EEC 216 Lecture #9: Leakage

Rajeevan Amirtharajah
University of California, Davis

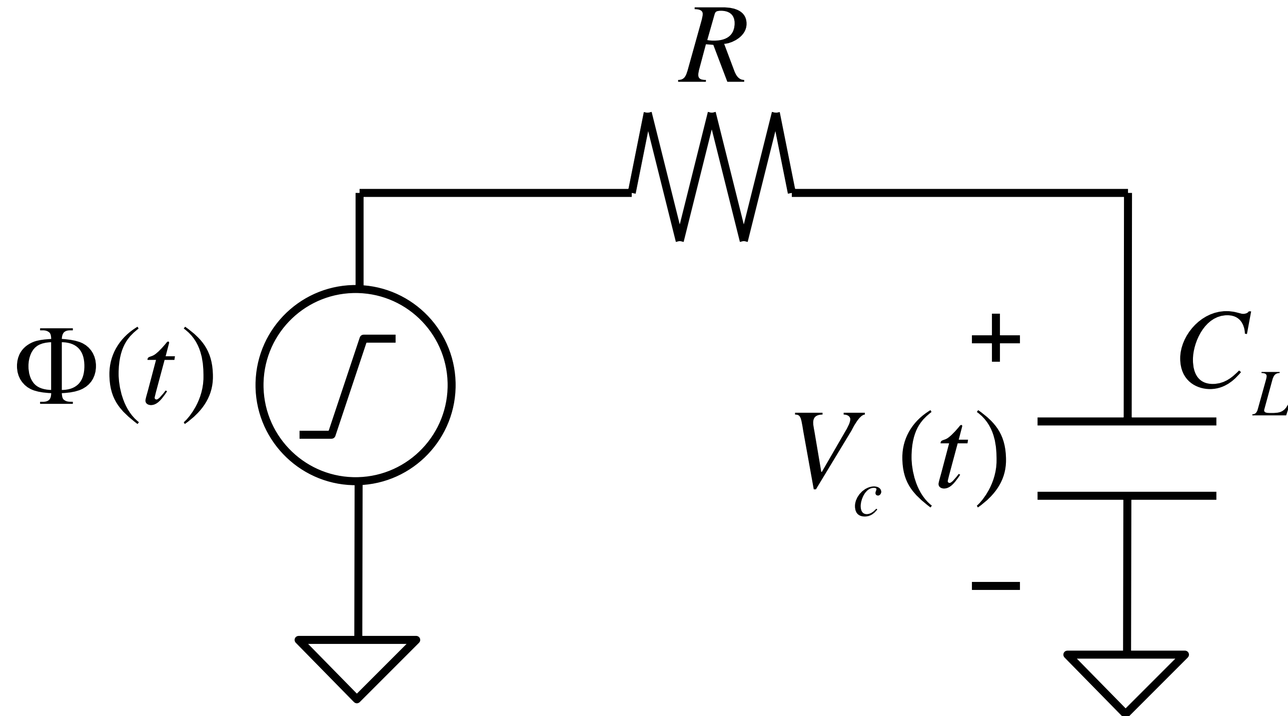
Outline

- **Announcements**
- Review: Energy Recovery Circuits
- Finish Lecture 8
- Leakage Mechanisms
- Circuit Styles for Low Leakage
- Cache SRAM Design Examples
- Next Topic: Subthreshold Circuits

Outline

- Announcements
- **Review: Energy Recovery Circuits**
- Finish Lecture 8
- Leakage Mechanisms
- Circuit Styles for Low Leakage
- Cache SRAM Design Examples
- Next Topic: Subthreshold Circuits

Adiabatic Charging Analysis



$$\Phi = RC\left(\frac{dV_c}{dt}\right) + V_c$$

- **Solve differential equation assuming input is voltage ramp with duration T**

Energy Dissipated With Ramp Driver

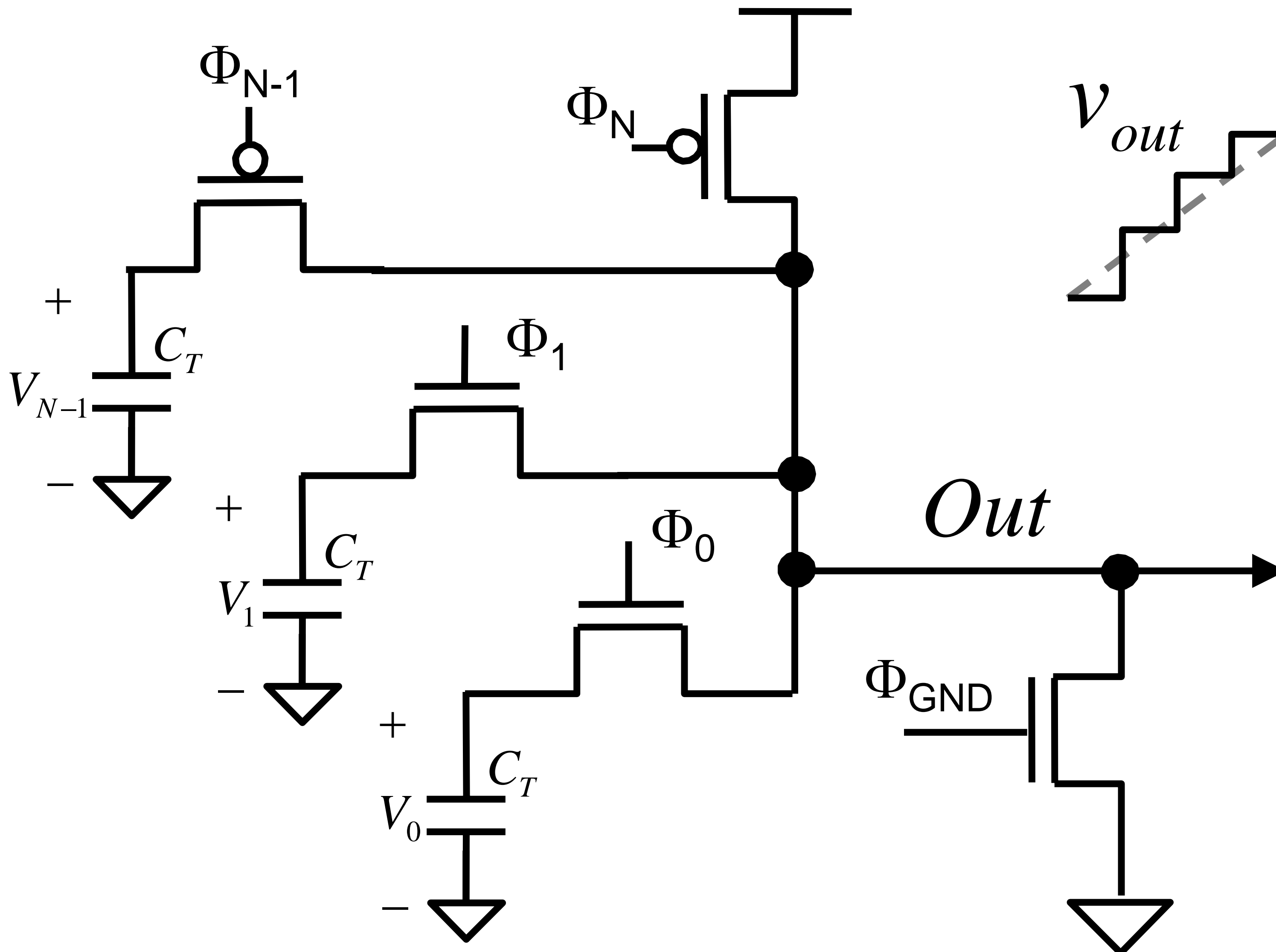
$$E_{diss} = \int_0^\infty i_R(t) V_R(t)\,dt = \int_0^\infty \frac{(\Phi - V_c(t))^2}{R}\,dt$$

$$= \int_0^T \frac{(\Phi - V_c(t))^2}{R}\,dt + \int_T^\infty \frac{(\Phi - V_c(t))^2}{R}\,dt$$

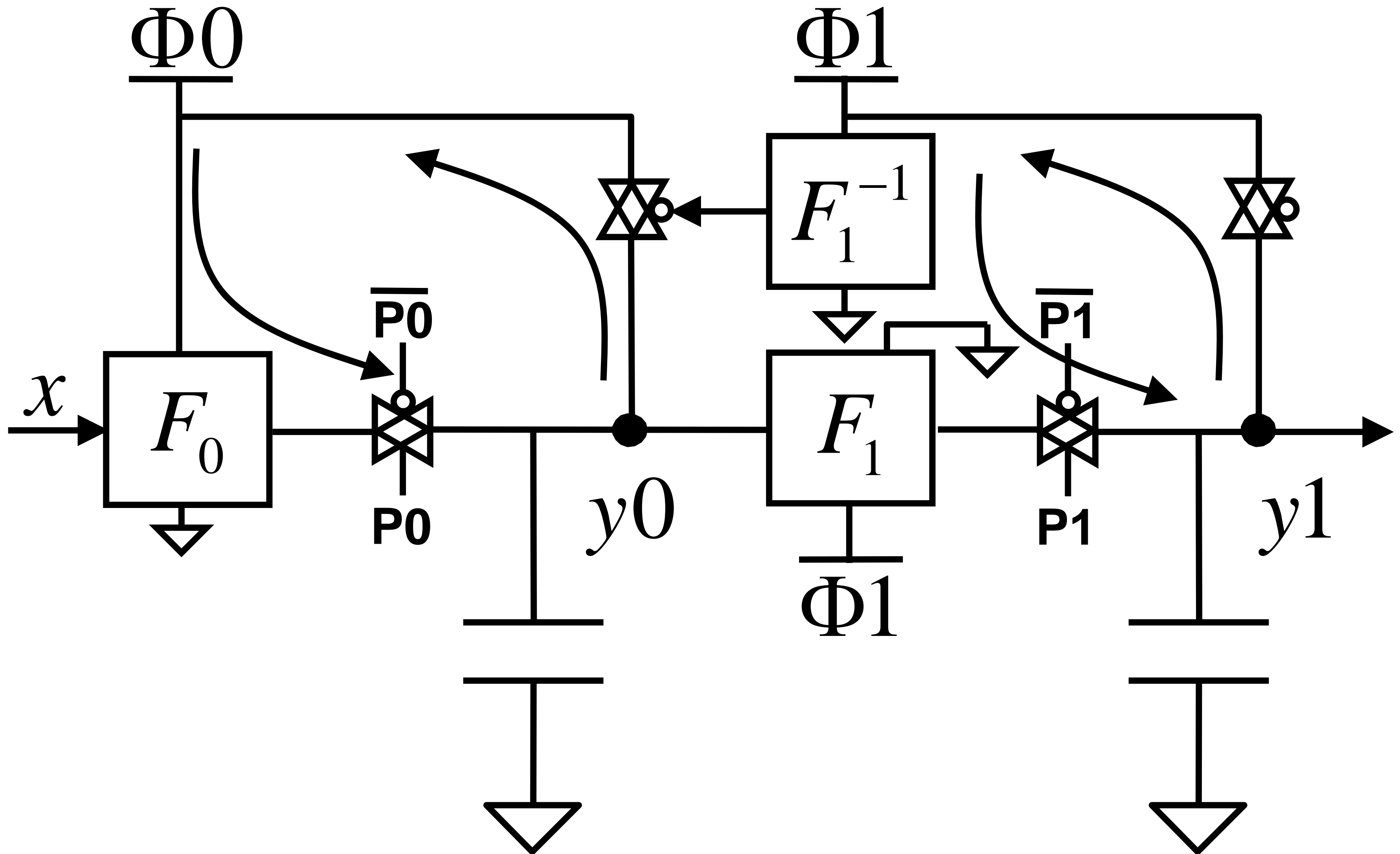
$$= \frac{RC}{T} C V_{DD}^2 \left[1 - \frac{RC}{T} + \frac{RC}{T} e^{-T/RC} \right]$$

- **Consider the extreme cases of *RC* with respect to *T***
  - *RC* << *T* implies less energy dissipation

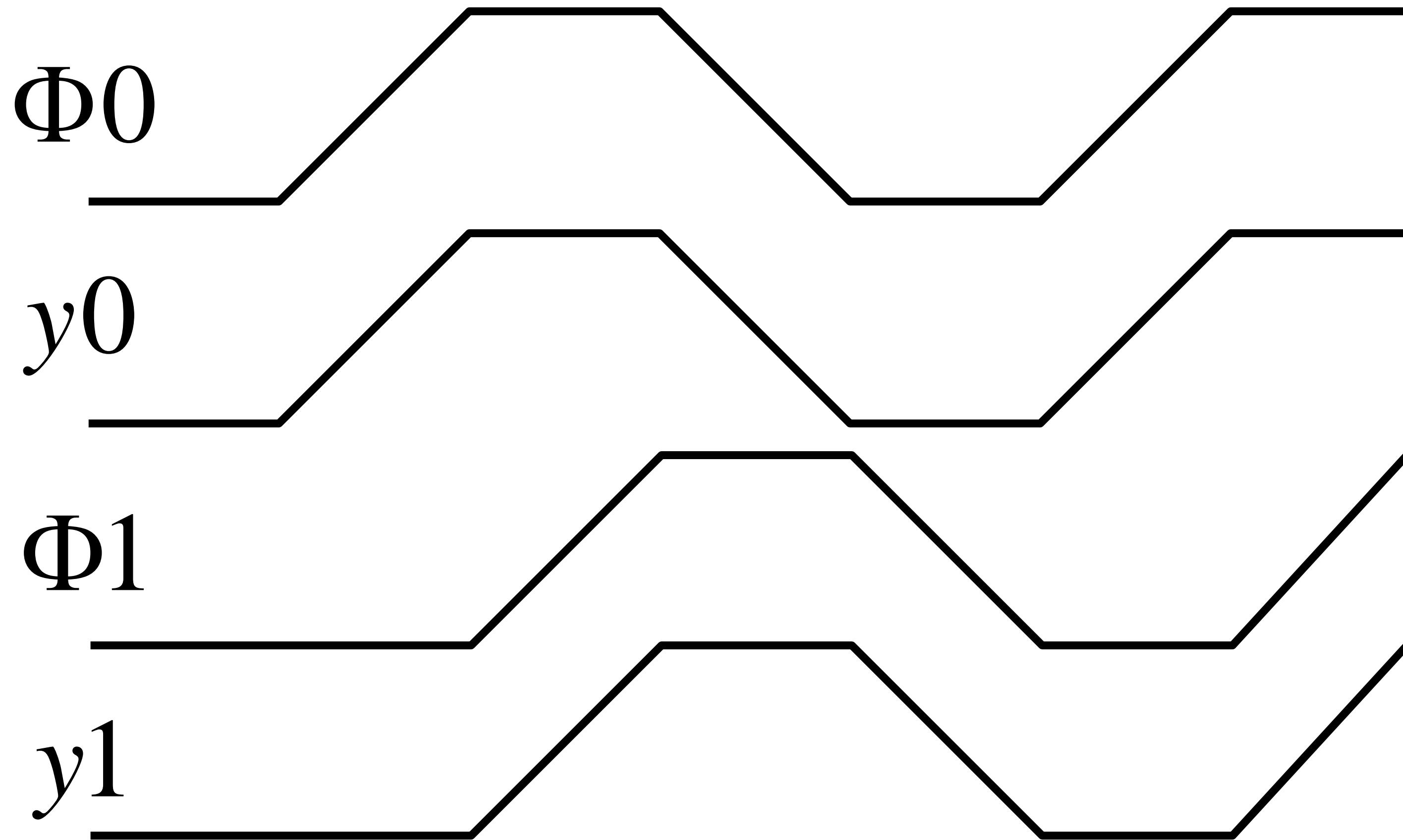
Example Voltage Ramp: Stepwise Charging



Next Stage Controlled Energy Recovery

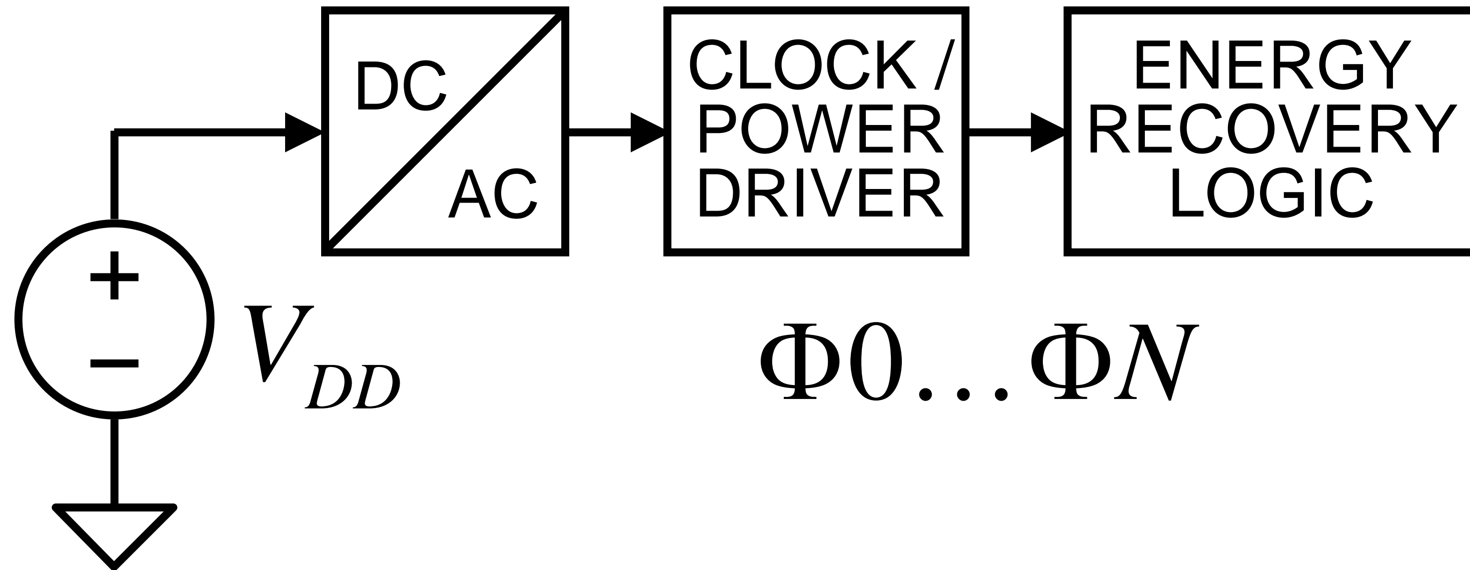


Cascaded Logic Energy Recovery Timing



- **Charge nth stage nodes and then discharge (n-1)th stage nodes**
- **How do we implement the energy recovery phase?**

Energy Recovery System Block Diagram

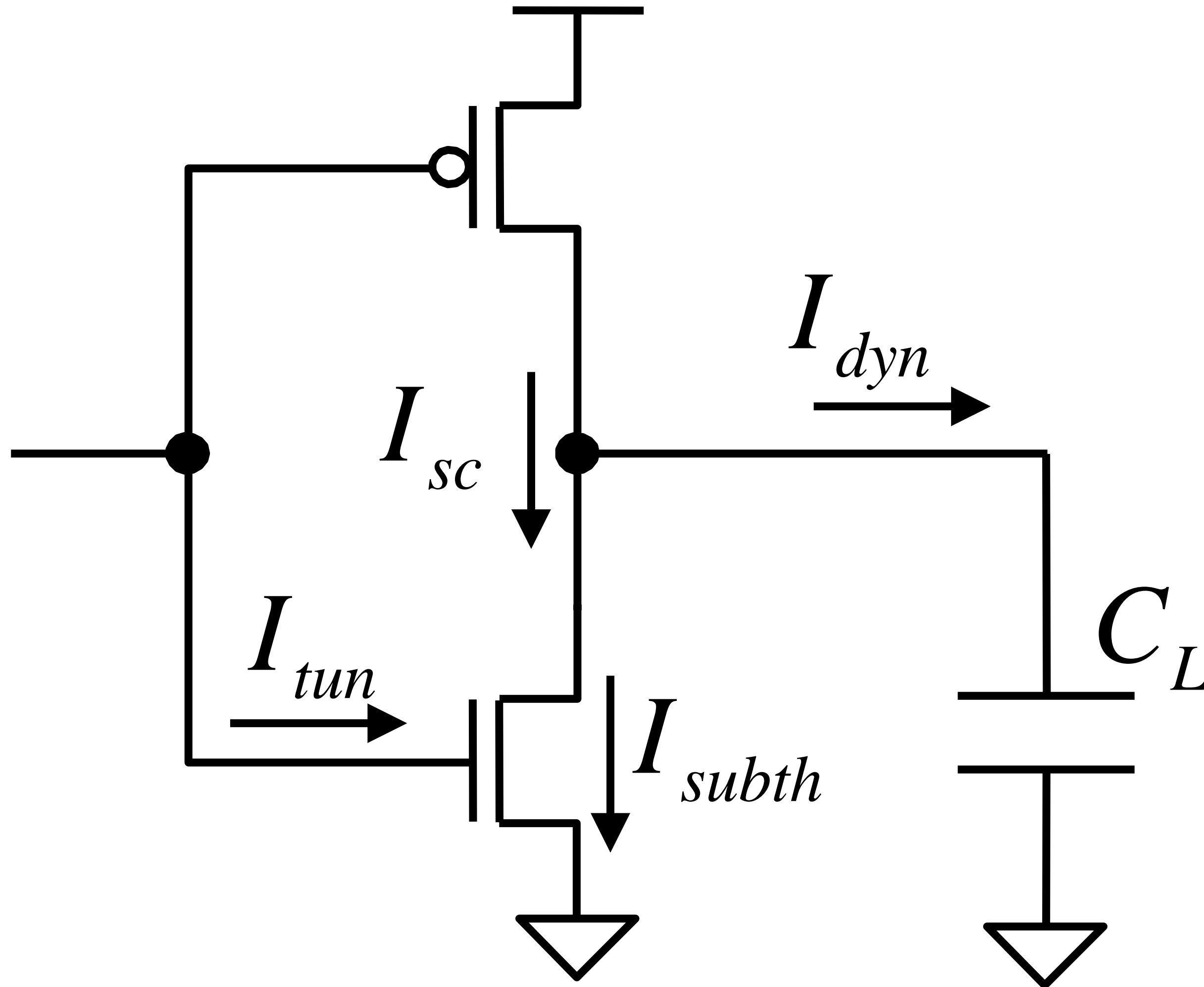


- **Use circuits to generate power / clock waveforms**
- **Generators must use as little power as possible**
  - Resonant RLC circuits often used in these applications
  - Minimize parasitic losses in power / clock generator

Outline

- Announcements
- Review: Energy Recovery Circuits
- Finish Lecture 8
- **Leakage Mechanisms**
- Circuit Styles for Low Leakage
- Cache SRAM Design Examples
- Next Topic: Subthreshold Circuits

CMOS Inverter Example



Components of CMOS Power Dissipation

- **Dynamic Power**
  - Charging and discharging load capacitances
- **Short Circuit (Overlap) Current**
  - Occurs when PMOS and NMOS devices on simultaneously
- **Static Current**
  - Bias circuitry in analog circuits
- **Leakage Current**
  - Reverse-biased diode leakage
  - Subthreshold leakage
  - Tunneling through gate oxide

Extremely Brief MOSFET Review

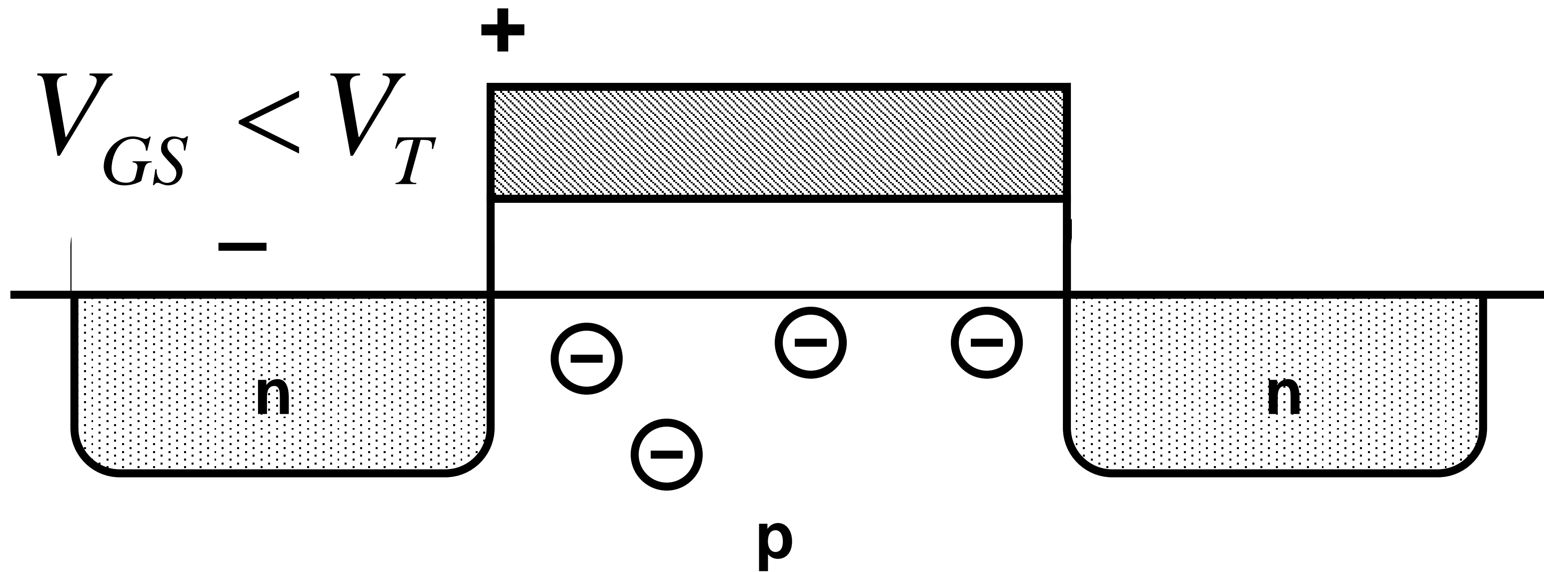
Saturation: $$I_D = \frac{\mu C_{ox}}{2}\frac{W}{L}\left(V_{GS} - V_T\right)^2\left(1 + \lambda V_{DS}\right)$$

Triode: $$I_D = \mu C_{ox}\frac{W}{L}\left(\left(V_{GS} - V_T\right)V_{DS} - \frac{V_{DS}^2}{2}\right)$$

Subthreshold: $$I_D = I_S e^{\frac{V_{GS}}{n\, kT/q}}\left(1 - e^{-\frac{V_{DS}}{kT/q}}\right)$$

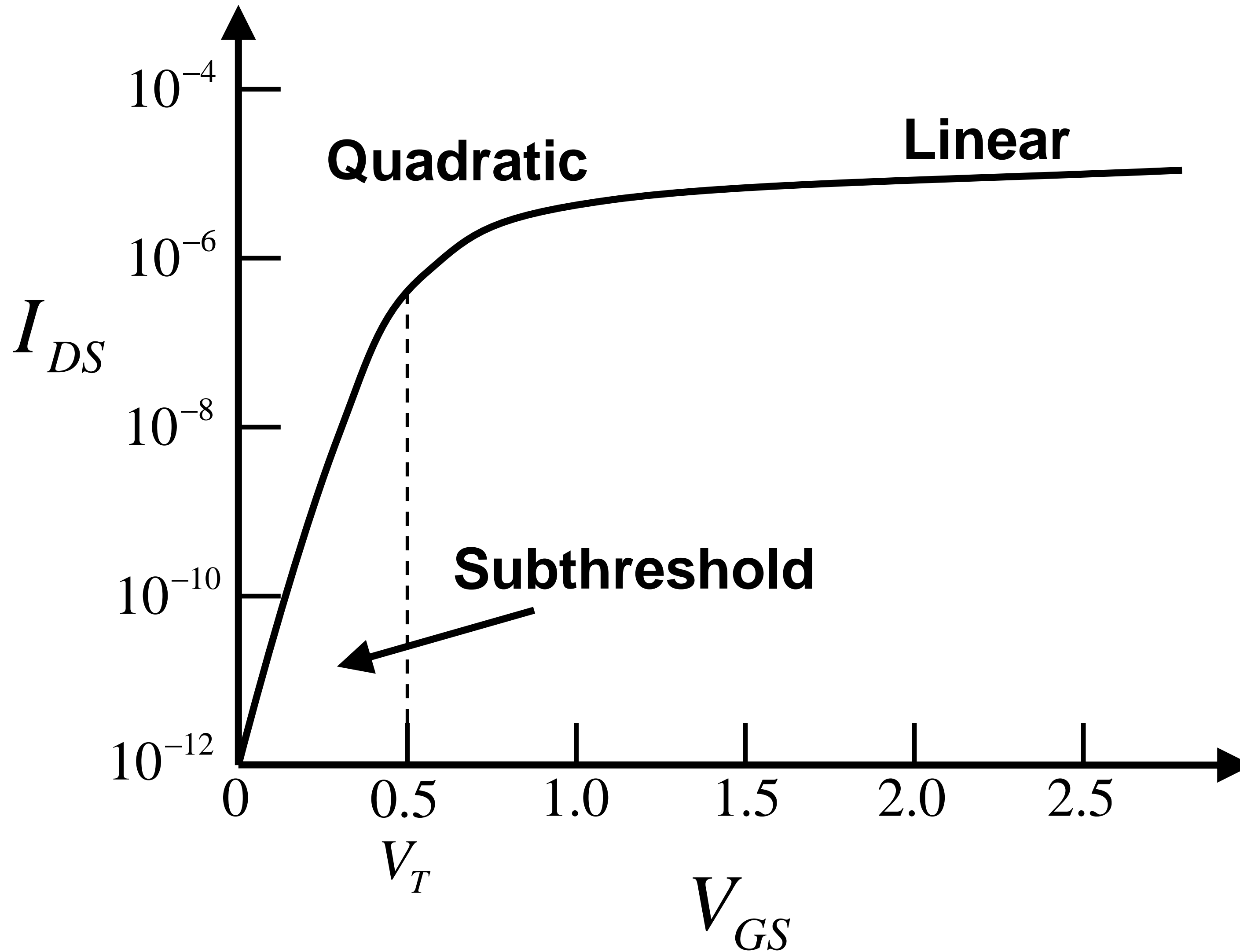
"Classical" MOSFET model, will discuss deep submicron modifications as necessary

Subthreshold Conduction



- **Weak inversion, MOSFET partially conducting**
- **n+ source – p+ bulk – n+ drain forms parasitic bipolar**
- **Approximate current with exponential function**

Drain Current vs. Gate-Source Voltage



Subthreshold Current Equation

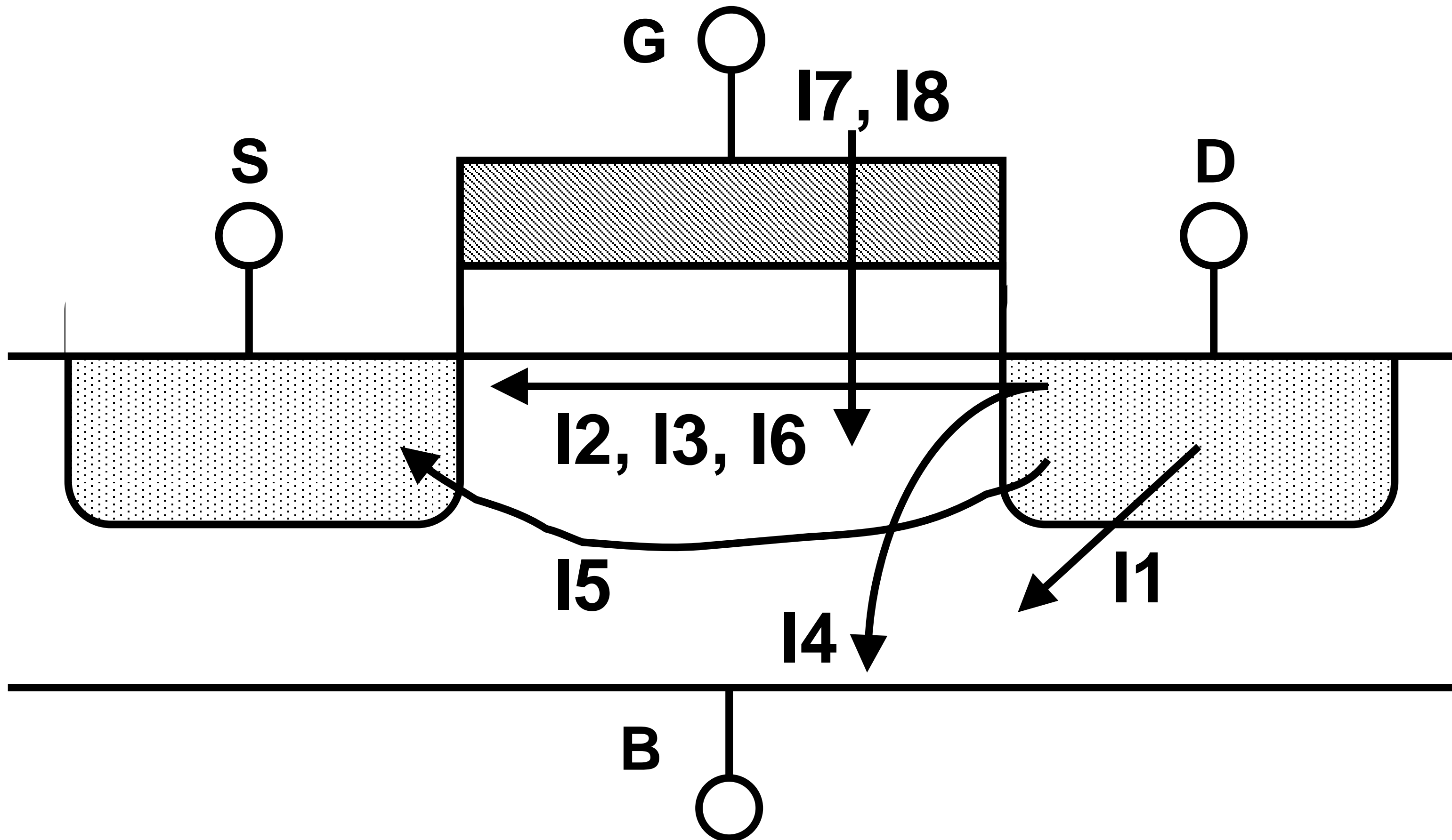
$$I_D = I_S e^{\frac{V_{GS}}{n^{kT}/_q}} \left(1 - e^{-\frac{V_{DS}}{^{kT}/_q}} \right) \left(1 + \lambda V_{DS} \right)$$

- ***I_s* and *n* are empirical parameters**
- **Typically,** $n \geq 1$ **often ranging around** $n \approx 1.5$
- **Usually want small subthreshold leakage for digital designs**
  - Define quality metric: inverse of rate of decline of current wrt V_{GS} below V_T
  - Subthreshold slope factor *S*: $S = n \frac{kT}{q} \ln(10)$

Subthreshold Slope Factor

- **Ideal case: *n* = 1**
  - *S* evaluates to 60 mV/decade (each 60 mV V_{GS} drops below V_T, current drops by 10X)
  - Typically *n* = 1.5 implies slower current decrease at 90 mV/decade
  - Current rolloff further decreased at high temperature, where fast CMOS logic tends to operate
- ***n* determined by intrinsic device topology and structure**
  - Changing n requires different process, like SOI

Leakage Currents in Deep Submicron



- **Many physical mechanisms produce static currents in deep submicron**

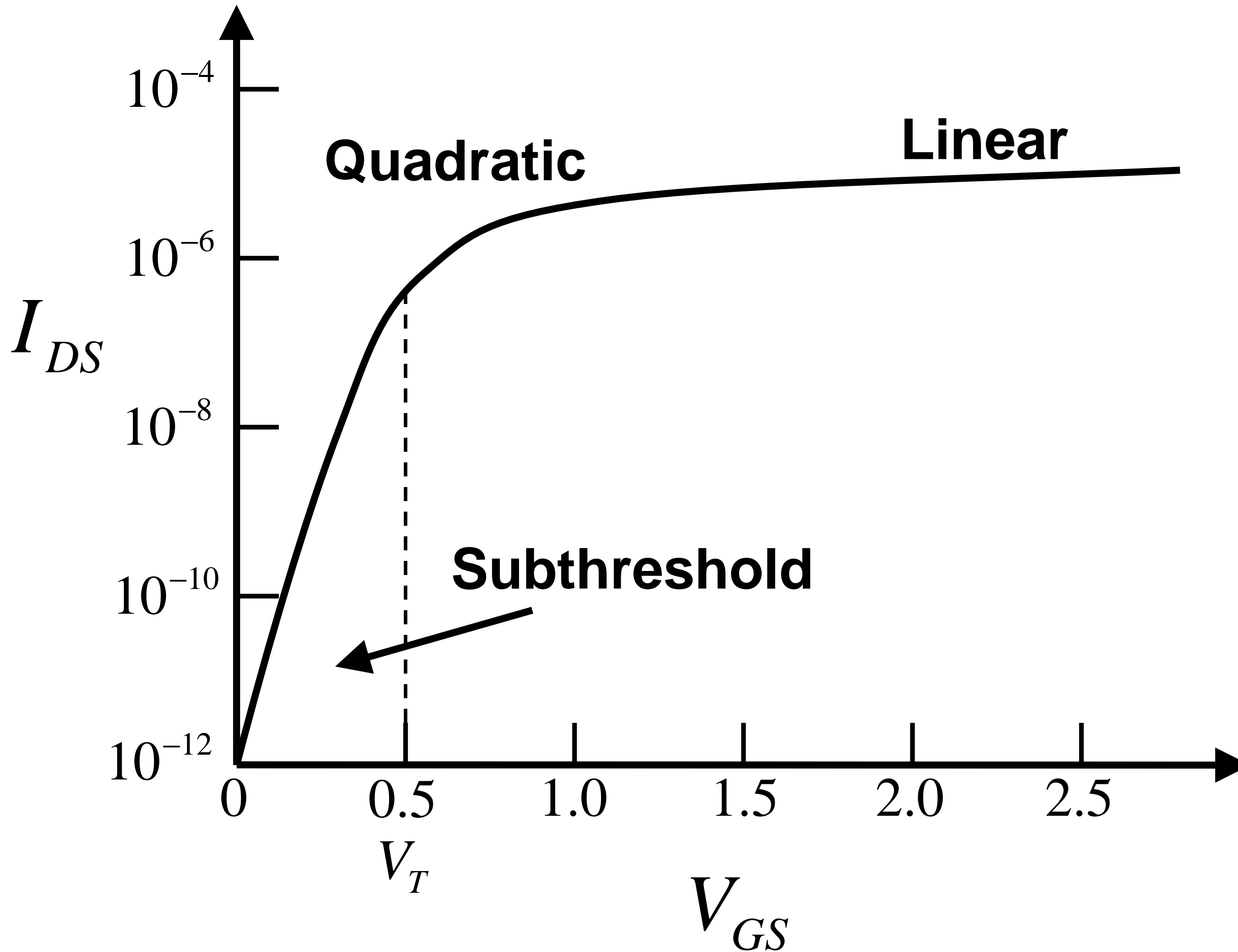
Transistor Leakage Mechanisms

1. pn Reverse Bias Current (I1)
2. Subthreshold (Weak Inversion) (I2)
3. Drain Induced Barrier Lowering (I3)
4. Gate Induced Drain Leakage (I4)
5. Punchthrough (I5)
6. Narrow Width Effect (I6)
7. Gate Oxide Tunneling (I7)
8. Hot Carrier Injection (I8)

pn Reverse Bias Current (I1)

- **Reverse-biased pn junction current has two main components**
  - Minority carrier drift near edge of depletion region
  - Electron-hole pair generation in depletion region of reverse-biased junction
  - If both n and p regions doped heavily, Zener tunneling may also be present
- **In MOSFET, additional leakage can occur**
  - Gated diode device action (gate overlap of drain-well pn junctions)
  - Carrier generation in drain-well depletion regions influenced by gate
- **Function of junction area, doping concentration**
- **Minimal contributor to total off current**

Drain Current vs. Gate-Source Voltage



Subthreshold Current Equation

$$I_D = I_S e^{\frac{V_{GS}}{n kT/q}} \left(1 - e^{-\frac{V_{DS}}{kT/q}} \right) \left(1 + \lambda V_{DS}\right)$$

- ***I_s* and *n* are empirical parameters**
- **Typically,** $n \geq 1$ **often ranging around** $n \approx 1.5$
- **Usually want small subthreshold leakage for digital designs**
  - Define quality metric: inverse rate of decline of current wrt V_{GS} below V_T
  - Subthreshold slope factor *S*: $S = n \frac{kT}{q} \ln(10)$

Detailed Subthreshold Current Equation

$$I_D = A \exp\left(\frac{q}{nkT}\left(V_{GS} - V_{T0} - \gamma V_S + \eta V_D\right)\right)\left(1 - \exp\left(\frac{-qV_{DS}}{kT}\right)\right)$$

$$A = \mu_0 C_{ox} \frac{W}{L}\left(\frac{kT}{q}\right)^2 e^{1.8}$$

- **V_{T0} = zero bias threshold voltage,**
- **μ0 = zero bias mobility**
- **Cox = gate oxide capacitance per unit area**
- **γ = linear body effect coefficient (small source voltage)**
- **η = DIBL coefficient**

Subthreshold Slope of Various Processes

Technology	Doping	S (mV / decade)
0.8 μm, 5 V CMOS	LDD	86
0.6 μm, 5 V CMOS	LDD	80
0.35 μm, 3.3 V BiCMOS	LDD	80
0.35 μm, 2.5 V CMOS	HDD	78
0.25 μm, 1.8 V CMOS	HDD	85

- **Roy & Prasad, p. 216**

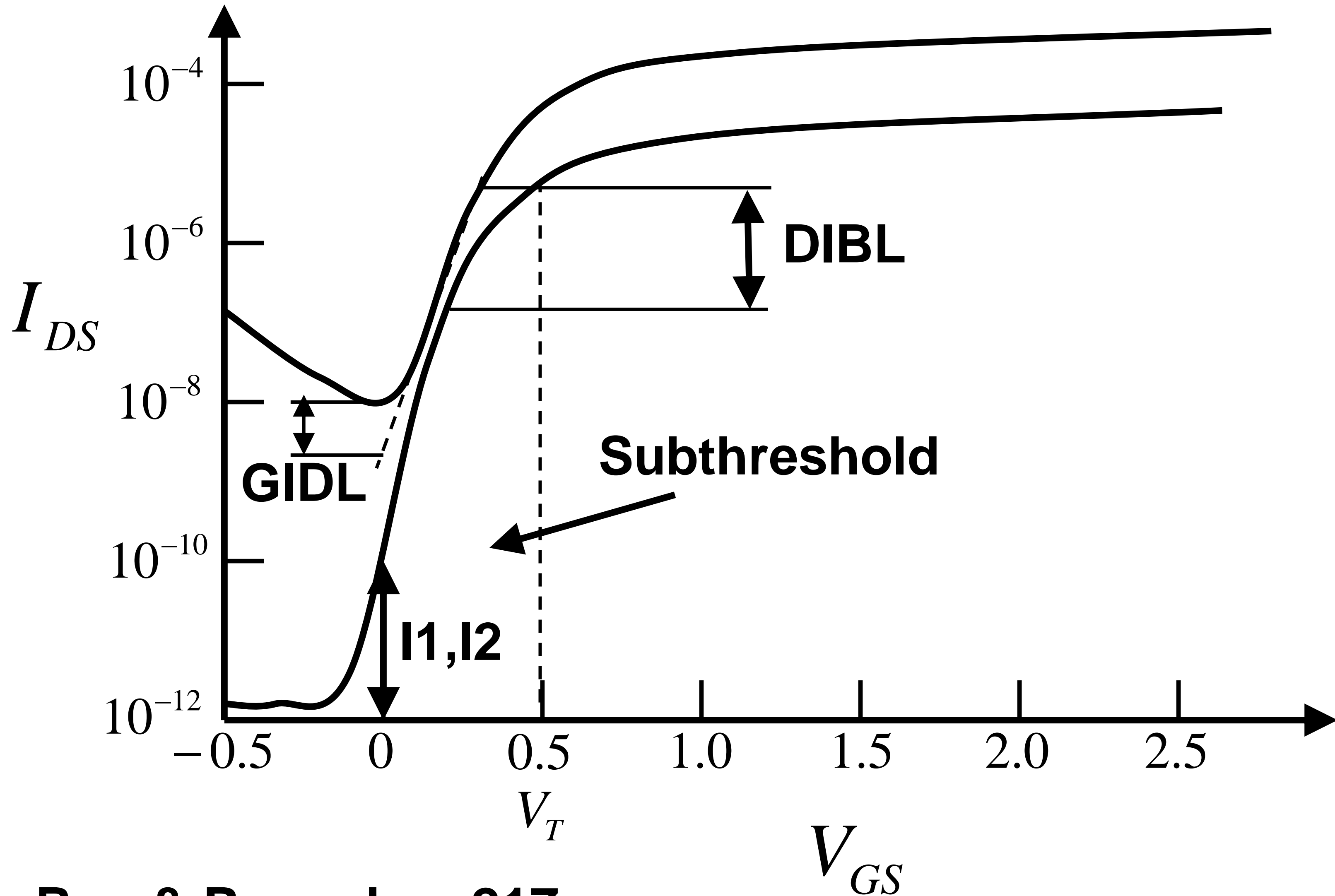
Drain Induced Barrier Lowering (I3)

- **DIBL occurs when drain depletion region interacts with source near channel surface**
  - Lowering source potential barrier
  - Source injects carriers into channel without influence of gate voltage
  - DIBL enhanced at higher drain voltage, shorter effective channel length
  - Surface DIBL happens before deep bulk punchthrough
- **DIBL does not change *S* but lowers V_T**
  - Higher surface, channel doping and shallow junctions reduce DIBL leakage current mechanism

Gate Induced Drain Leakage (I4)

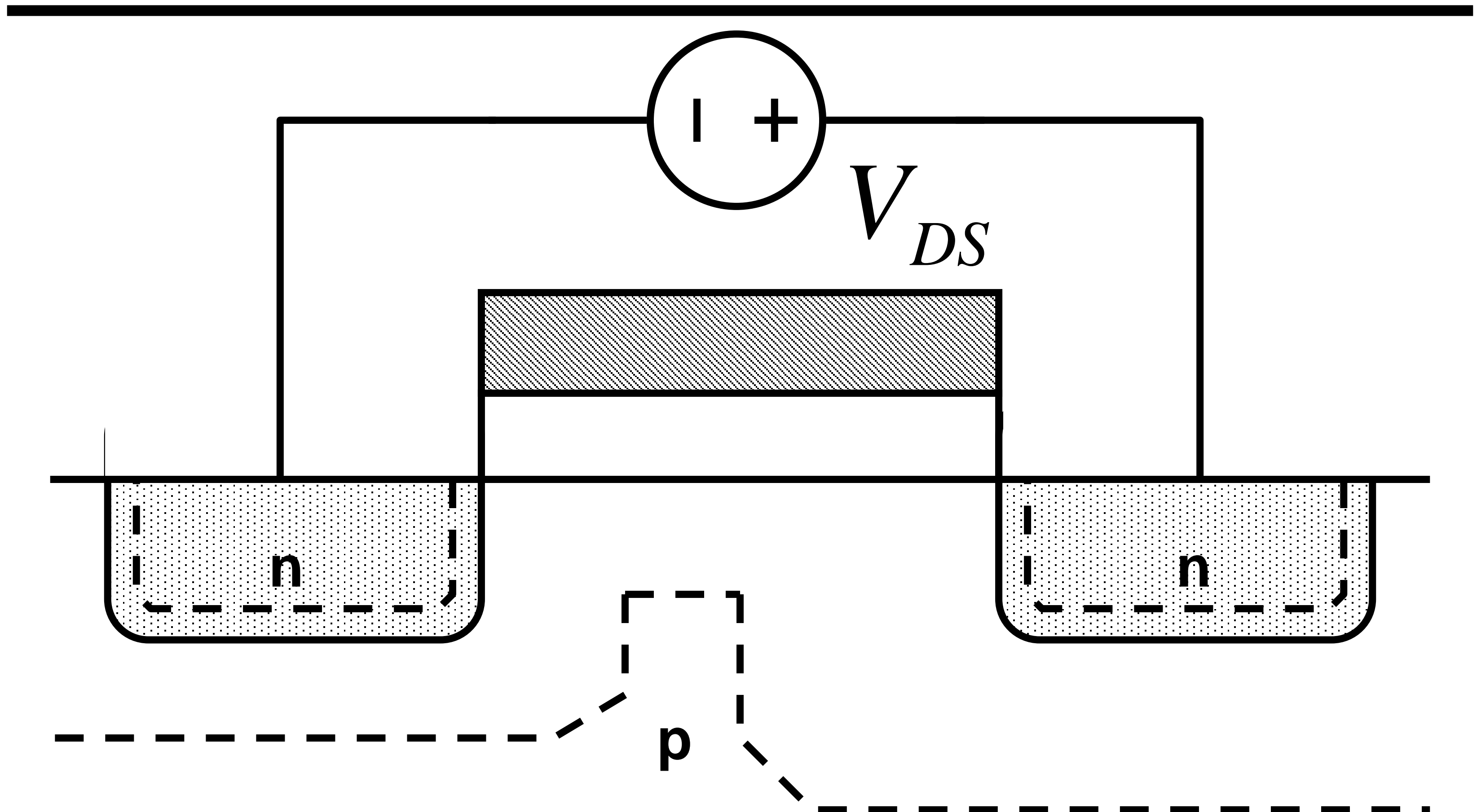
- **GIDL current appears in high E-field region under gate / drain overlap causing deep depletion**
  - Occurs at low V_G and high V_D bias
  - Generates carriers into substrate from surface traps, band-to-band tunneling
  - Localized along channel width between gate and drain
  - Seen as "hook" in I-V characteristic causing increasing current for negative V_G
  - Thinner oxide, higher VDD, lightly-doped drain enhance GIDL
- **Can be major obstacle to reducing off current**

Revised Drain Current vs. Gate Voltage



- **Roy & Prasad, p. 217**

Punchthrough



- Source / Drain depletion regions “touch” deep inside channel

Punchthrough Channel Current (I5)

- **Space-charge condition allows channel current to flow deep in subgate region**
  - Gate loses control of subgate channel region
- **Current varies quadratically with drain voltage**
  - Subthreshold slope factor *S* increases to reflect increase in drain leakage
- **Regarded as subsurface version of DIBL**

Gate Oxide Tunneling (I7)

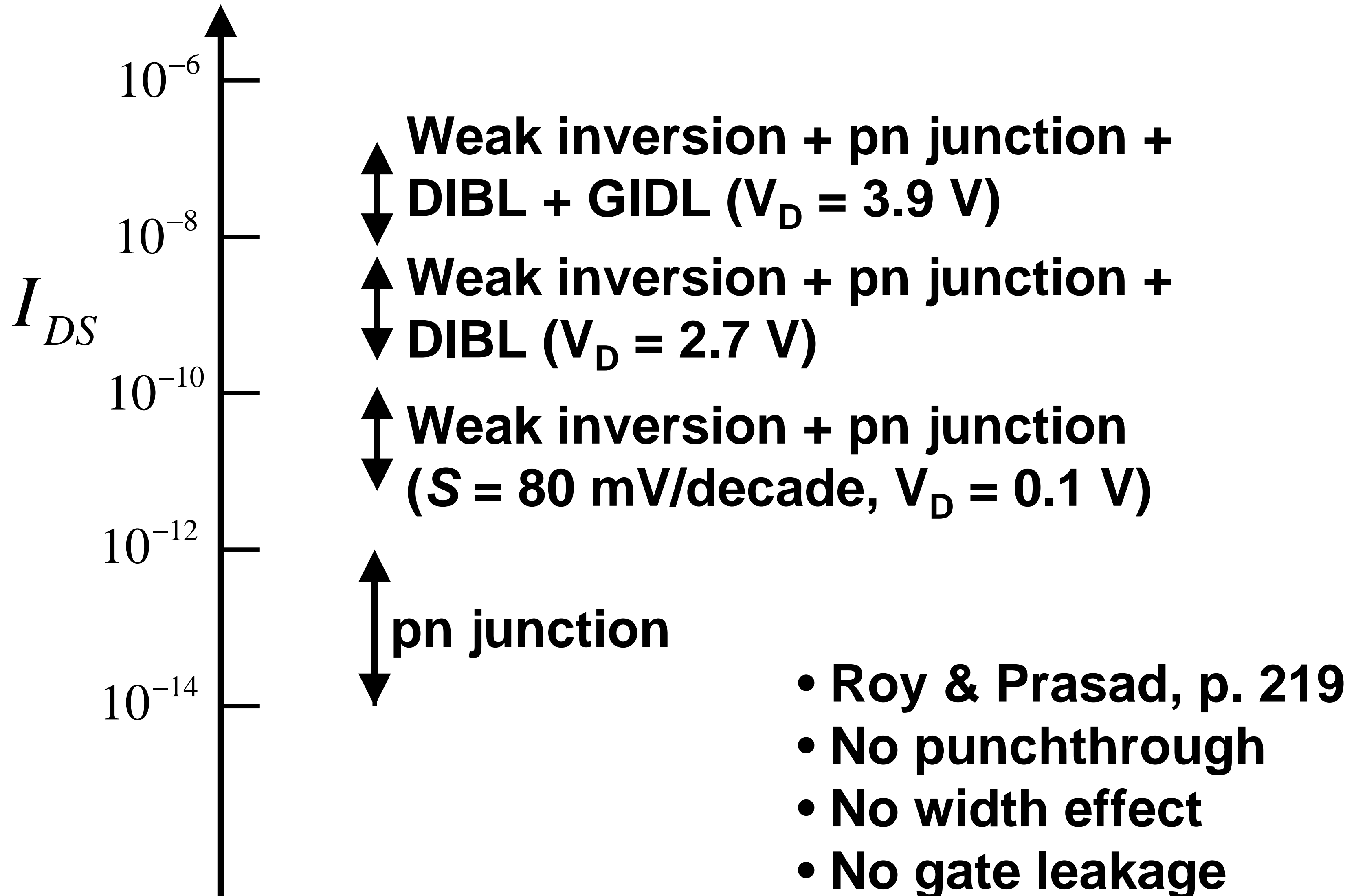
$$I_{OX} = AE_{OX}^2 e^{-B/E_{OX}}$$

- **High E-field E_{OX} can cause direct tunneling through gate oxide or Fowler-Nordheim (FN) tunneling through oxide bands**
- **Typically, FN tunneling at higher field strength than operating conditions (likely remain in future)**
- **Significant at oxide thickness < 50 Angstroms**
- **Could become dominant leakage mechanism as oxides get thinner**
  - High K dielectrics might make better
  - Interesting circuit design issues (see ISSCC 2004)

Other Leakage Effects

- **Narrow Width Effect (I6)**
  - V_T increases for geometric gate widths around 0.5 μm in non-trench isolated technologies
  - Opposite effect in trench isolated technologies: V_T decreases for widths below 0.5 μm
- **Hot Carrier Injection (I8)**
  - Short channel devices susceptible to energetic carrier injection into gate oxide
  - Measurable as gate and substrate currents
  - Charges are a reliability risk leading to device failure
  - Increased amplitude as length reduced unless V_{DD} scaled accordingly

Leakage Summary



- Roy & Prasad, p. 219
- No punchthrough
- No width effect
- No gate leakage

Leakage Current Estimation

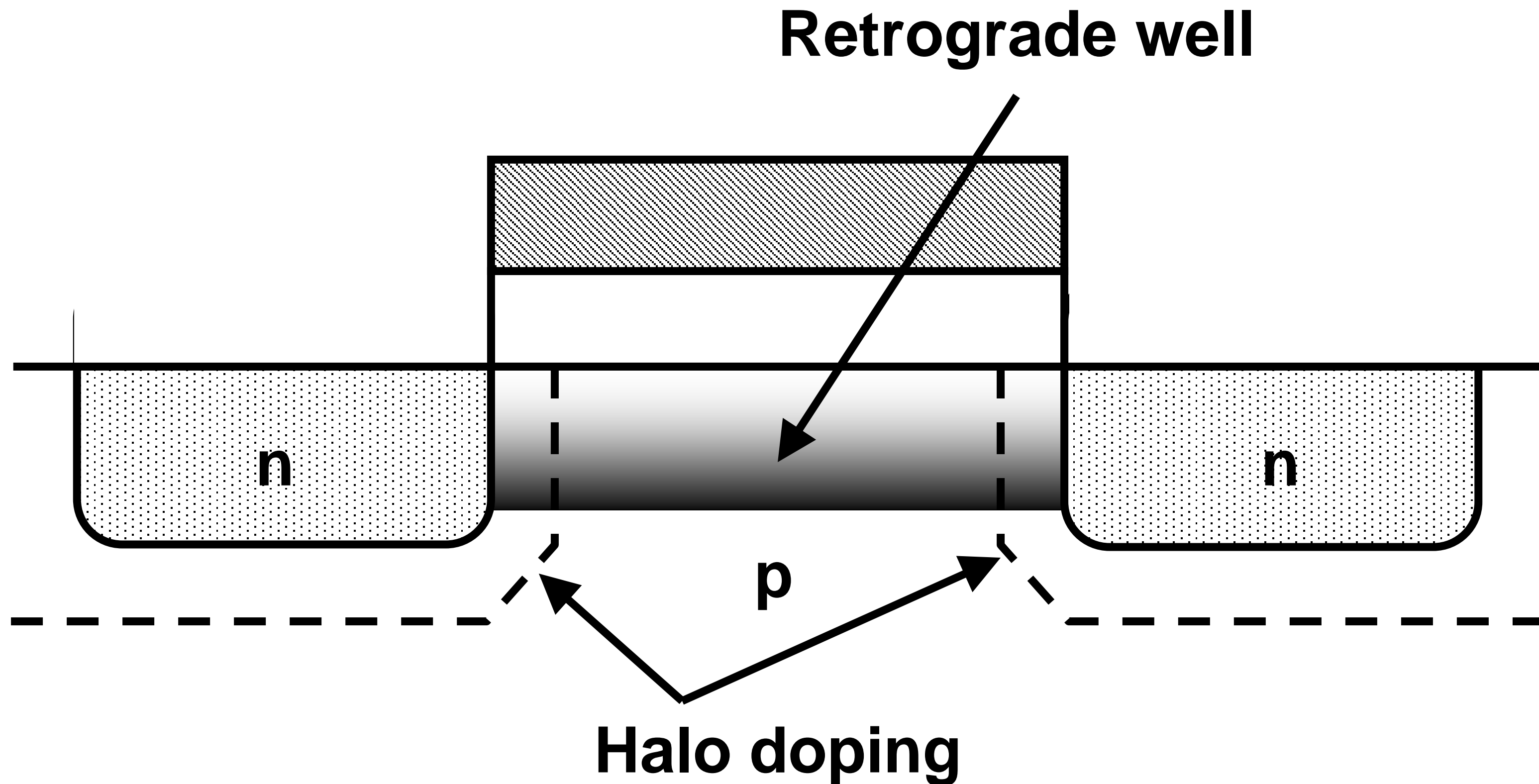
$$P_{leak} = \sum_i I_{DS_i} V_{DS_i}$$

- **Parallel transistors, simply add leakage contributions for each one**
- **For series connected devices, calculating leakage currents more complex**
  - Equate subthreshold currents through each device in series stack
  - Solve for V_{DS1} (first device in series stack) in terms of V_{DD} assuming source voltage small
  - Remaining voltages must sum to total voltage drop across series stack

Outline

- Announcements
- Review: Energy Recovery Circuits
- Finish Lecture 8
- Leakage Mechanisms
- **Circuit Styles for Low Leakage**
- Cache SRAM Design Examples
- Next Topic: Subthreshold Circuits

Channel Engineering for Reduced Leakage

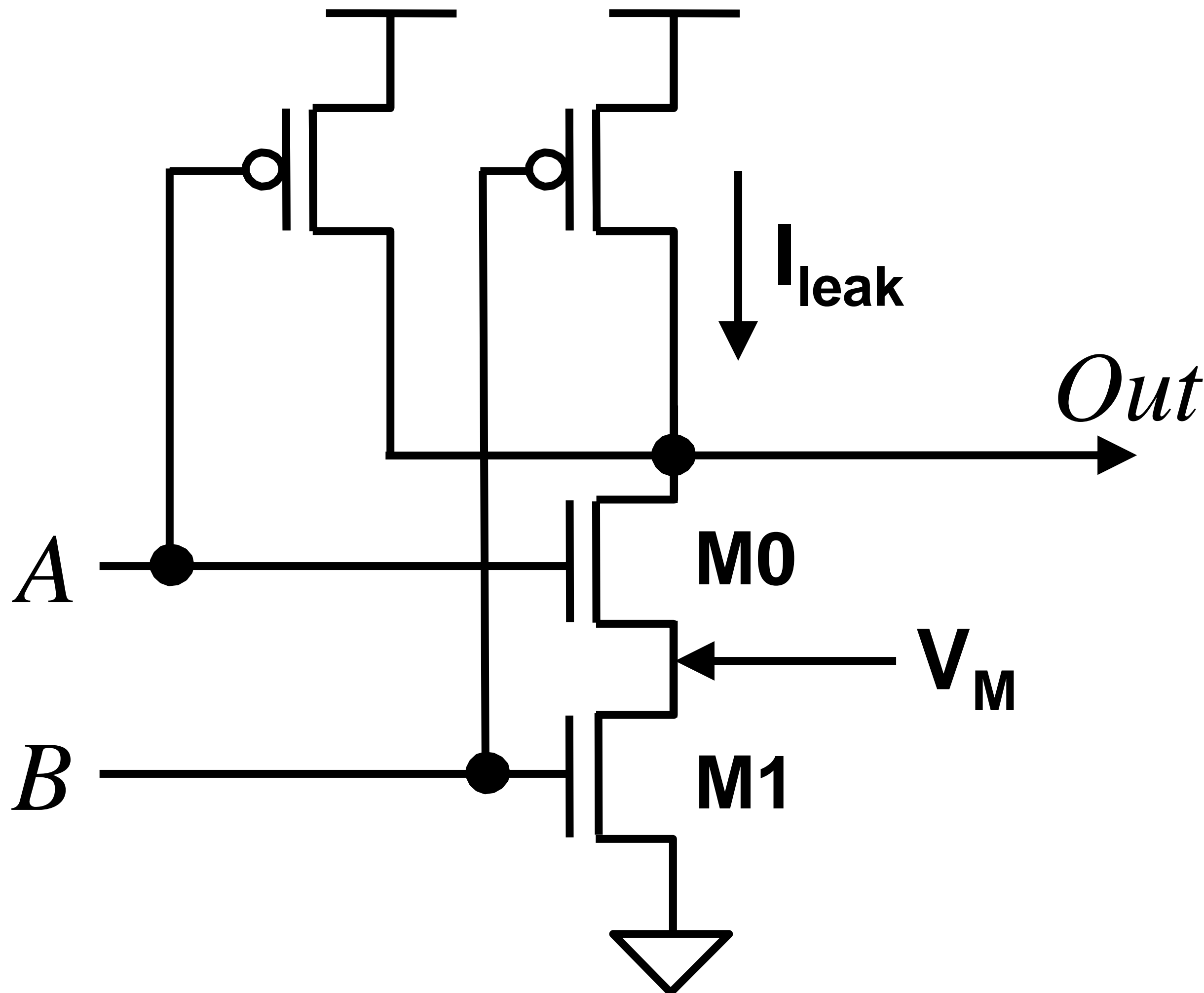


- Goal: optimize channel profile
- Minimize leakage while maximizing drive current

Modifying Channel for Leakage Reduction

- **Process modifications can be used to decrease subthreshold leakage**
- **Retrograde doping**
  - Vertically non uniform, low to high doping concentration going deeper into the substrate
  - Increase mobility near channel surface
  - Creates barrier to punchthrough in bulk
  - Reduce impact of short channel length on V_T
- **Halo doping**
  - Highly doped p-type implanted near channel ends
  - Reduces charge-sharing effects from source and drain fields, decreases DIBL and punchthrough

Stacking Effect in Two-Input NAND Gate



- **Multiple off transistors dramatically cuts leakage**

Stacking Transistors Leakage Effects

- **Intermediate node voltage $V_M > 0$ V**
- **Positive source voltage for device M0 has three major effects:**

1. $V_{GS0} = V_{in} - V_M = 0\text{ V} - V_M < 0$ V reduces subthreshold current exponentially
2. Body to source potential ($V_{BS0} = 0\text{ V} - V_M < 0$ V) becomes negative, increasing V_{TH} through increased body effect, thus decreasing I_{leak}
3. Drain to source potential ($V_{DS0} = V_{DD} - V_M$) decreases, increasing V_{TH} through reduced DIBL, thus decreasing I_{leak} further

Stacking Effect Impact

- **Leakage current drops by an order of magnitude**
- **Leakage highly dependent on input vector**
  - Can reduce leakage power by choosing input bits carefully
  - Large search space (2^N possible vectors), so exhaustive search impossible for large fan-in logic
  - Can use genetic algorithm to find near-optimal input vector
- **Can effectively control leakage in standby mode**

Multiple Threshold Voltages

- **If process implements two threshold devices, can control leakage by mixing both types of devices**
  - Use high V_T transistors to interrupt leakage paths
  - Use low V_T devices for high performance
- **Implementing multiple thresholds**
  - Multiple channel doping densities
  - Multiple gate oxides
  - Multiple channel lengths
  - Multiple body biases

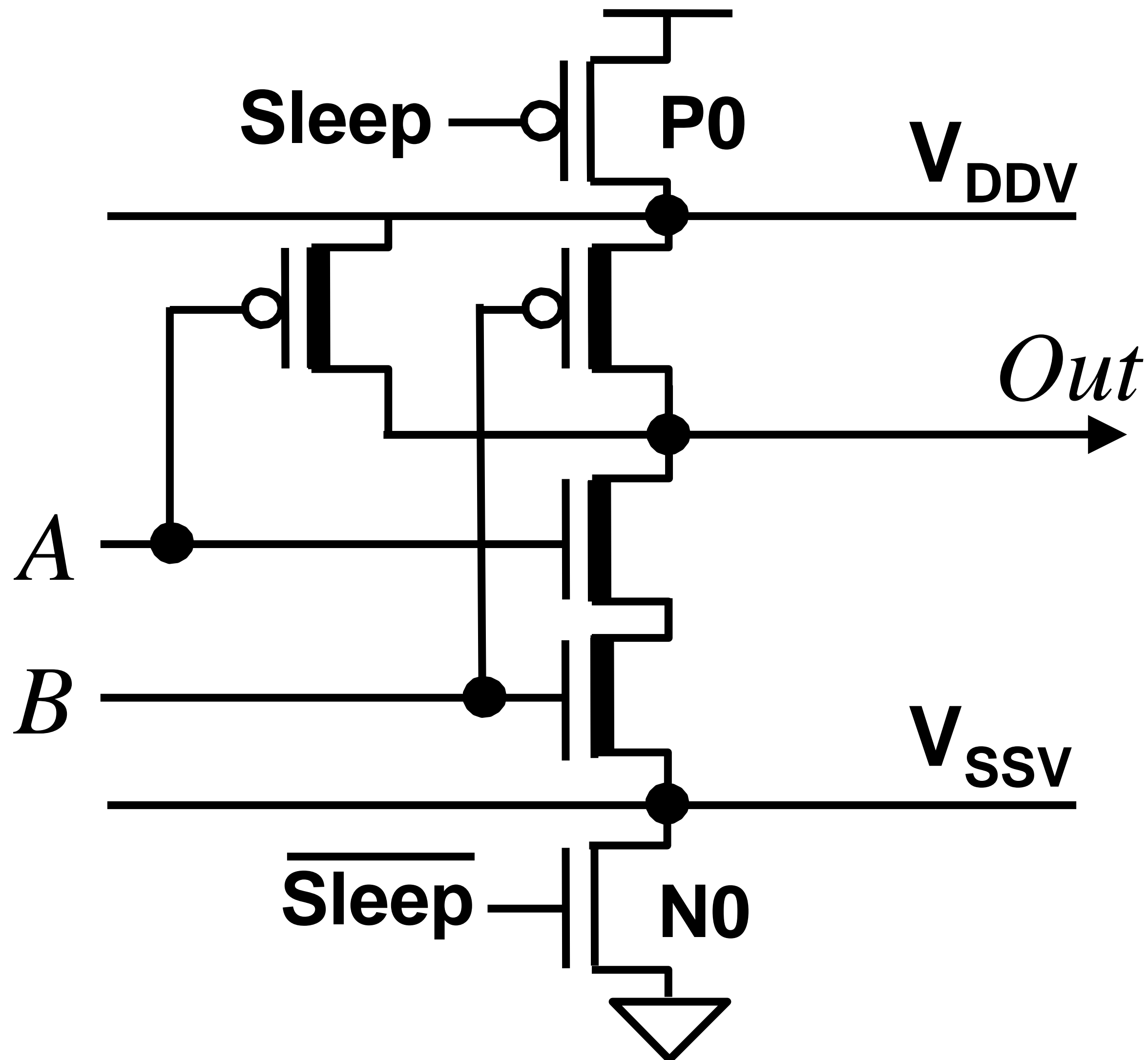
Implementing Multiple Threshold Voltages I

- **Multiple Channel Doping**
  - Varying channel dopant concentration shifts V_T
  - Requires additional mask steps
  - Threshold voltage variation makes it challenging to achieve consistently, esp. when thresholds close
  - Increasingly difficult in future deep submicron
- **Multiple Gate Oxides**
  - Grow two different oxide thicknesses
  - Thicker oxide results in higher V_T, lower subthreshold leakage and gate tunneling current, lower dynamic power through reduced gate capacitance
  - Must increase channel length with oxide thickness

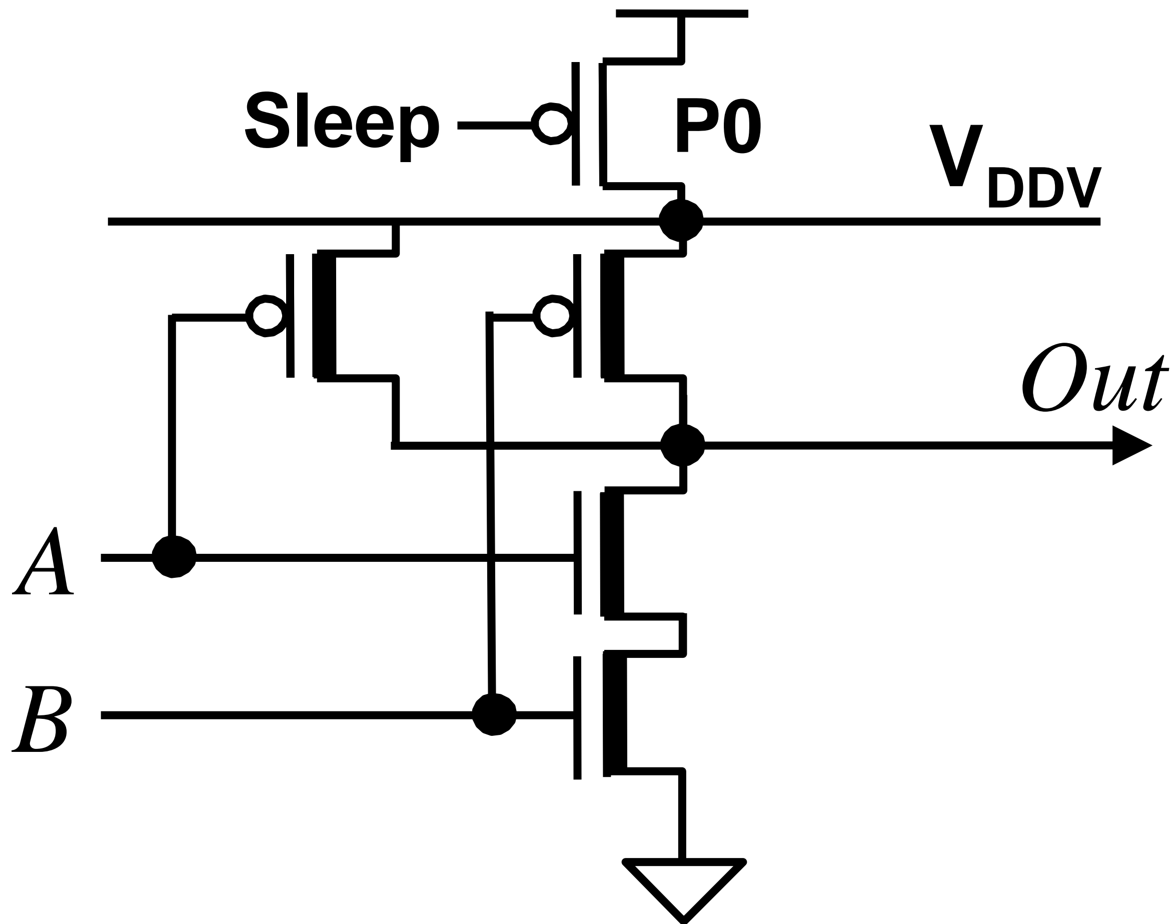
Implementing Multiple Threshold Voltages II

- **Multiple Channel Lengths**
  - Decreasing channel length reduces “V_T” (consider threshold to be V_{GS} which results in a fixed current)
  - Achieved in conventional CMOS technology
  - Longer channels increase gate capacitance and dynamic power
- **Multiple Body Biases**
  - Body (substrate) voltage changed to modify V_T
  - For individual transistor control, requires triple well process since devices cannot share same well
  - Easy to include in Silicon-on-Insulator (SOI) processes since devices automatically isolated

Multiple Threshold CMOS

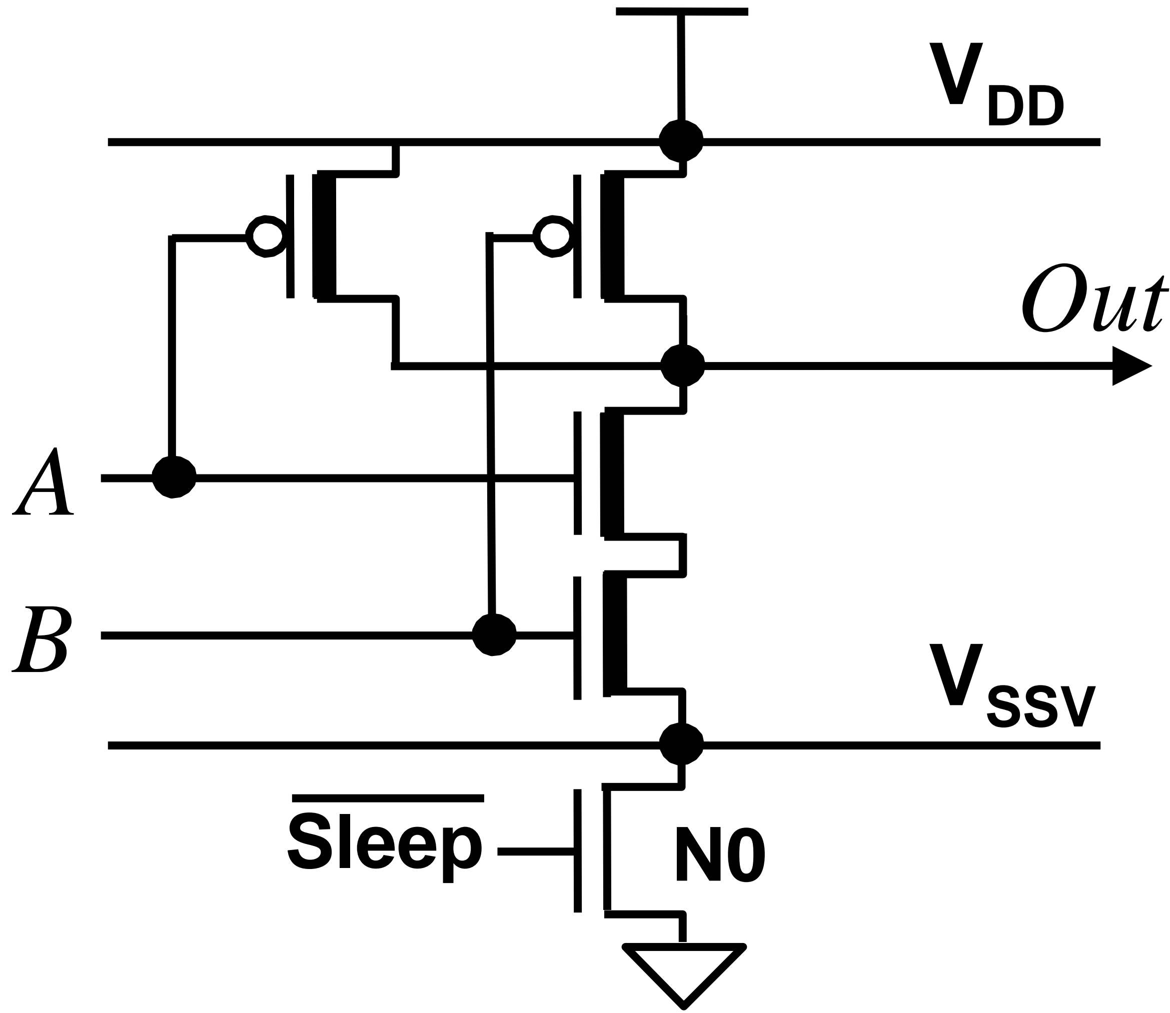


PMOS Insertion Multiple Threshold CMOS



- **Use only PMOS high V_T device to limit leakage current**

NMOS Insertion Multiple Threshold CMOS



- Use only NMOS high V_T device to limit leakage current

Multiple Threshold CMOS Design

- **Must design sleep transistors with low on resistance so virtual supplies almost function like real supplies**
  - NMOS insertion better since a narrower device results in same on resistance
  - Easy to implement based on existing circuits
- **MTCMOS only reduces standby leakage**
  - Active mode leakage also a concern
- **Large inserted FETs increase area and delay**
- **Data retention in standby mode requires extra high V_T memory circuit**

Super Cutoff CMOS

$V_{DD} + \Delta V$

P0

V_{DDV}

Out

A

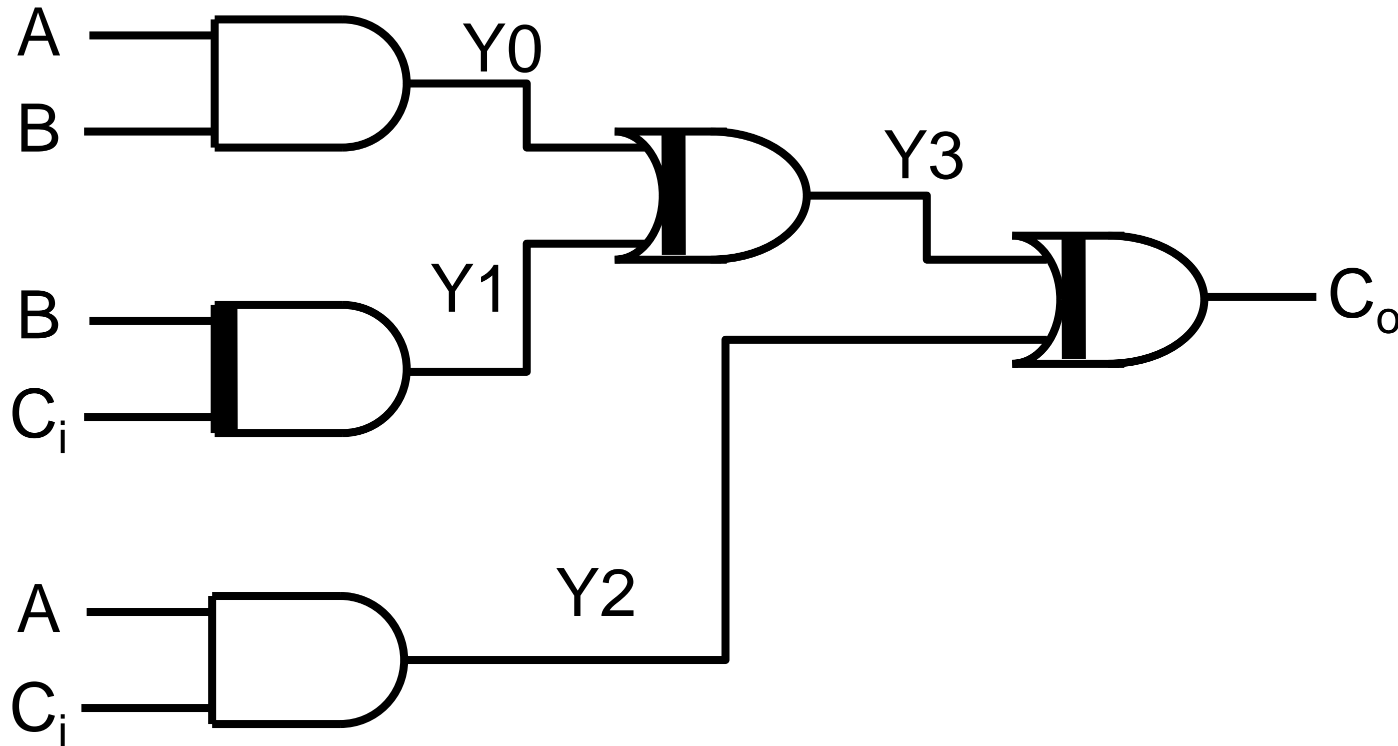
B

V_{SSV}

$V_{SS} - \Delta V$

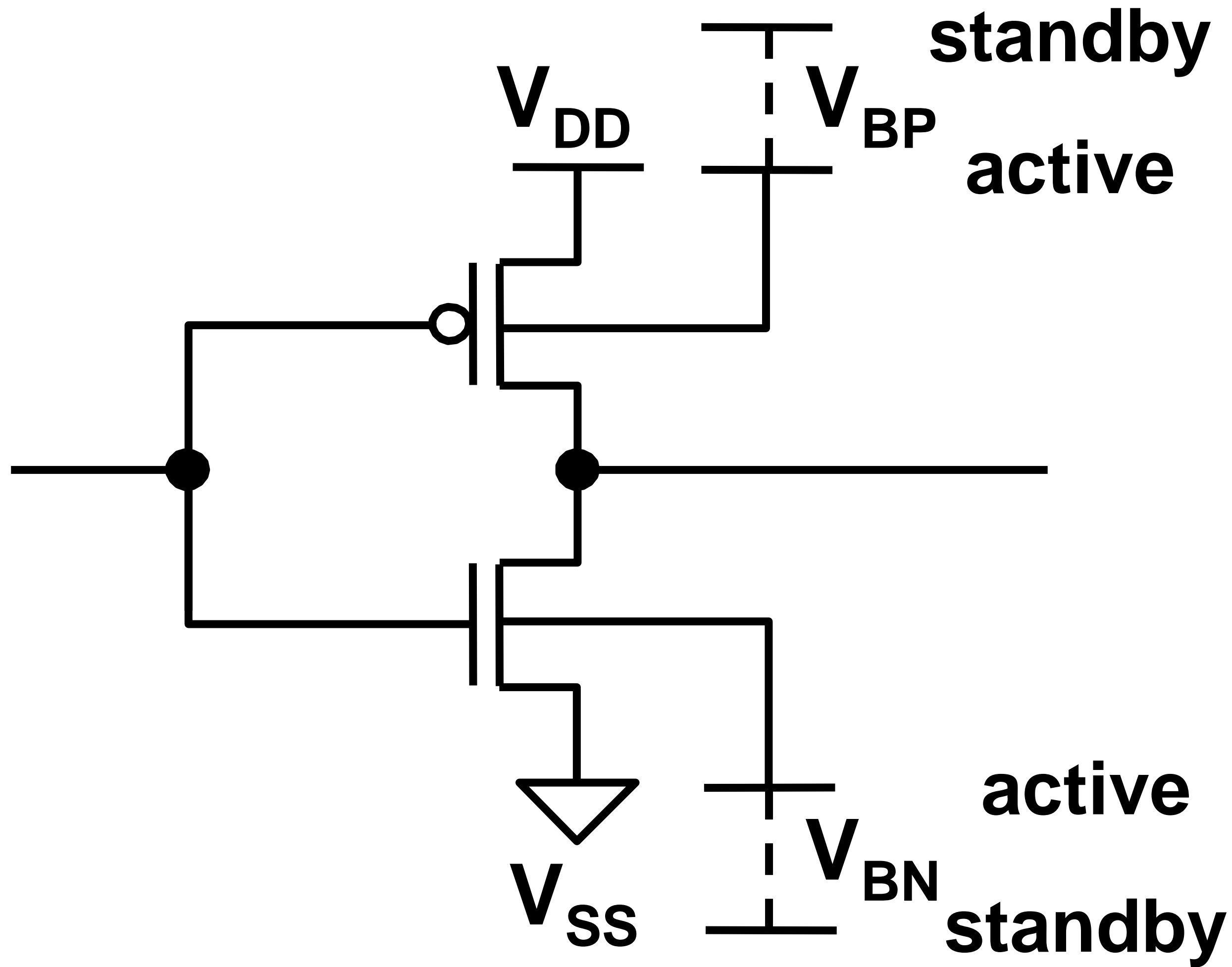
N0

Dual Threshold Datapath



- **Assign high V_T devices to non critical path gates (e.g.,Y0 = A * B for Full Adder carry logic)**
- **Use low V_T in critical path (e.g., carry in from preceding adder stages)**

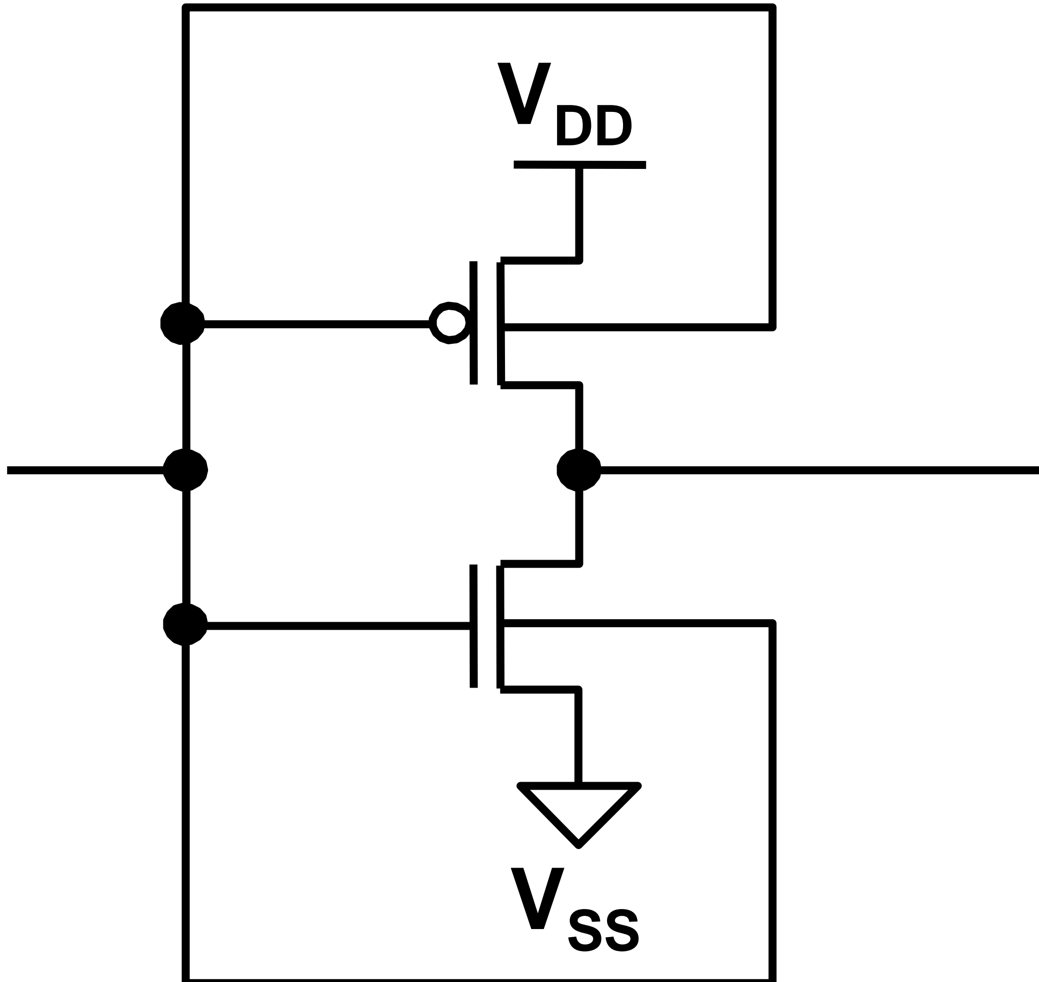
Variable Threshold CMOS



Variable Threshold CMOS Design

- **Body biasing technique**
  - Self-substrate bias circuit used to control body bias and adjust threshold voltage
- **Active mode: nearly zero applied bias**
  - Slightly forward substrate bias can increase speed in active mode
- **Standby mode: deep reverse bias applied**
  - Increases threshold voltage
  - Reduces subthreshold leakage current
- **Routing body net adds to overall area**

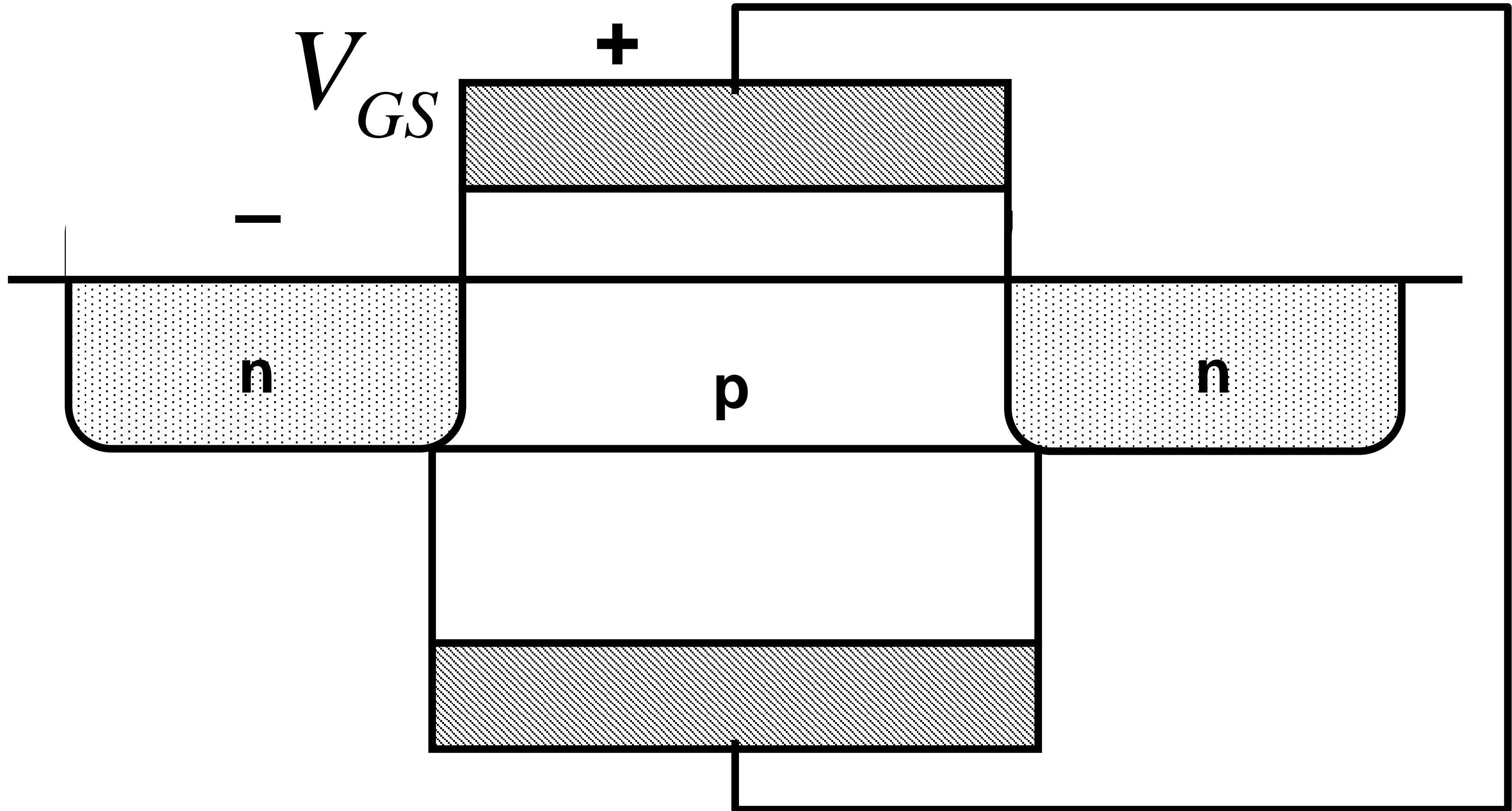
Dynamic Threshold CMOS



Dynamic Threshold CMOS Design

- **Threshold voltage adjusted dynamically with operating state of circuit**
  - Want high threshold in standby mode for low subthreshold leakage
  - Want low threshold in active mode for high drive current
- **Implement by tying body terminal to input**
  - Requires triple well technology in bulk CMOS
  - Supply voltage limited by diode built-in potential (source-body pn diode should be reverse biased)
  - Ultra low supply voltage (V_{DD} < 0.6 V)
- **Stronger advantages in partially depleted SOI**

Dual Gated SOI MOSFET

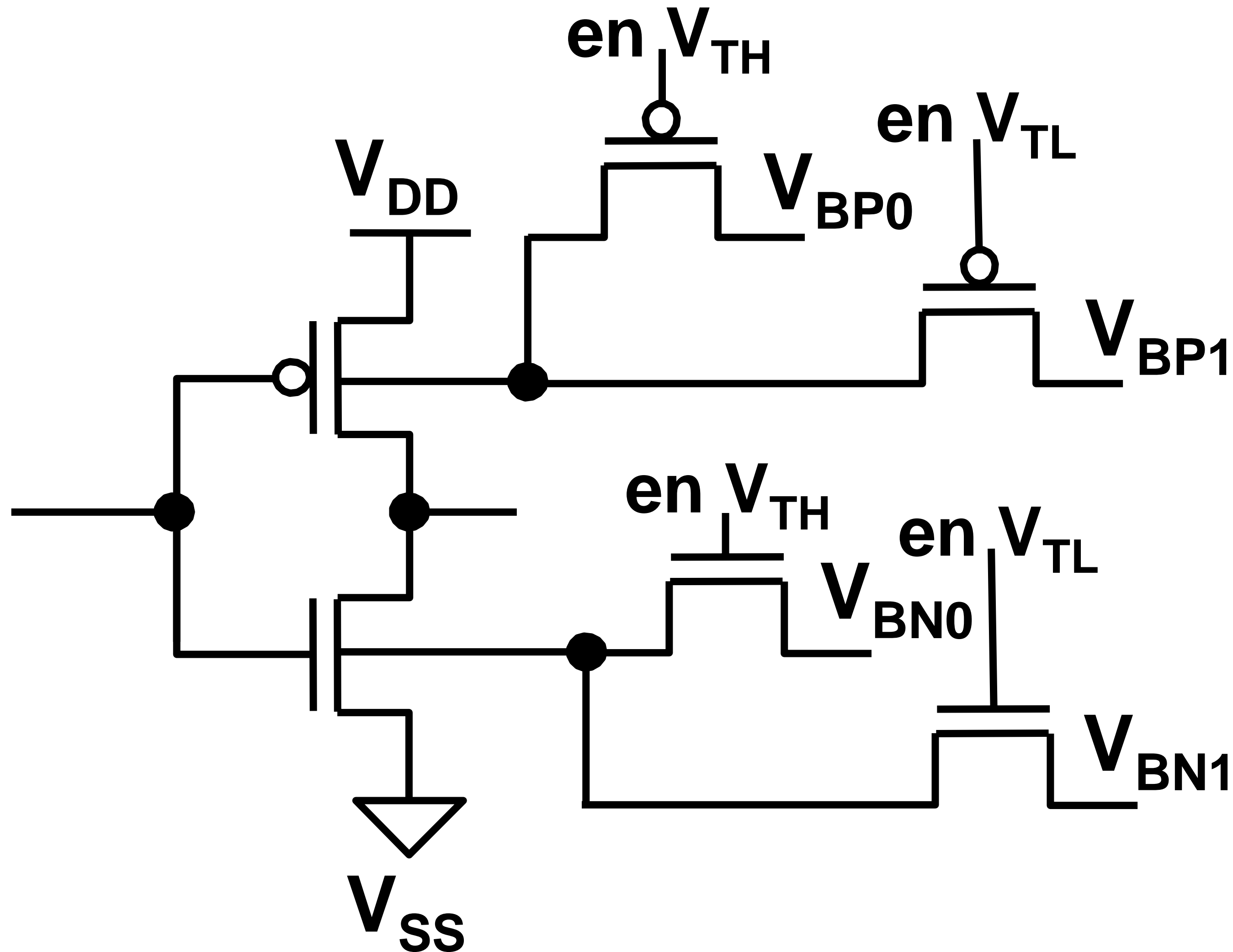


- **Double gate for dynamic threshold adjustment**

DG Dynamic Threshold SOI Design

- **Asymmetrical double gate SOI MOSFET**
  - Back gate oxide thicker than front gate oxide
  - Threshold voltage of back gate larger than supply voltage
  - Front gate threshold voltage changes dynamically with back gate voltage
- **Nearly ideal symmetric subthreshold characteristics**
- **Power delay product better (smaller) than symmetric double gate SOI CMOS**

Threshold Voltage Hopping Scheme



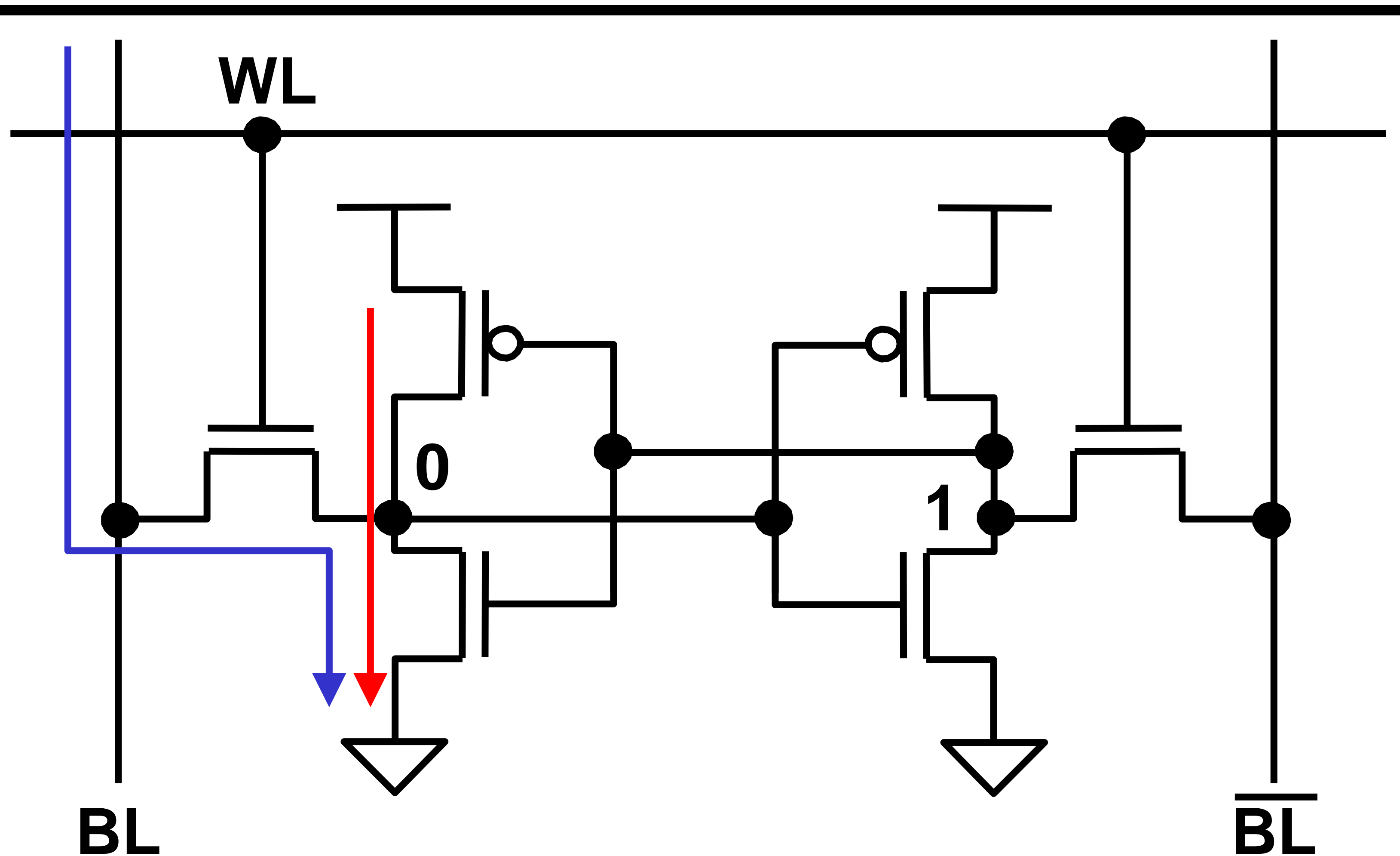
Outline

- Announcements
- Review: Energy Recovery Circuits
- Finish Lecture 8
- Leakage Mechanisms
- Circuit Styles for Low Leakage
- **Cache SRAM Design Examples**
- Next Topic: Subthreshold Circuits

Leakage Power Significant for Caches

- **Large fraction of current processor die devoted to memory structures (caches, TLB, etc.)**
  - Alpha 21264: 30% of area, StrongARM 60%
- **Caches account for large component of total leakage power**
- **At 130 nm process node:**
  - Leakage equals 30% of total L1 cache energy
  - Leakage accounts for 80% of total L2 cache energy
- **Explore techniques for reducing leakage power in caches (SRAMs in general)**

SRAM Cell Leakage Paths

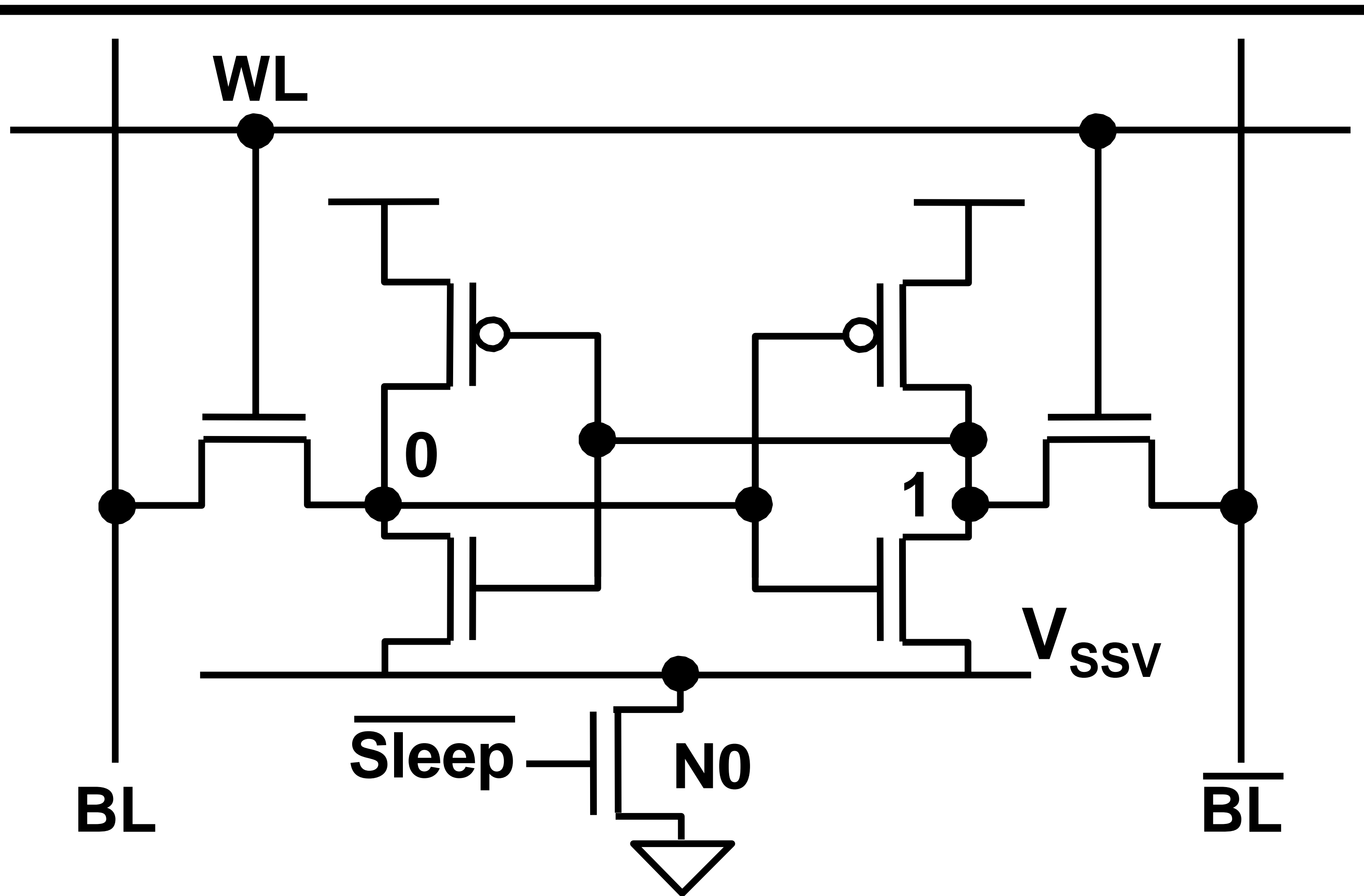


- **Dominant leakage paths: bitline to GND, V_{DD} to GND**

Gated Ground SRAM Cell Operation

- **Extra NMOS device gates power on and off**
  - Can be shared among all cells in a row, amortizing area and power overhead
  - Sizing strongly affects power, performance, and data retention capability of SRAM cell
  - Must be large enough to sink read/write current and maintain SRAM state, but if too large reduces stacking effect and increases leakage
- **Word line decoder controls bottom NMOS**
- **Turning off NMOS device cuts off leakage**
  - Also allows virtual ground to float higher in standby, reducing noise immunity
  - Simulate to verify SRAM data maintained in standby

Data Retention Gated-Ground SRAM Cell



- **Sleep device turned ON in active rows, OFF in inactive**

Drowsy SRAM Cell Operation

- **Idea: during access, SRAM cell in high power and performance mode, otherwise in low power "drowsy" mode**
- **Approach: switch between low and high V_{DD}**
  - Rely on short channel effects to decrease leakage at low supply voltage
- **Implement using two high V_T PMOS devices to switch between supplies**
  - High V_T required to reduce leakage between supplies
  - Requires a separate V_{DD} mux for each cache line
- **Scale V_{DD} to about 1.5 times V_T and still maintain state (0.3 V in 70 nm)**

Drowsy SRAM Cell

WL

V_{DDV}

0

1

BL

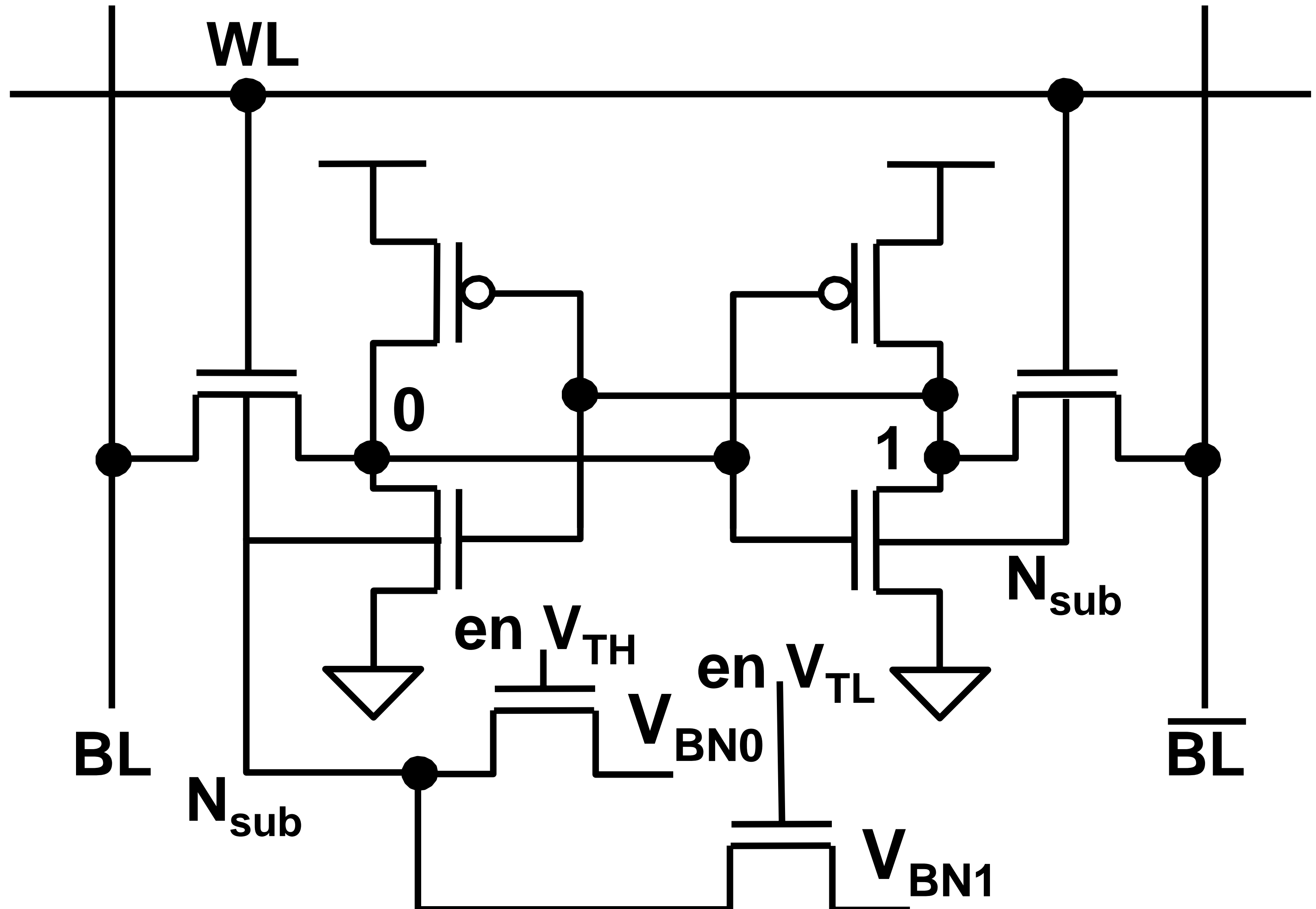
$\overline{BL}$

- **Virtual V_{DD} switched by high V_T PMOS devices**

Dynamic Threshold SRAM Cell Operation

- **Idea: during access, SRAM cell in high power and performance mode (low V_T), otherwise in low leakage (high V_T) mode**
- **Approach: switch between low and high body bias**
  - Body bias at 0 V for high speed
  - Body bias negative to increase V_T and cut leakage
- **Energy for single substrate transition greater than saved in leakage for a single clock cycle**
  - Body bias updates must be at larger time increments
  - Exploit spatial and temporal locality to bias one cache line at a time based on program accesses

Dynamic Threshold SRAM Cell



Conclusions

- **Leakage is an increasingly important component of total power dissipation**
  - A variety of physical mechanisms cause leakage currents
  - Subthreshold conduction probably most important
- **Many proposed circuit techniques to deal with it**
  - Multiple thresholds available in process: use stacked devices or low leakage series devices
  - Adjust thresholds using bulk terminal dynamically
- **Techniques starting to appear in commercial designs, especially for large memories such as on-chip cache**