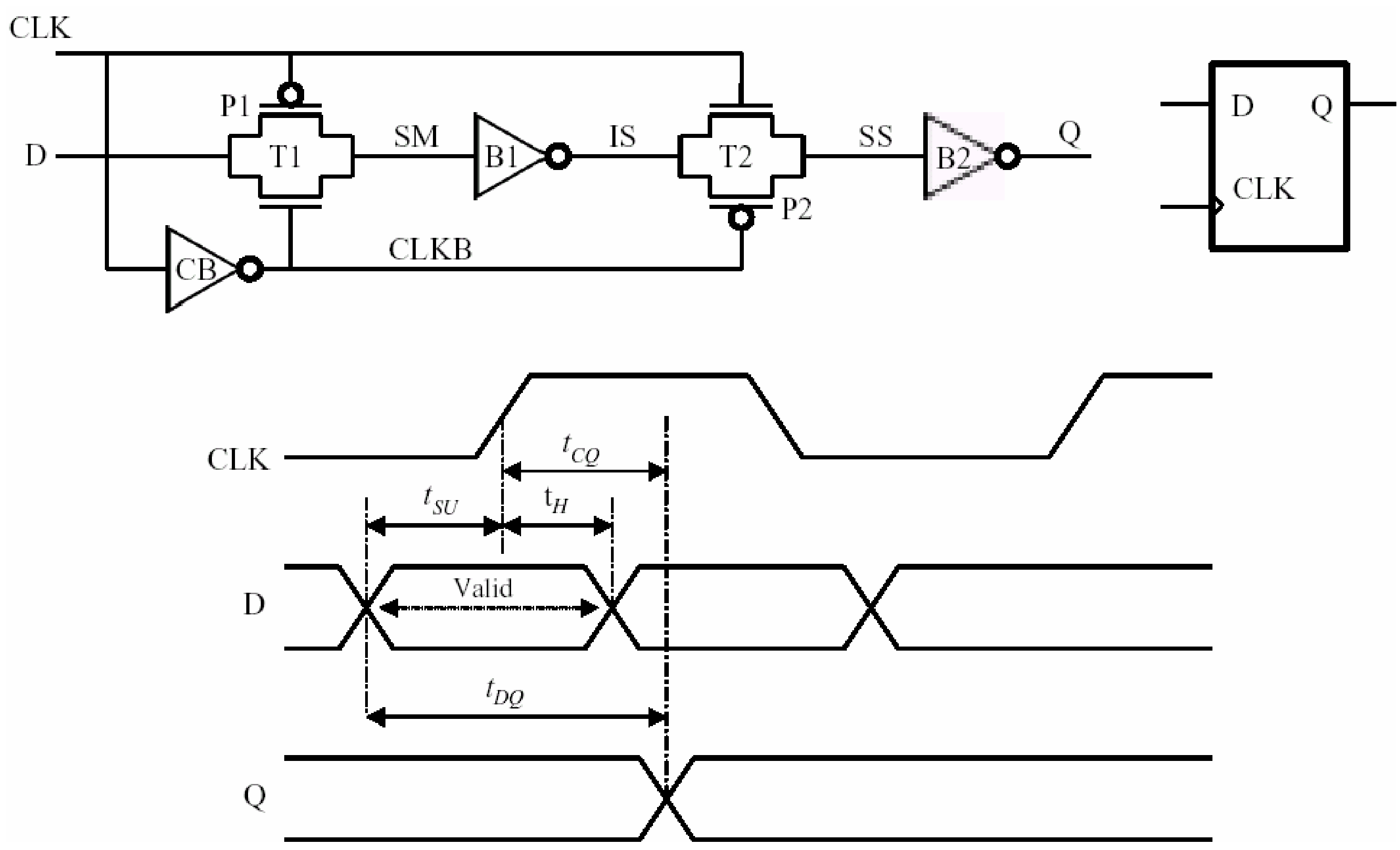
Modified Svensson Latch of 21064



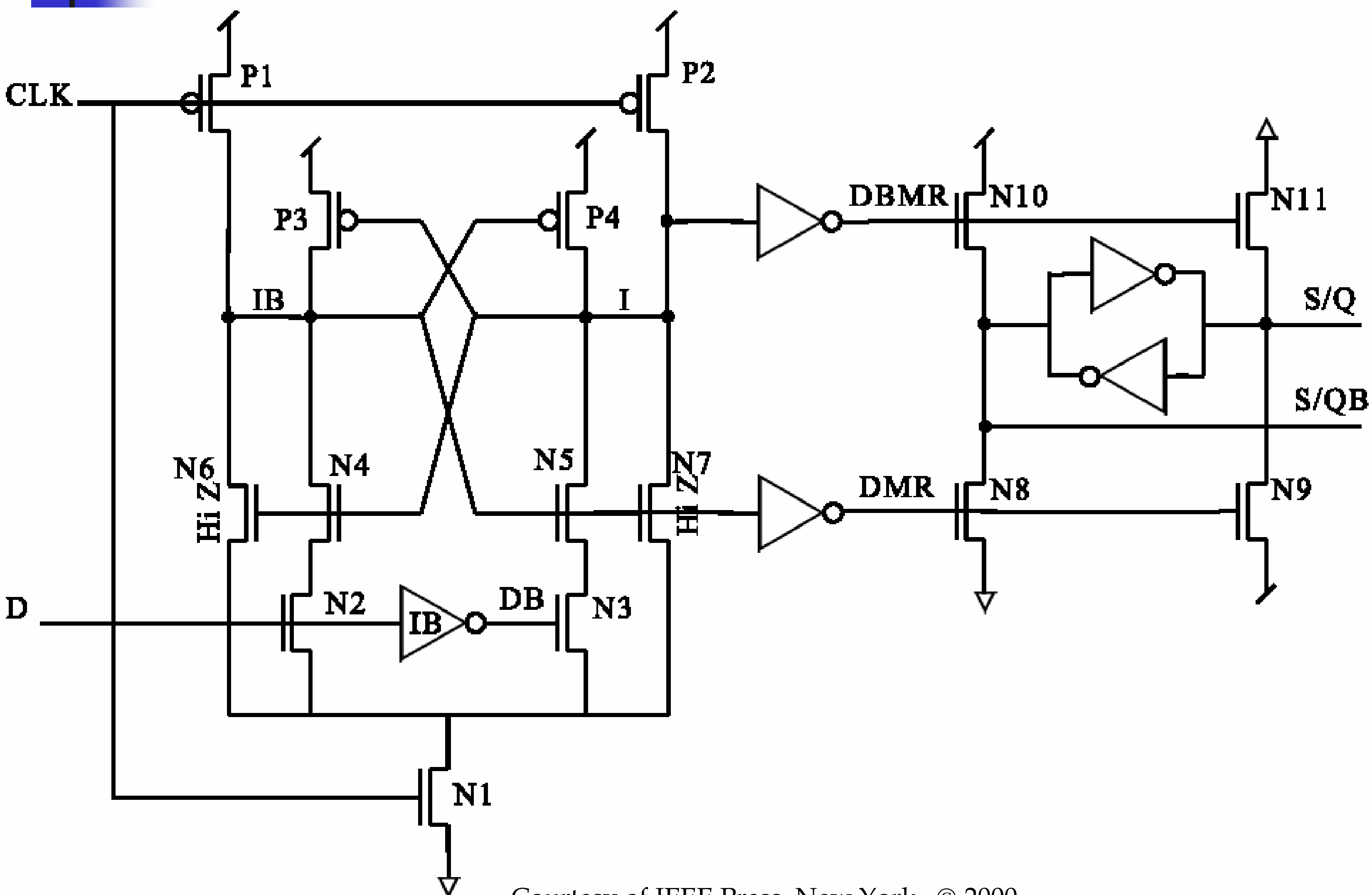
Tri-state based static latch



Master-slave (Dynamic) FF



Sense Amplifier-Based Flip-Flop



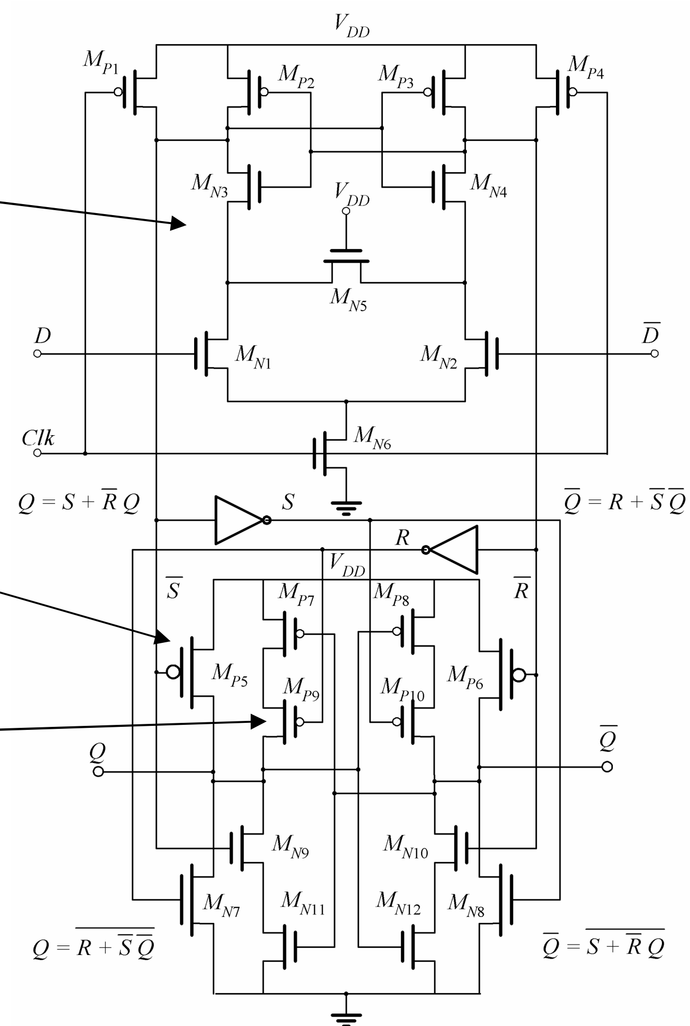
Sense Amplifier-Based Flip-Flop

The first stage is unchanged sense amplifier

Second stage is sized to provide maximum switching speed

Driver transistors are large

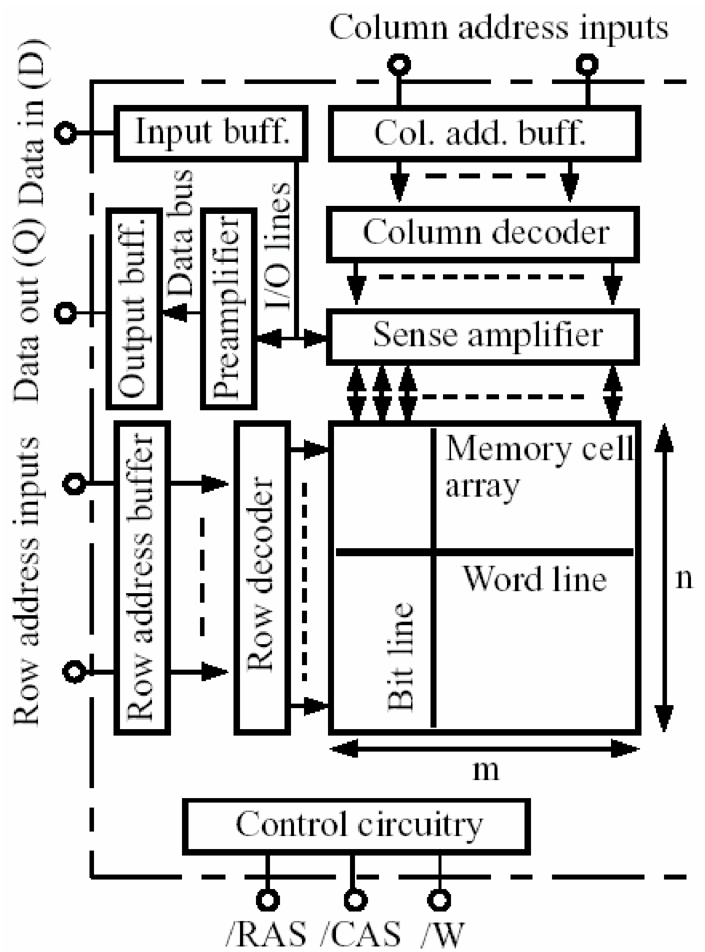
Keeper transistors are small and disengaged during transitions



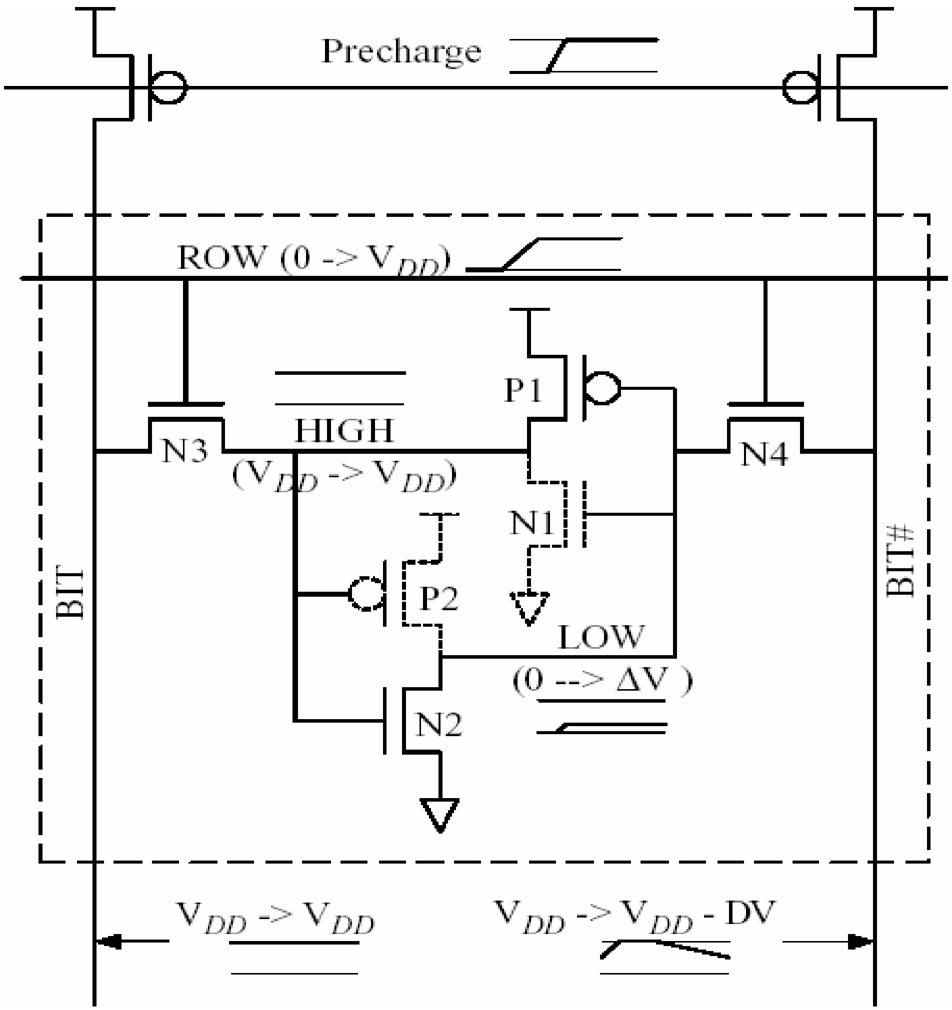
On-chip Memory

- Typically largest fraction of chip area
- Nearly always topologically organized (low Rent parameter <0.6)
- Simple wire/area planning rules

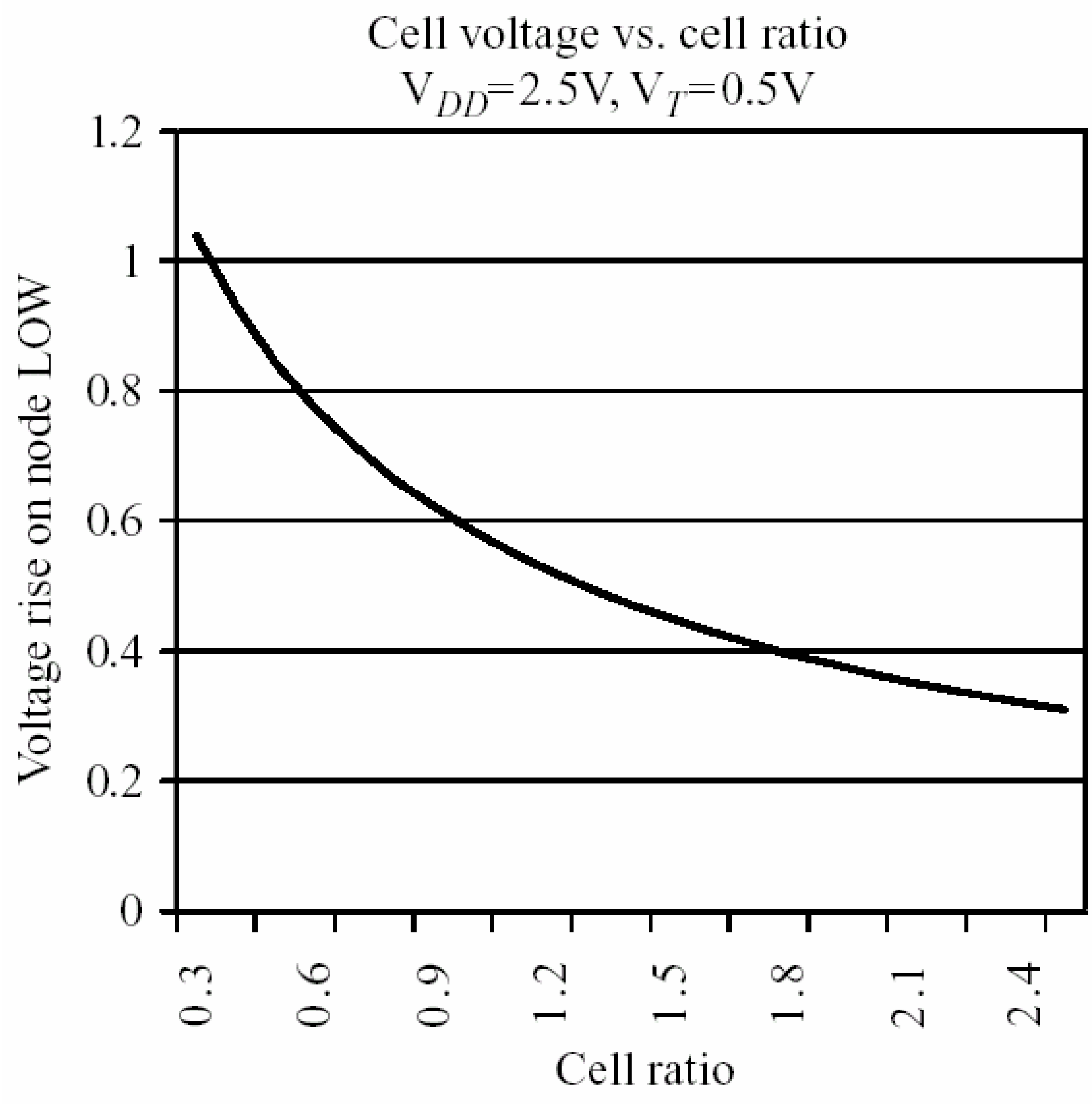
Generic memory block diagram



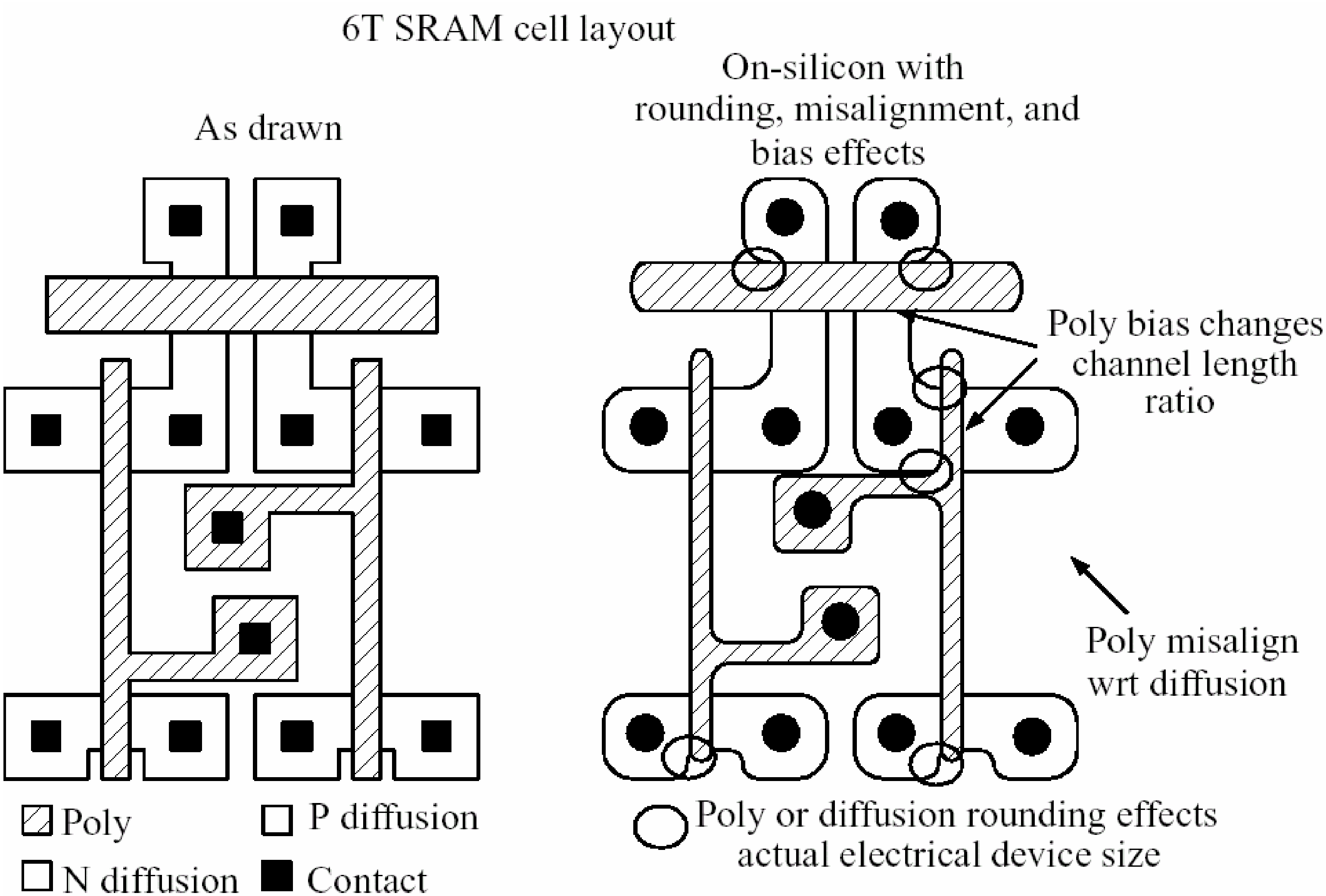
SRAM read operation



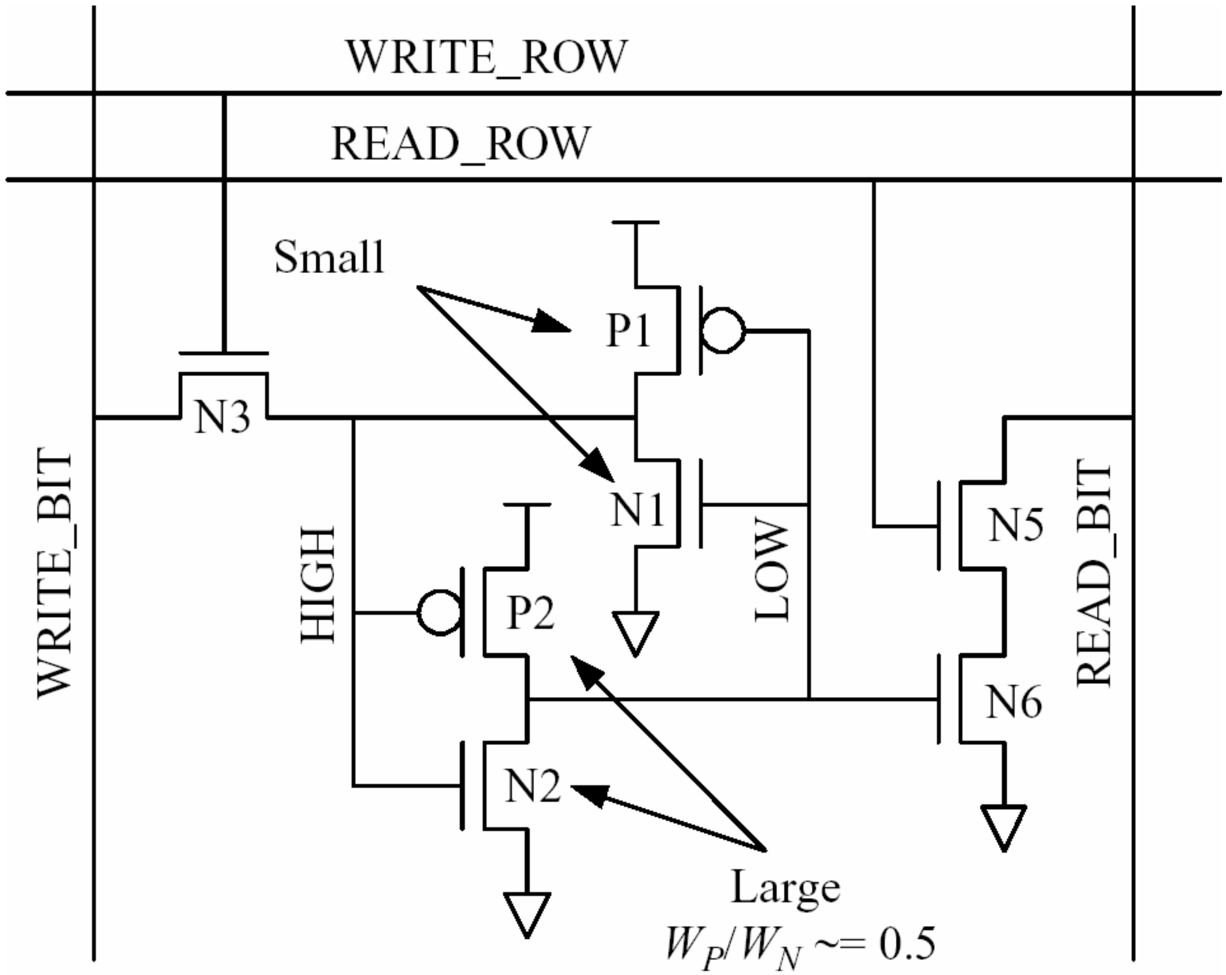
SRAM cell sized to avoid read-disturbance



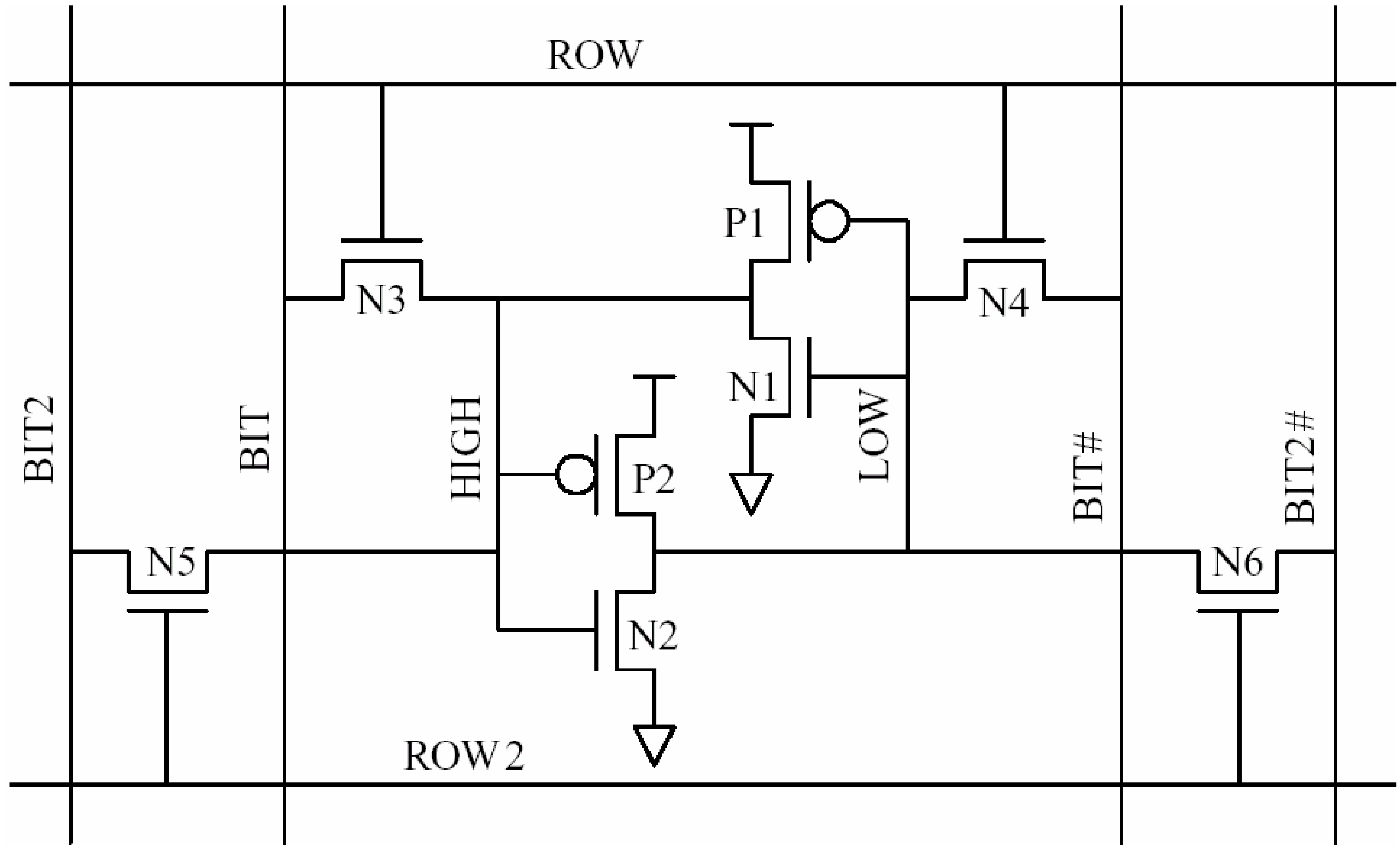
Realistic layout issues in SRAM cell



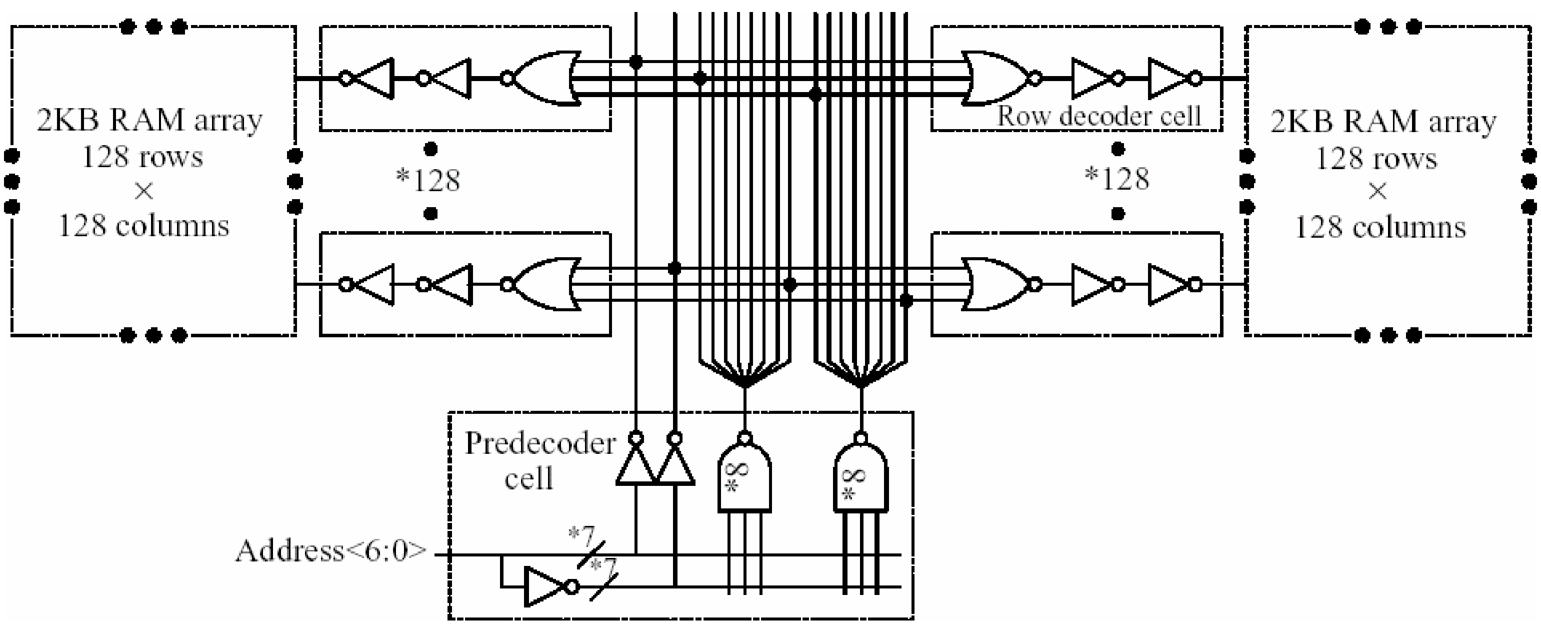
Asymmetric Read/Write Ports



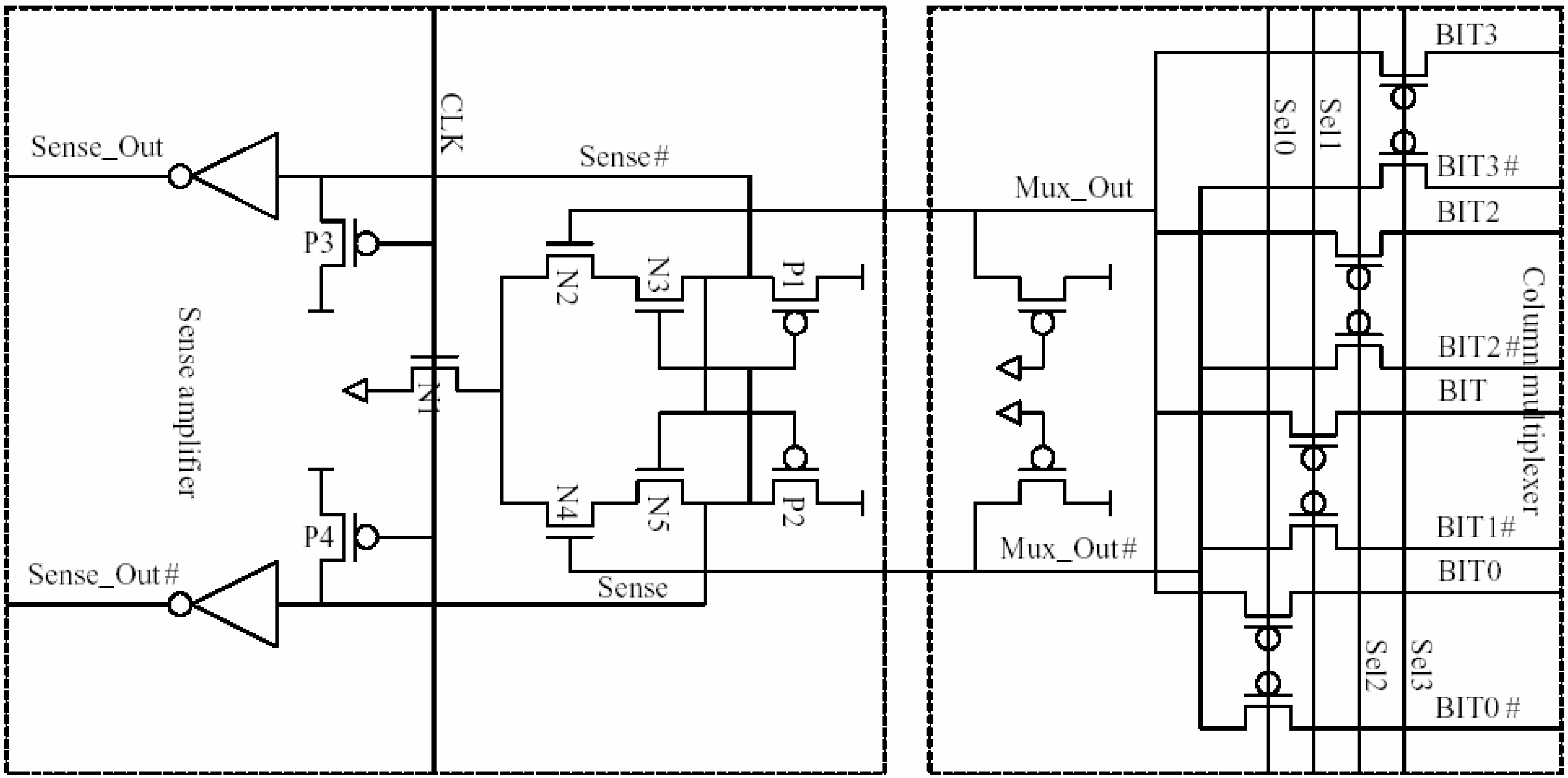
Multi-porting



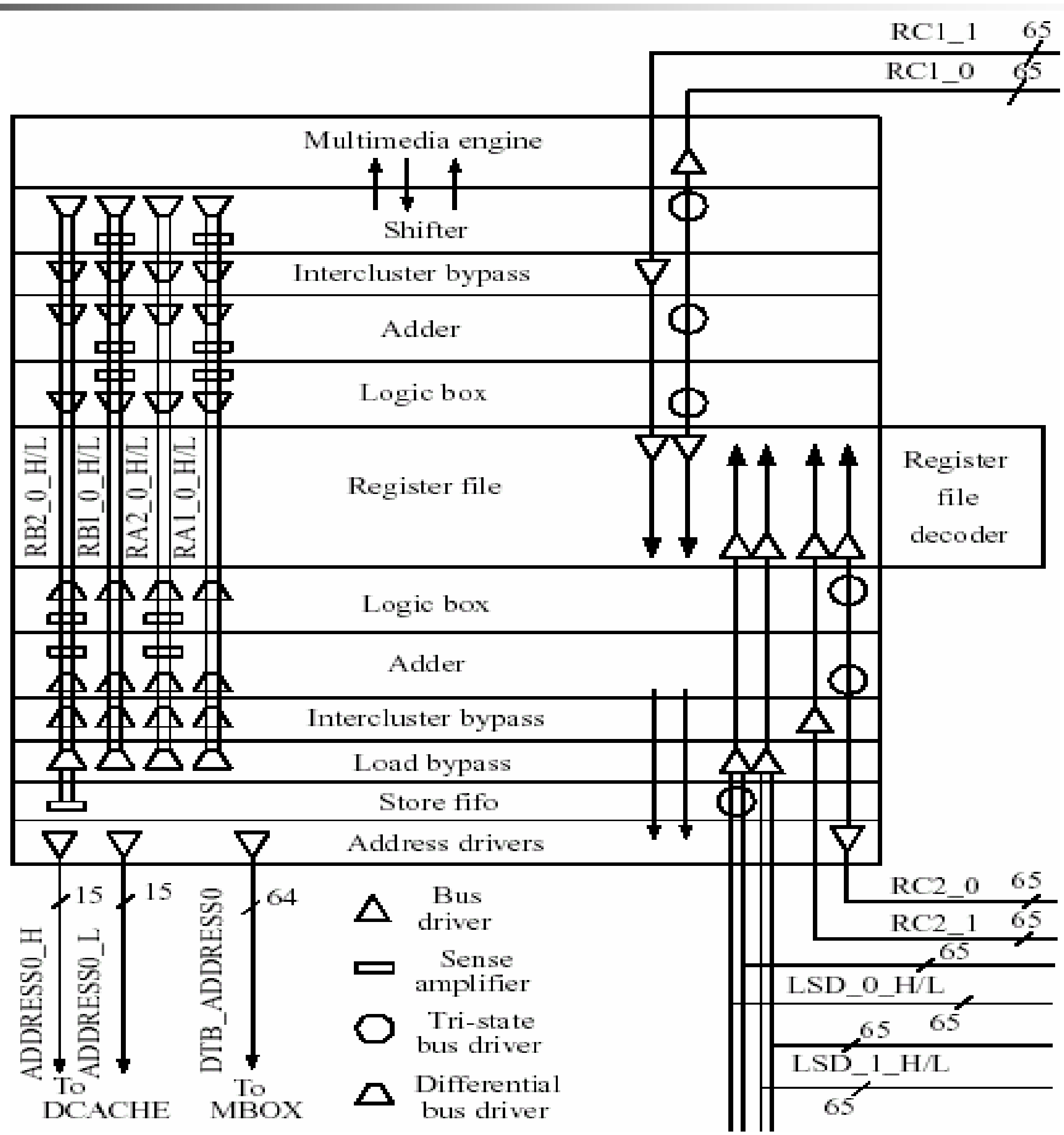
Split Row Decoder



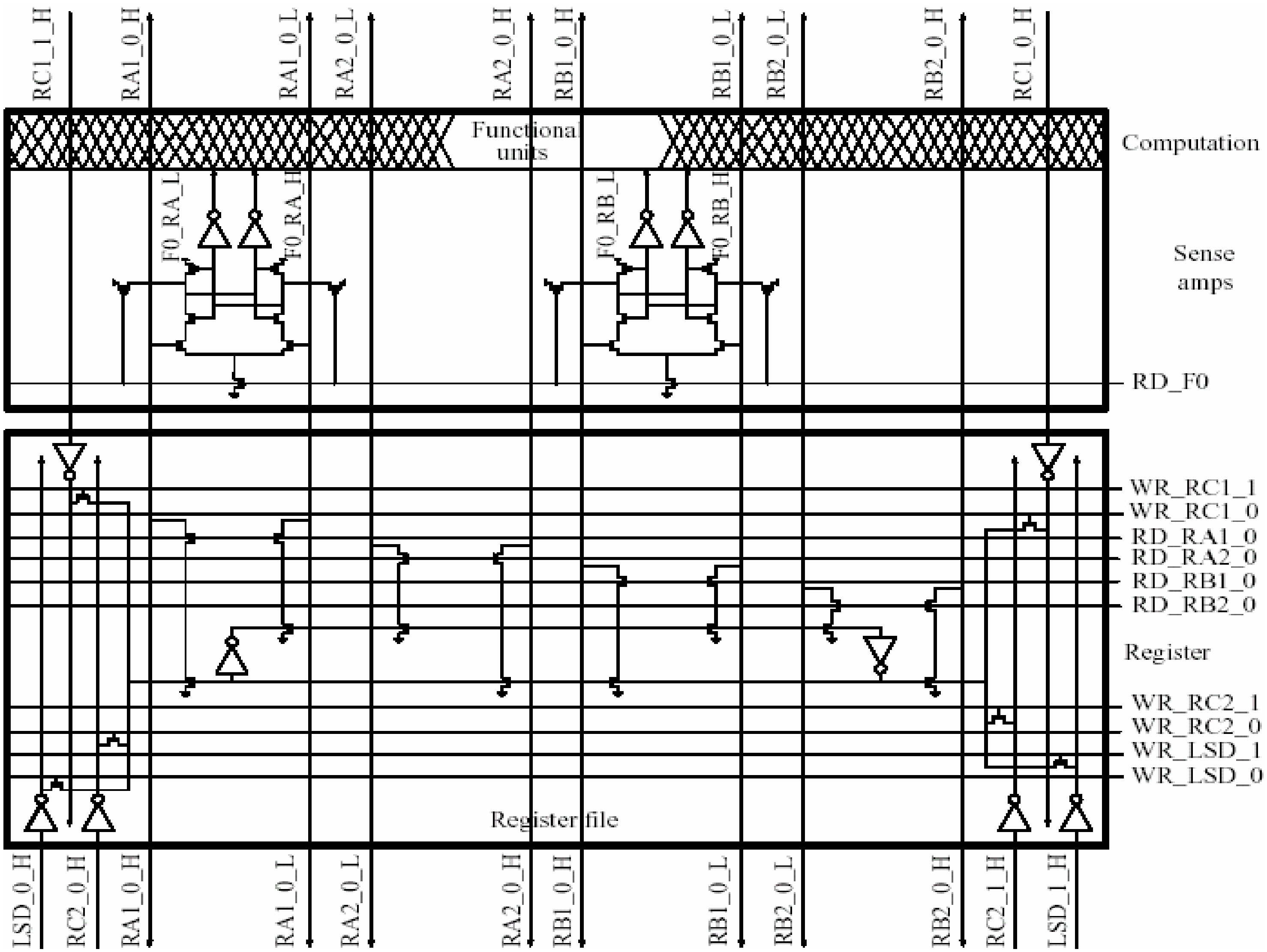
Column mux and sense-amp



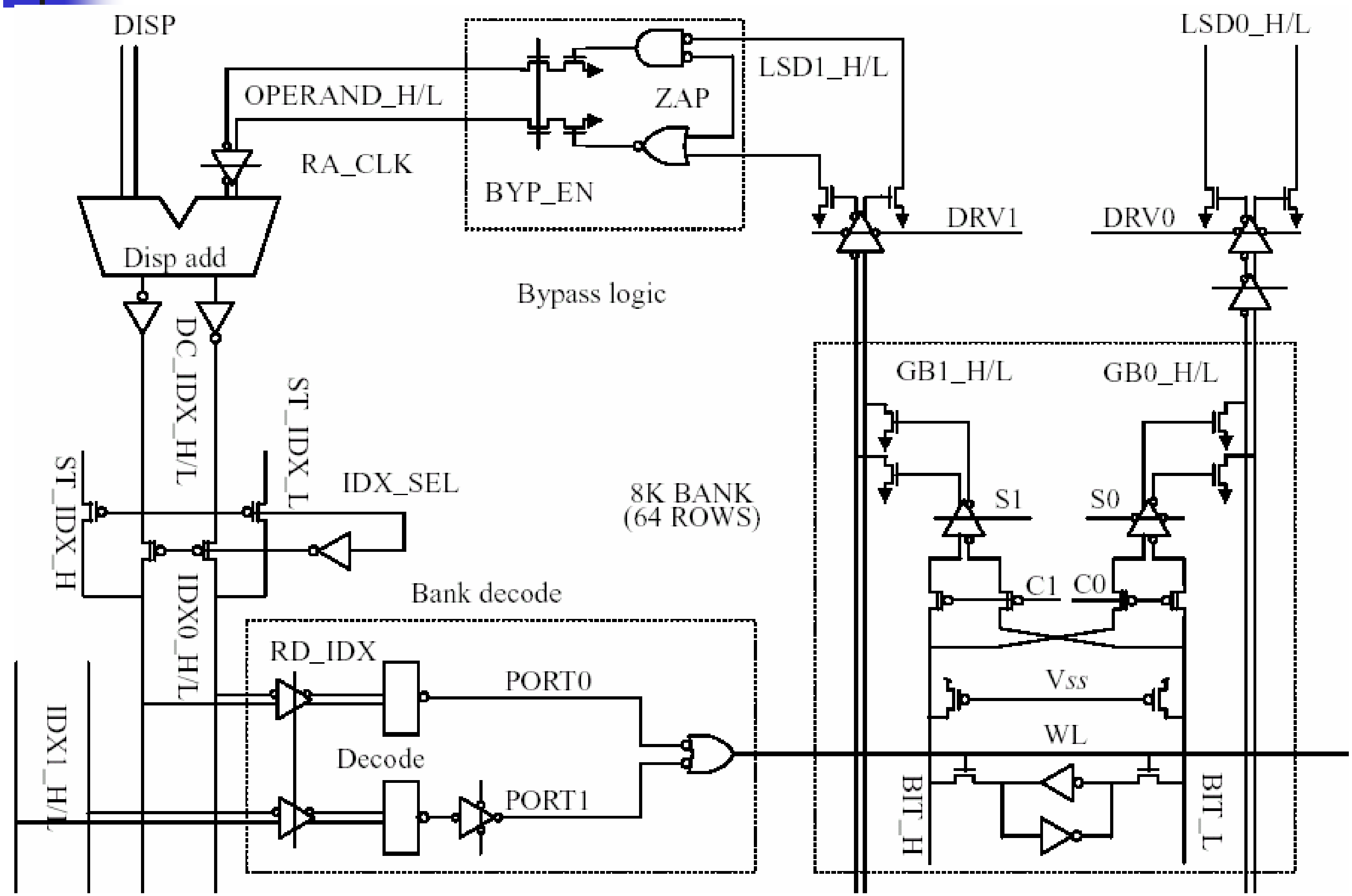
21264 Integer Unit floorplan



21264 Integer Register File cells



21264 L1 Dcache



21164 L2 Cache