Charge Recycling in MTCMOS Circuits: Concept and Analysis

Ehsan Pakbaznia, Farzan Fallah+
Massoud Pedram

University of Southern California
+ Fujitsu Labs of America

IEEE SSCS DL Series
Seoul, South Korea
10/26/2006

Realities

- Power has emerged as the #1 limiter of design performance beyond the 65nm generation.
- Dynamic and static power dissipation limit achievable performance due to fixed caps on chip or system cooling capacity.
- Power related signal integrity issues (IR drop, L di/dt noise) have become major sources of design re-spins.

Transistors (and silicon) are free.
Power is the only real limiter.
Optimizing for frequency and/or area may achieve neither.

Pat Gelsinger, Senior Vice President & CTO, Intel

Industry Views (Intel)

EETIMES

THE INDUSTRY SOURCE FOR ENGINEERS & TECHNICAL MANAGERS WORLDWIDE

SUBSCRIBE | NEW

Tektronix presents Net Seminar series Winter 2003

Grove calls leakage chip designers' top problem

By Ron Wilson and David Lammers
EE Times
December 13, 2002 (4:20 p.m. EST)

PRINT THIS STORY | SEND AS EMAIL

SAN FRANCISCO — Power consumption, particularly off-state current leakage, is the major technical problem facing the semiconductor industry, said Andrew Grove, chairman of the board at Intel Corp.

In a luncheon address at the International Electron Devices Meeting (IEDM) here, Grove said that as chip densities increase to a billion transistors or more, power is "becoming a limiter of integration."

Industry Views

BusinessWeek online

BW HOME | BW MAGAZINE | TOP NEWS | INVESTING | GLOBAL BIZ | TECHN

OCTOBER 4, 2004 · Editions: N. America | Europe | Asia | Edition Prefe

Full Table of Contents
Cover Story
International Cover Story
Up Front
The Great Innovators
Readers Report
Corrections &
Clarifications
Books
Technology & You
Economic Viewpoint
Business Outlook

MLB.com™ knows

TECHNOLOGY & YOU

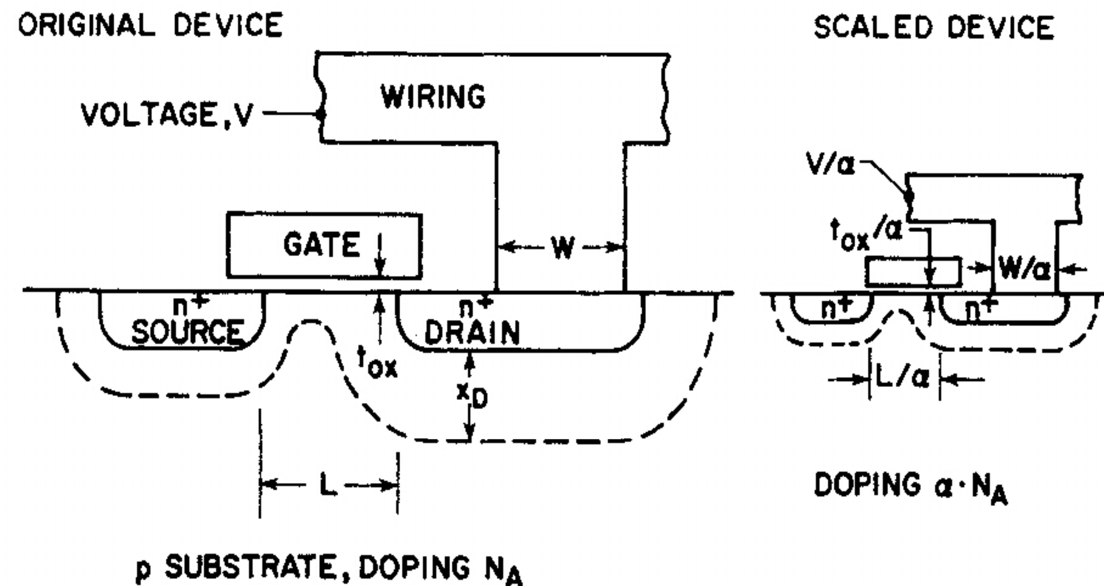
Those Superfast Chips: Too Darn Hot

Without cooler new processors, PC makers could hit a speed bump

Intel's (INTC) recent announcement that it plans to produce new "dual-core" processors that amount to two Pentiums on a single chip drew attention mainly from hard-core techies. But it was an admission that the company's strategy for making PCs ever cheaper and faster has hit a wall: The chips are simply getting too hot. Further progress will require new technologies.

Find
help

Constant-Field MOSFET Scaling

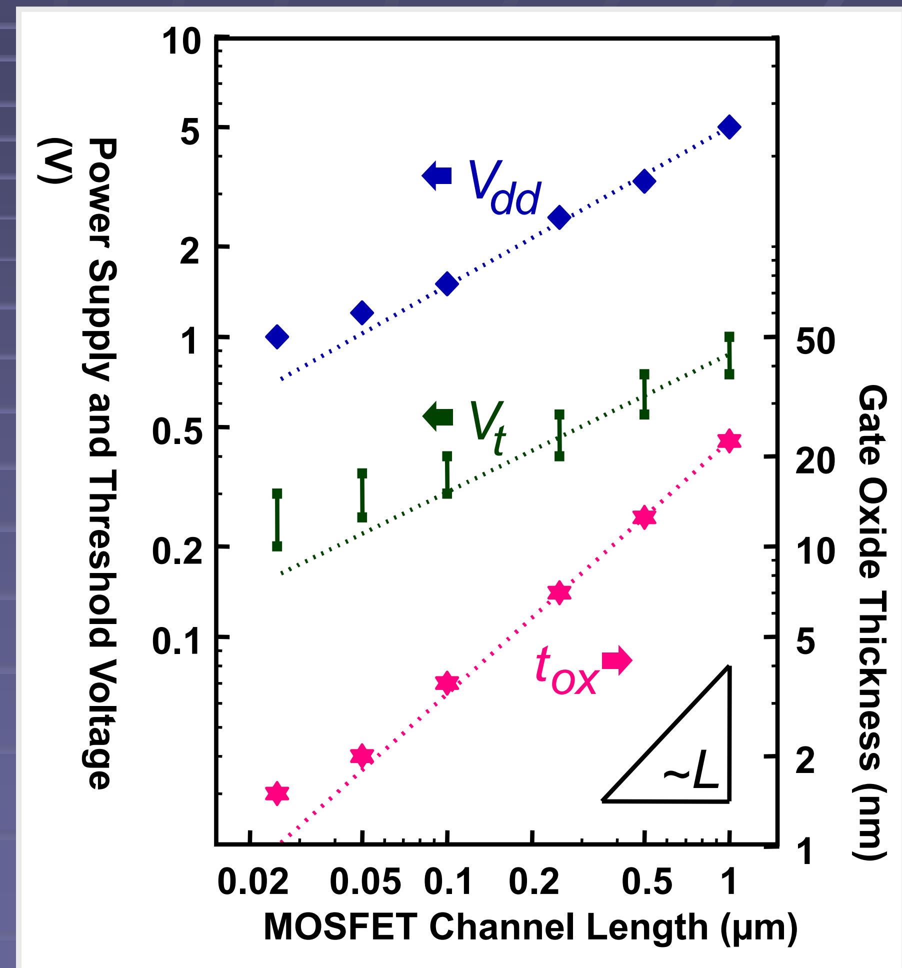


Source: B. Davari, IBM, 1999

- L, W, t_{ox}, x_D, V_{DD}, V_T, C, I, and τ scale by $1/\alpha$.
- Area, power dissipation, and charges scale by $1/\alpha$.
- Power dissipation and charges per unit area do not scale.

V_{dd}, V_{th} and t_{ox} Scaling

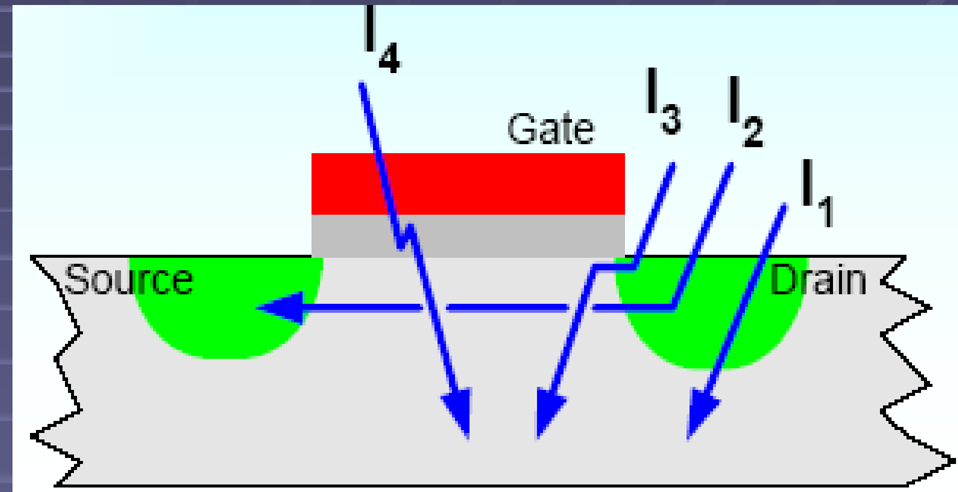
- V_{dd} scaling needed to reduce power and maintain device reliability
  - V_{th} scaling needed to maintain switching speeds
  - t_{ox} scaling needed to maintain the current drive and keep V_{th} variations under control when dealing with short-channel effects.
- V_{th} does not scale much since the inverse subthreshold slope, which represents transistor turn-off rate, is dominated by temperature, not V_{th} or V_{dd}.

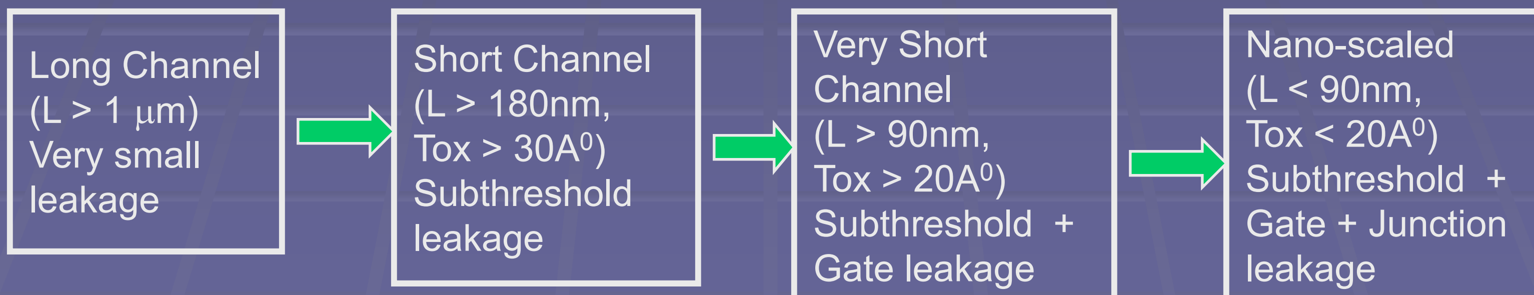


Source: Y Taur, 2002

Leakage Components in CMOS

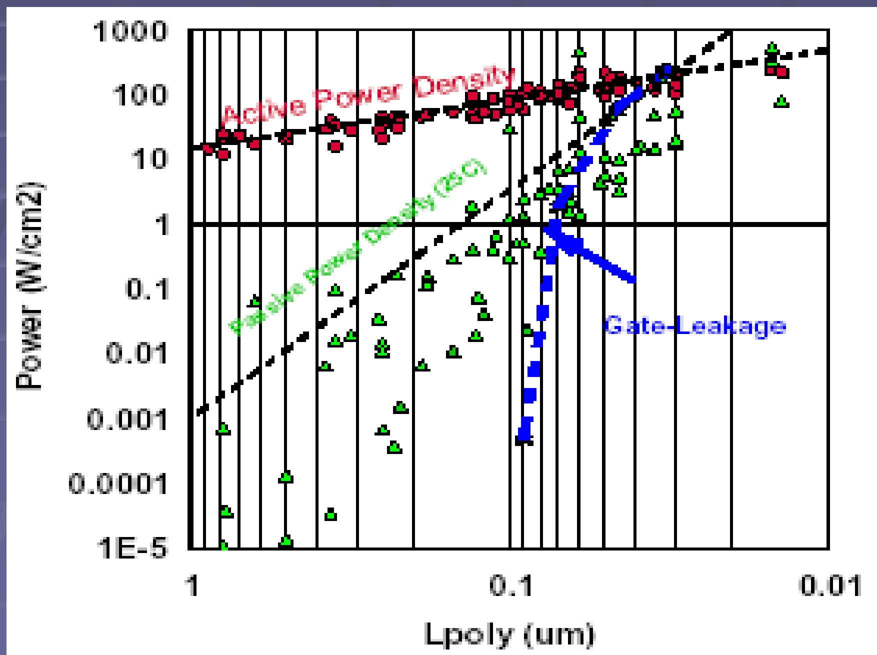
- I_1 Diode reverse bias current
- I_2 Subthreshold current
- I_3 Gate induced drain leakage
- I_4 Gate oxide tunneling



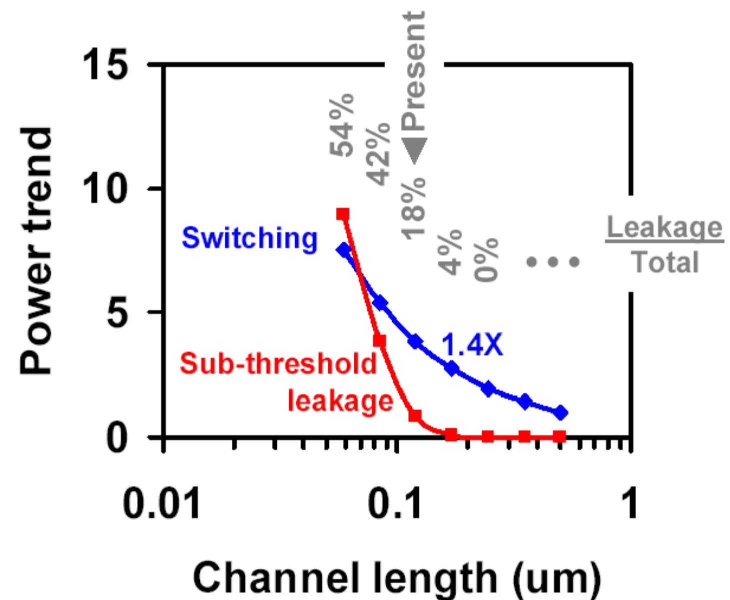


Leakage vs. Total Power

- A significant part of total power at 90nm and below
  - Sub-threshold leakage is increasing due to V_{th} scaling.
  - Gate leakage is increasing due to gate oxide scaling.
- Leakage in active mode is a major issue.



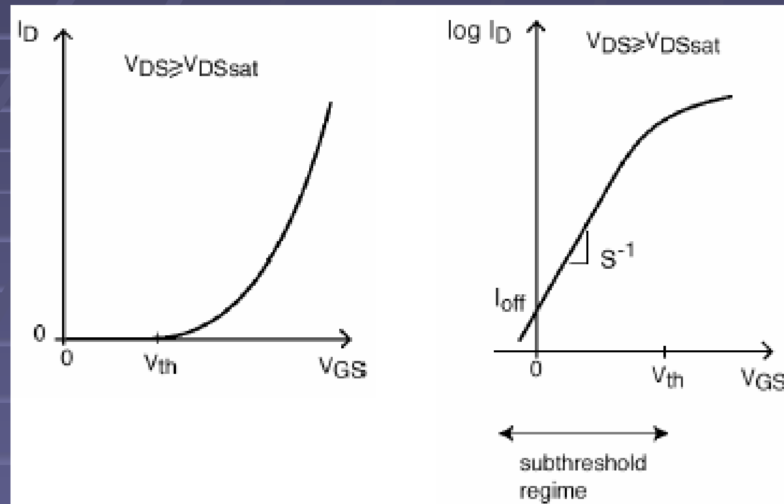
Source: Puri, et al 2003

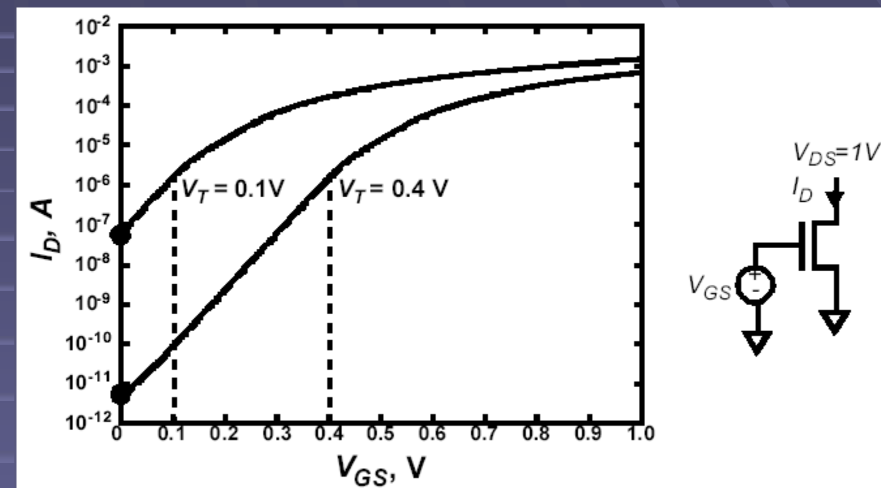


Source: Chandrakasan, et al 2002

Subthreshold Leakage Current

Transfer characteristics of MOSFET for V_{GS} near V_{th}:

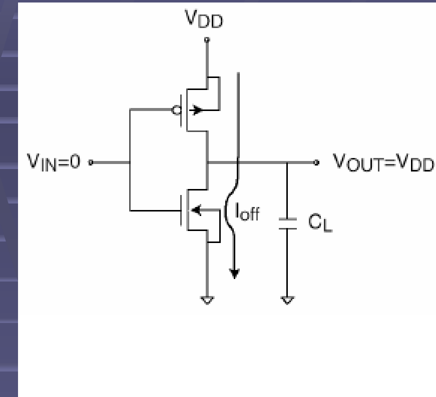




$$I_{sub} = \frac{W}{L}\mu_e v_T^2 C_{sth}\, e^{\frac{V_{GS}-V_{th}+\eta V_{DS}}{n v_T}}\left(1-e^{\frac{-V_{DS}}{v_T}}\right) \propto e^{\frac{V_{GS}-V_{th}+\eta V_{DS}}{n v_T}} = 10^{\frac{V_{GS}-V_{th}+\eta V_{DS}}{S}}$$

- The inverse subthreshold slope, S, is equal to the voltage required to increase I_D by 10X, i.e., $S = \frac{nkT}{q}\ln 10$
  - If n = 1, S = 60 mV/dec at 300 K
  - We want S to be small to shut off the MOSFET quickly
  - In well designed devices, S is 70 - 90 mV/dec at 300 K.

Modeling I_{sub} and I_{off} Currents



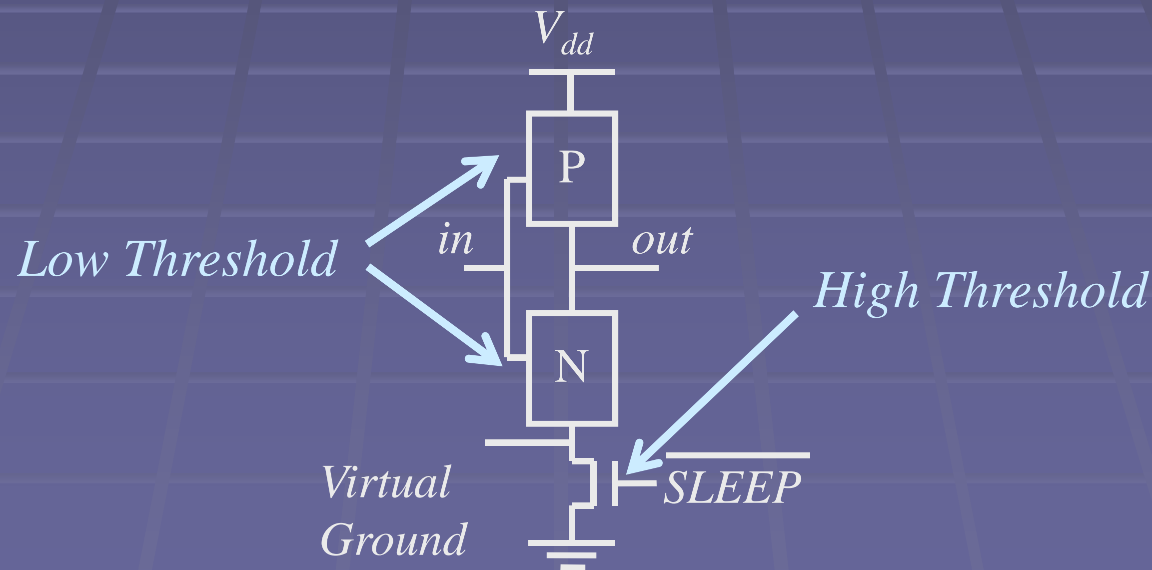
- Increases exponentially with reduction in V_{th}.
- Modulation of V_{th} in a short channel transistor.
  - $L \downarrow \Rightarrow V_{th} \downarrow$: "$V_{th}$ Rolloff"
  - $V_{DS} \uparrow \Rightarrow V_{th} \downarrow$: "Drain Induced Barrier Lowering"
  - $V_{SB} \uparrow \Rightarrow V_{th} \uparrow$: "Body Effect".
- $V_{DS} = 0 \quad \Rightarrow I_{sub} = 0$
- Long-channel device w/ $V_{DS} > 3nv_T \Rightarrow I_{sub} = \frac{W}{L} \mu_e v_T^2 C_{sth} e^{\frac{V_{GS}-V_{th}}{nv_T}}$
- With $n = 1 + \frac{\gamma}{2\sqrt{2\Phi_f}} = 1 + \frac{C_{sth}}{C_{ox}} = 1 + \frac{C_{dep} + C_{it}}{C_{ox}}$ $\quad I_{off} = I_{sub}(V_{GS} = 0) = \frac{W}{L} \mu_e v_T^2 C_{sth} e^{-\frac{V_{th}}{nv_T}}$
- Key dependencies of the subthreshold slope:
  - $T_{ox} \downarrow \Rightarrow C_{ox} \uparrow \Rightarrow n \downarrow \Rightarrow$ sharper subthreshold
  - $N_A \uparrow \Rightarrow C_{sth} \uparrow \Rightarrow n \uparrow \Rightarrow$ softer subthreshold
  - $V_{SB} \uparrow \Rightarrow C_{sth} \downarrow \Rightarrow n \downarrow \Rightarrow$ sharper subthreshold
  - $T \uparrow \Rightarrow$ softer subthreshold.

Leakage Reduction Techniques

- Device engineering
- Lowering and/or turning off V_{dd} (voltage islands and power domains)
- Non-minimum channel length transistors
- Dual-V_{th} design
- Transistor stacking
- Body bias control (static and/or adaptive)
- Cooling and/or refrigeration
- MTCMOS (sleep transistors, power gating)

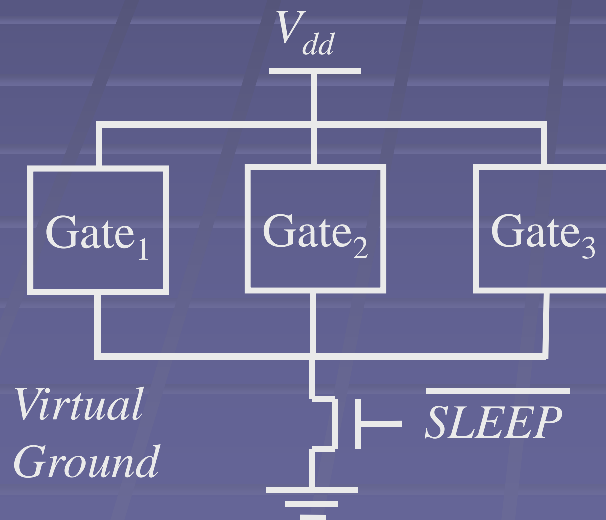
Multi-Threshold CMOS (MTCMOS)

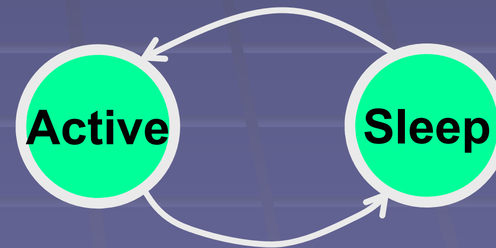
- It is also called power gating, using sleep transistor, etc.
- A high-V_{th} sleep transistor is used to disconnect low-V_{th} transistors from the ground (V_{dd}).

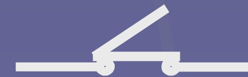


MTCMOS Technology

- MTCMOS technology has proven to be one of the most effective technique for reducing subthreshold leakage in the standby mode of circuit operation.

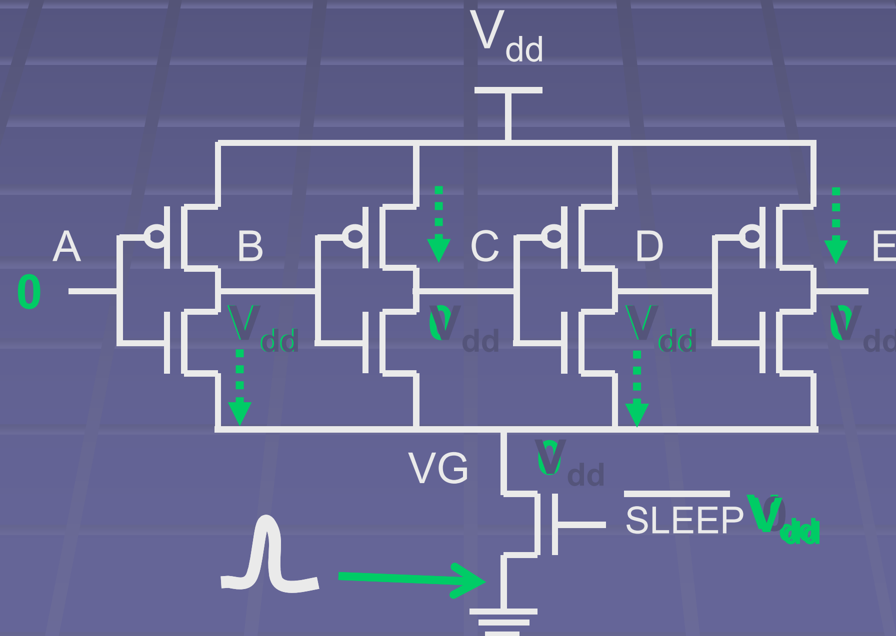






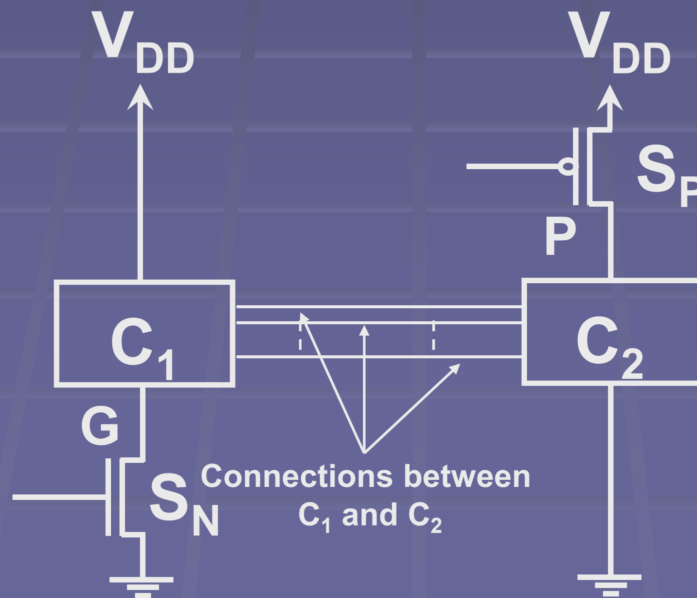
Some Drawbacks of MTCMOS

- There is potential for large ground bounce (noise).
- Energy is wasted when switching between the sleep mode and active mode of circuit operation
  - This means energy cannot be saved unless the sleep time is long enough.

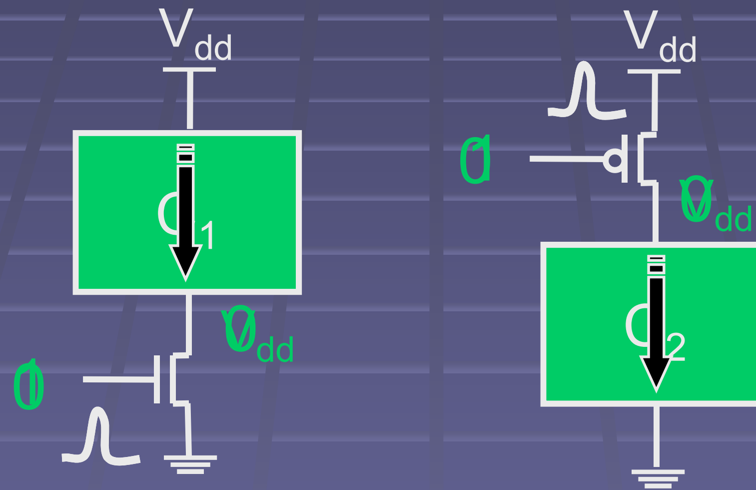


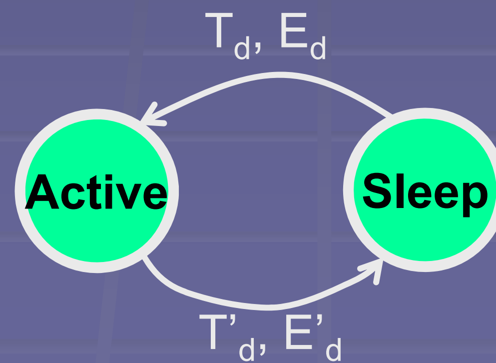
Our Solution: Charge Recycling

- Charge recycling technique uses both NMOS and PMOS sleep transistors.
- Circuit C is divided into 2 sub-circuits:
  - Sub-circuit C_1 is connected to S_N
  - Sub-circuit C_2 is connected to S_P

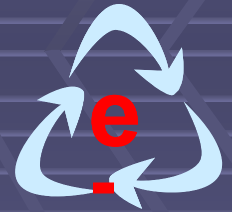


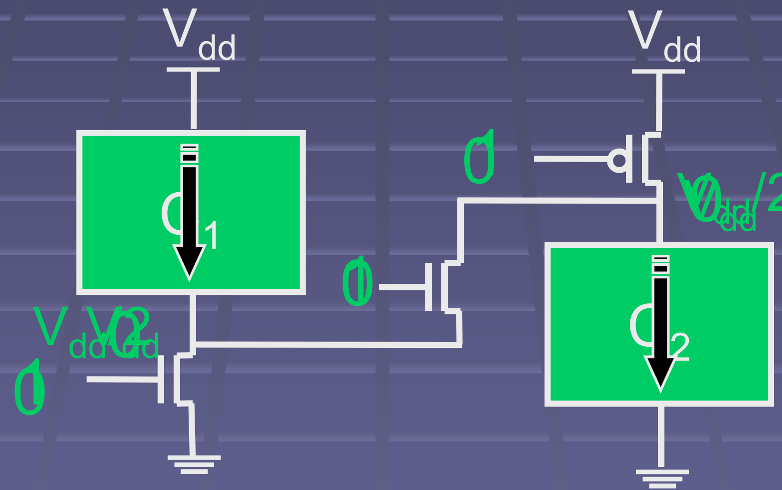
Mode Transitions in This Configuration

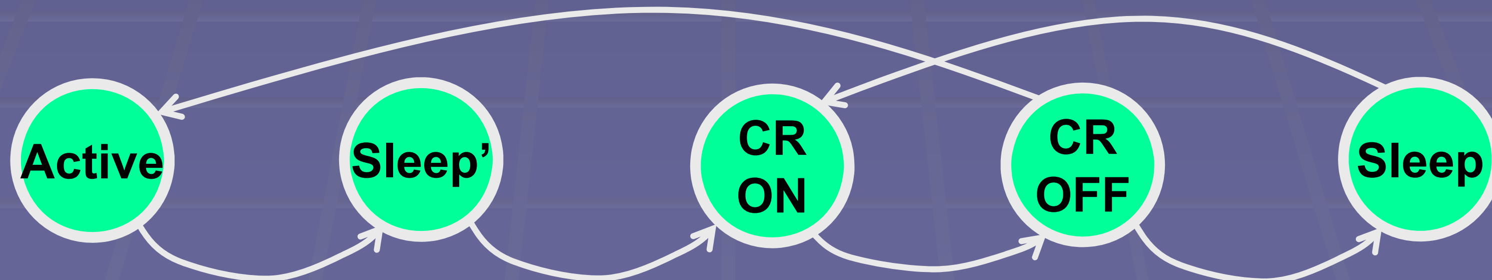




Our Solution: Charge Recycling (CR)

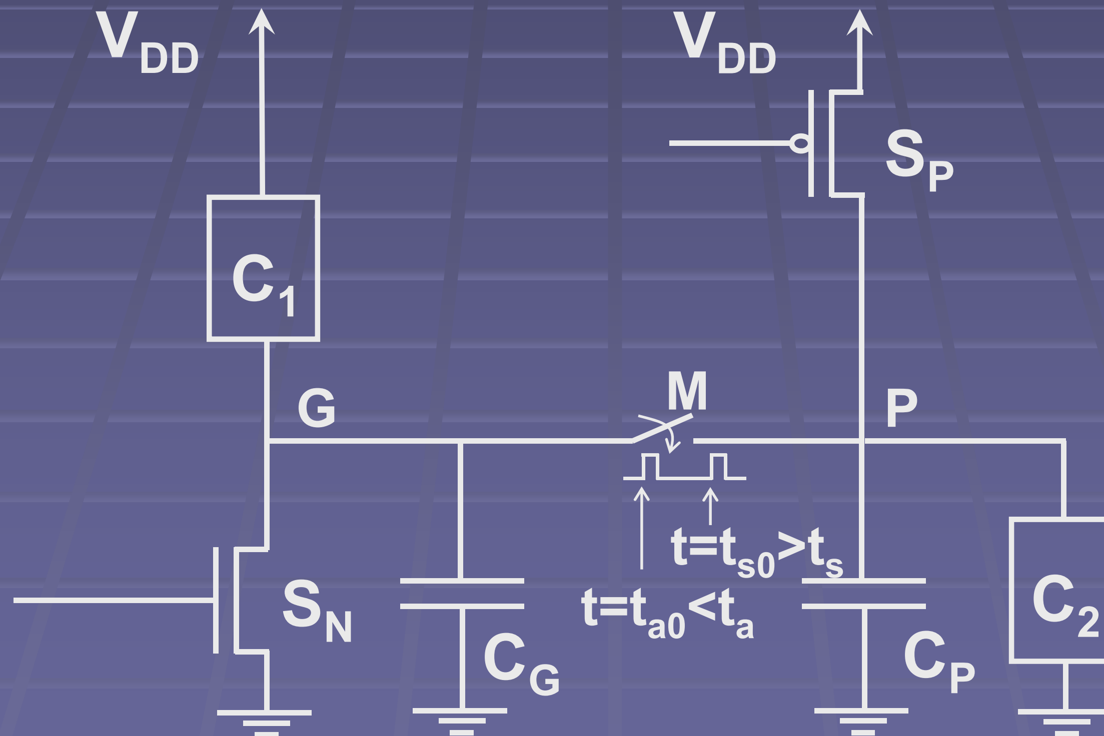






Energy Consumption in CR

- Replacing CR element with an ideal switch, M:



Energy Consumption in CR (cont.)

- Energy consumption during mode transition:

$$E_{sleep-active} = (1-\alpha) C_P V_{DD}^2$$

$$E_{active-sleep} = (1-\beta) C_G V_{DD}^2$$

where we have:

$$\alpha = \frac{C_G}{C_G + C_P} \quad \text{and} \quad \beta = \frac{C_P}{C_G + C_P}$$

$C_P =$ Total Virtual Power Capacitance

$C_G =$ Total Virtual Ground Capacitance

Energy Saving Ratio (ESR) in CR

- Energy consumption in one cycle for the conventional MTCMOS and CR-MTCMOS:

$$E_{conv.} = C_G V_{DD}^2 + C_P V_{DD}^2$$

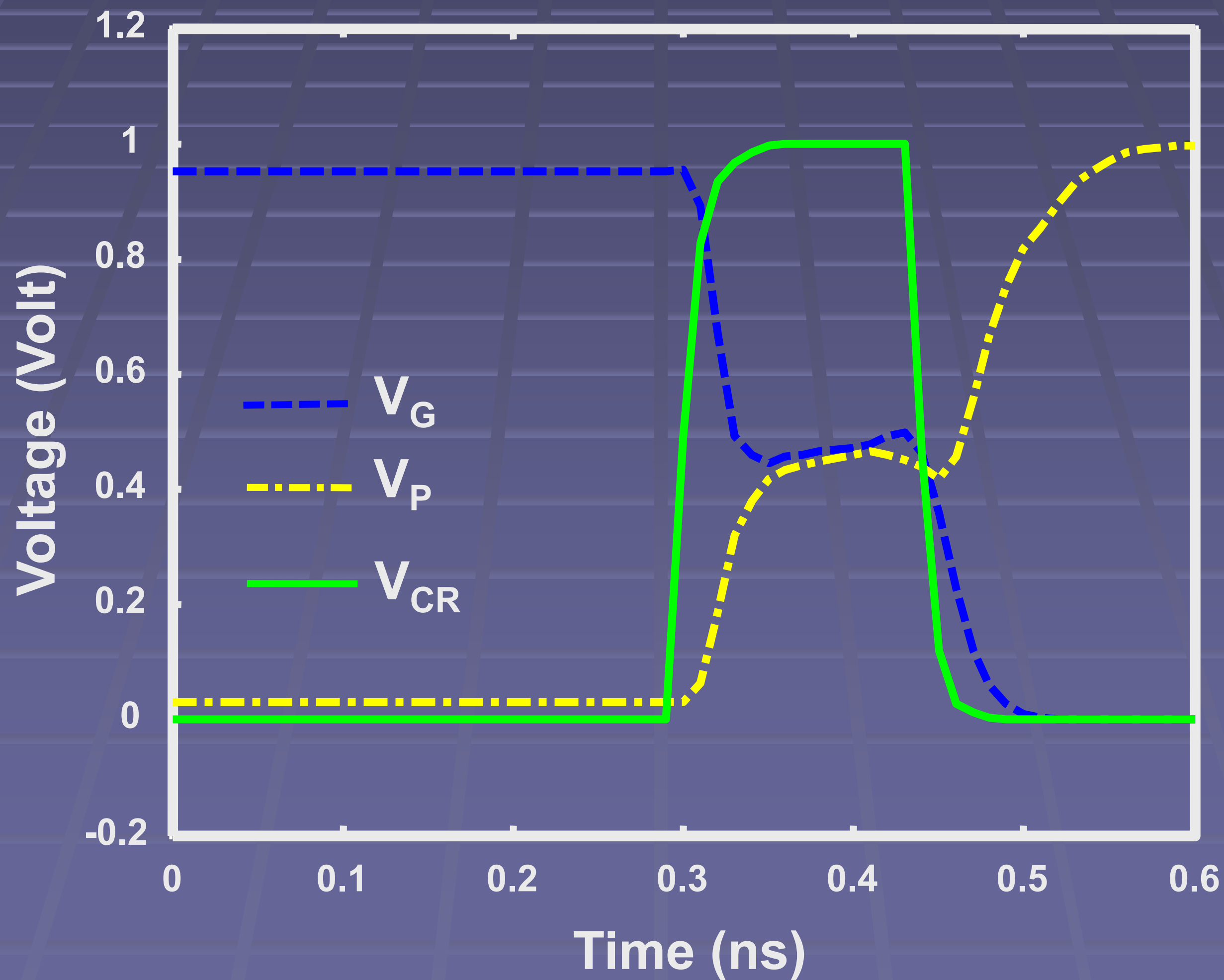
$$E_{CR} = \alpha C_G V_{DD}^2 + \beta C_P V_{DD}^2$$

- The energy saving ratio is:

$$ESR(X) = \frac{E_{total} - E_{cr_total}}{E_{total}} = \frac{2X}{(1+X)^2}$$

- ESR is maximum when $X=1$, i.e., when $C_G = C_P$.

Charge Recycling Operation



Effect of the Threshold Voltage

- Condition for a complete CR (assuming $C_G = C_P$):

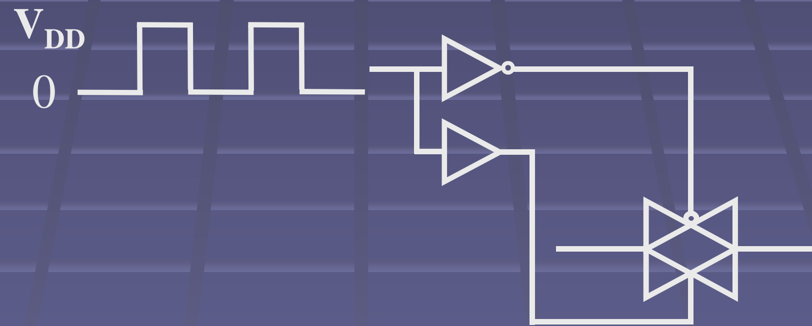
$$Min\left\{V_{tn}, \left|V_{tp}\right|\right\} \le \frac{V_{DD}}{2}$$

  - If $V_{tn} = |V_{tp}|$, then a simple NMOS will be adequate.

- Solution: decrease V_t
  - Trade off: leakage current increases.

Effect of Transistor Sizing

- The larger the transmission gate (TG), the faster the charge recycling operation

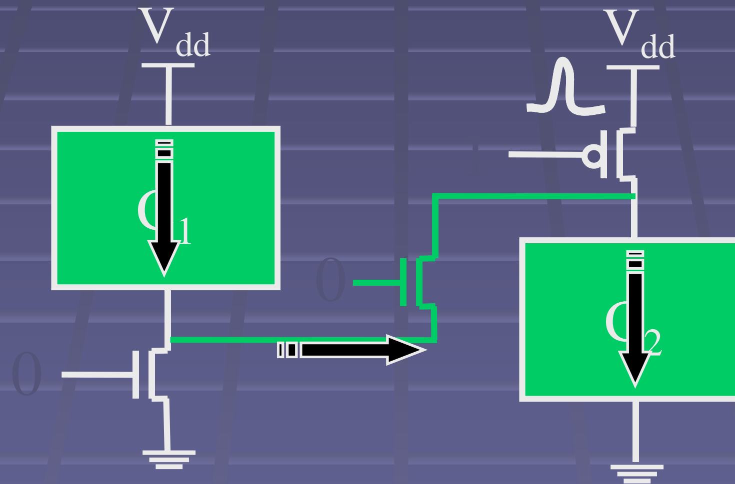


$$\frac{E_{tg-total}}{E_{total}} = \frac{4C_{tg}V_{DD}^2}{(C_G + C_P)V_{DD}^2} = \frac{4C_{tg}}{C_G + C_P}$$

- Trade off: larger TG switching power penalty
  - C_{tg} denotes the input cap of NMOS and PMOS in TG.

Leakage Analysis

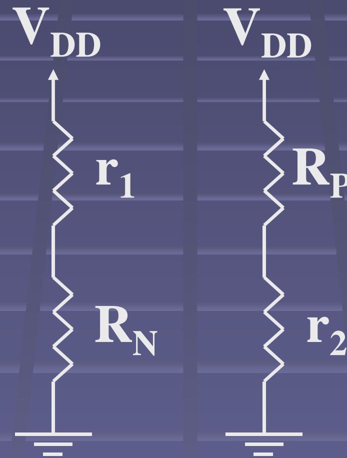
- TG adds a new leakage path:



- Transistors in the TG must be high V_t transistors.

Leakage Paths in Conventional Technique

- The equivalent leakage model for the sleep mode:

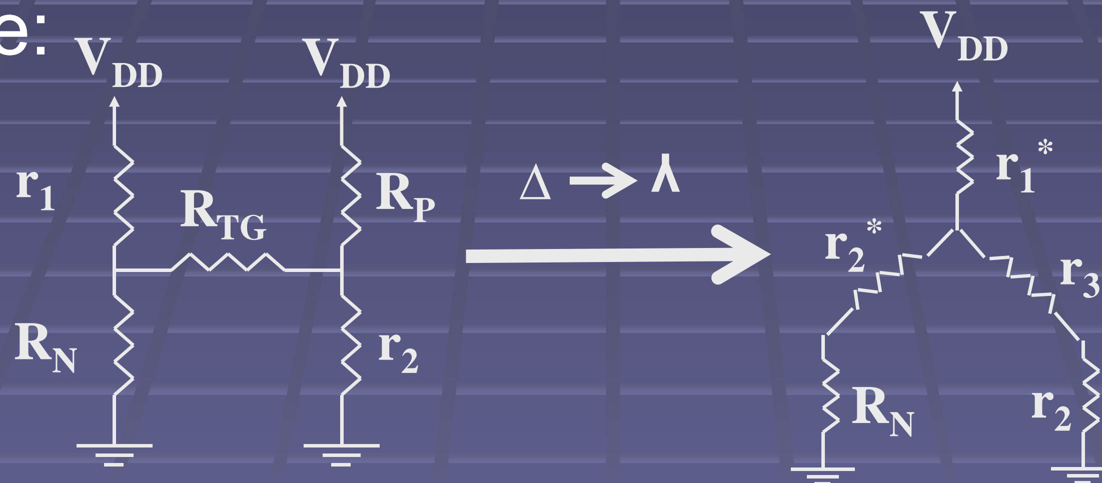


- Leakage is calculated by writing KVL equations (R_N=R_P=R):

$$P_{leakage-conv.} = \frac{2V_{DD}^2}{R}$$

Leakage Paths in CR Technique

- The equivalent leakage model in the sleep mode:



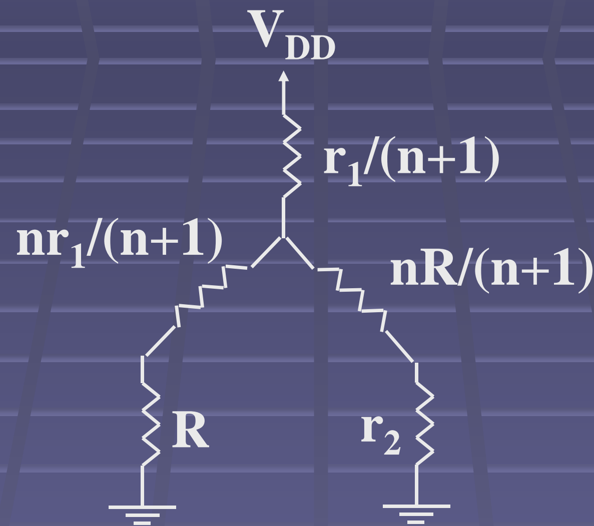
where (R_N=R_P=R and R_{TG}=nR):

$$r_1^* = \frac{r_1 R_P}{r_1 + R_{TG} + R_P} = \frac{1}{n+1} r_1$$

$$r_2^* = \frac{r_1 R_{TG}}{r_1 + R_{TG} + R_P} = \frac{n}{n+1} r_1$$

$$r_3^* = \frac{R_P R_{TG}}{r_1 + R_{TG} + R_P} = \frac{n}{n+1} R$$

Leakage in CR Technique



- Leakage is calculated by writing KVL equations:

$$P_{leakage-CR} = \left(2 + \frac{1}{n}\right)\frac{V_{DD}^2}{R}$$

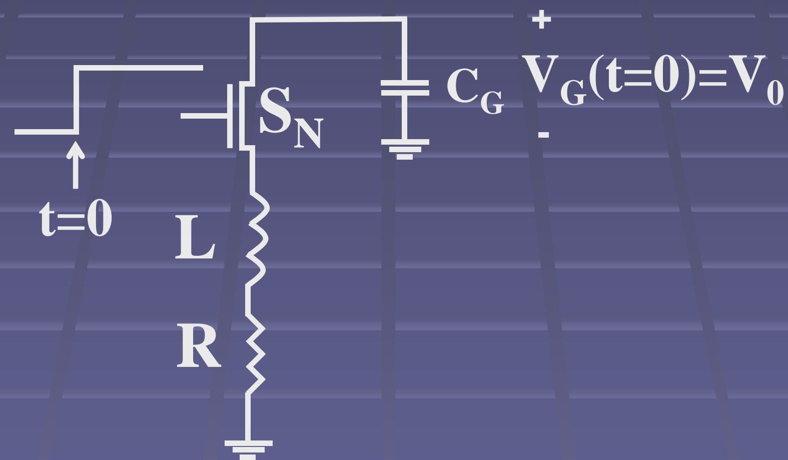
- Leakage has increased by a factor of 1/2n .

Leakage in CR Technique (cont.)

- If n=2, there is 25% increase in the leakage:
  - For short and medium sleep periods, this increase is negligible compared to the saving that we get from the CR technique
  - For long sleep periods, we must use larger *n* by choosing transistors with smaller W/L ratios in the TG
  - This is also beneficial from the layout area point of view
  - Potential disadvantage: CR takes longer to complete.

Ground Bounce (GB) Analysis

- Simple wake-up circuit model for GB analysis:



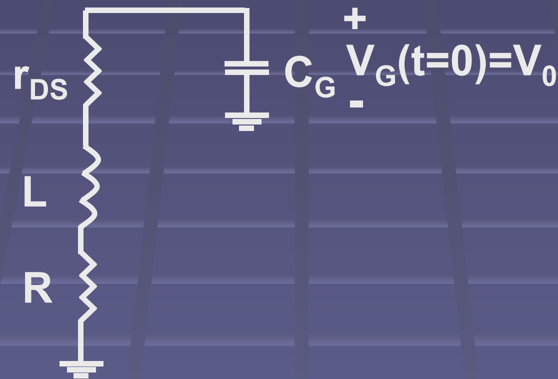
- For conventional MTCMOS: $V_0=V_{DD}$
- For CR MTCMOS: $V_0=V_{DD}/2$

Ground Bounce Analysis (cont.)

- It is well known that:
  - The positive GB peak occurs when S_N operates in saturation region
  - When operating in the saturation region, drain-source current of S_N, and thus the GB value, dose not depend on the V_0 value
  - In CR-MTCMOS, the positive GB peak value remains unchanged
  - The negative GB peak occurs when S_N operates in linear region. This changes in CR-MTCMOS.

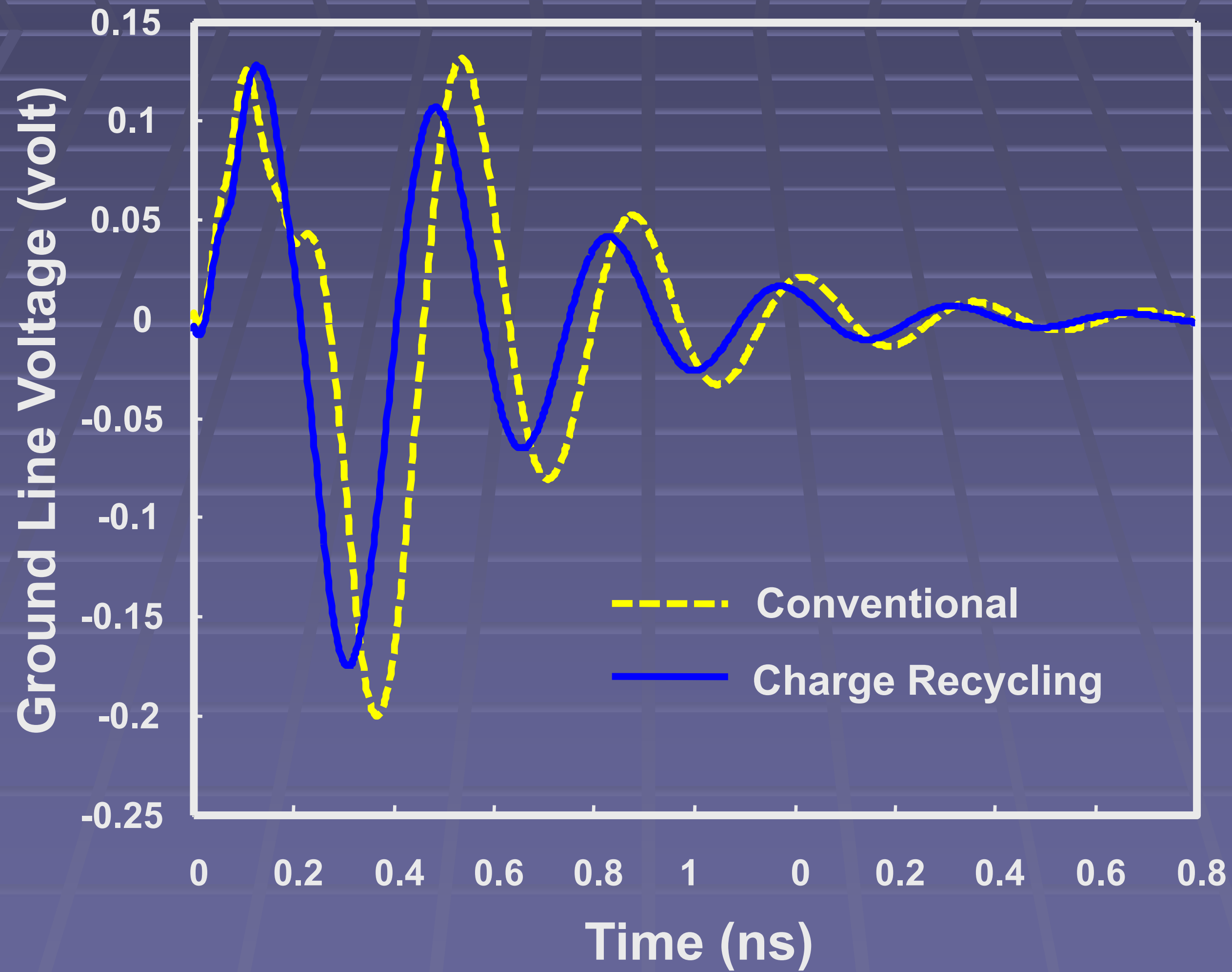
Ground Bounce Analysis (cont.)

- Equivalent circuit model when the negative GB peak happens (r_{DS} is the ON resistance of S_N):



- RLC circuit with initial voltage value V_0 on C_G.
- In CR-MTCMOS:
  - V_0 is reduced by 50%
  - Negative GB peak value is reduced
  - Settling time is lowered.

Ground Bounce

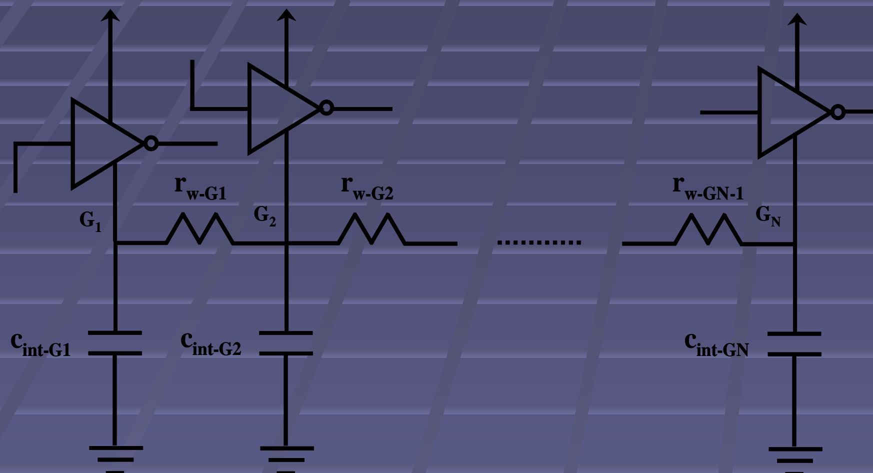


Experimental Results for the 90nm CMOS Node

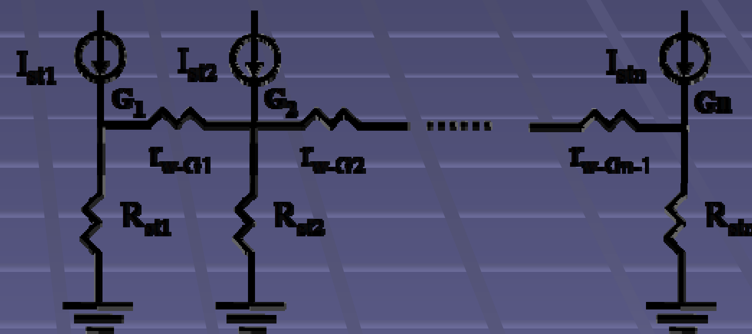
Circuit	Wake up Time (ps)		Mode Transition Energy Cons. (pJ)		Energy Saving (%)	Wake Up Time Reduction (%)
	Conv.	CR	Conv.	CR		
9Sym	494	489.61	29	16	45%	0.9%
C432	240	232.73	10	5.7	43%	3%
C1355	132	125.42	12	7.2	40%	5%
C1908	267	275.63	38	20.5	46%	-3%
C2670	578	573	123	72.6	41%	0.9%
C3540	1500	1545	490	276.9	43%	-3%
C5315	1320	1307	638	357.3	44%	0.1%
C6288	2100	2047	1047	628.2	40%	2.5%
C7552	2310	2402	1532	842.6	45%	-4%

Application to Row-based Designs

Circuit Model of a MTCMOS Row



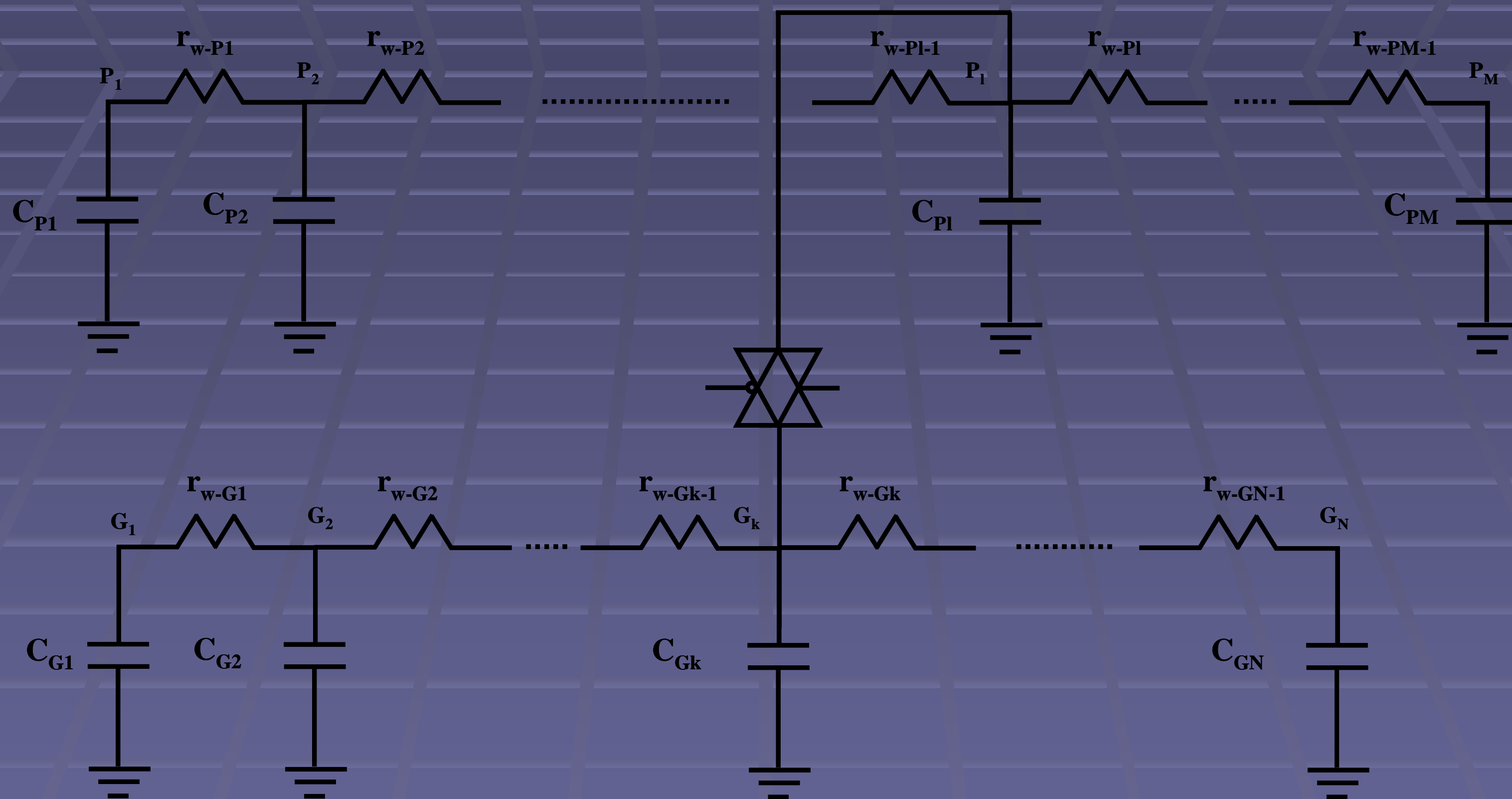
Sleep mode Model



Active mode Model

- We need to decide whether we need a sleep transistor for a specific node G_i, and if needed, what the “optimal” size of the sleep transistor is.

Circuit Model of CR-MTCMOS Rows



- We must decide on the number and positions of the connection points of charge recycling transistors connecting two adjacent cell rows.

Conclusion

- The only known method for reducing energy consumed during transition from the sleep mode to active mode and vice versa
  - Results in about 45% saving in transition mode energy
- It reduces the negative peak of the ground bounce and the settling time of the ground bounce.
  - It does not, in most cases, increase the wakeup time of the circuit.
- Careful sizing needed to control the leakage current path through CR path in sleep mode.