Lecture 6

Leakage and Low-Power Design

R. Saleh
Dept. of ECE
University of British Columbia
res@ece.ubc.ca
Methods of Reducing Leakage Power

- So far we have discussed dynamic power reduction techniques which result from switching-related currents
- The transistor also exhibits many current leakage mechanisms that cause power dissipation when it is not switching
- In this lecture, we will explore the different types of leakage currents and their trends
- We will then describe ways to limit various types of leakage
- We will also re-examine the DSM transistor in more detail as a side-effect of this study
- Readings:
  - Sections of Chapter 2 and 3 in HJS
  - Many books and papers on DSM leakage power
  - Alvin Loke Presentation (SSCS Technical Seminar, 2007)
Basic CMOS Transistor Structure

- Typical process today uses twin-tub CMOS technology
- Shallow-trench isolation, thin-oxide, lightly-doped drain/source
- Salicided drain/source/gate to reduce resistance
- Extensive channel engineering for $V_T$-adjust, punchthrough prevention, etc.
- Need to examine some details to understand leakage
Sources of Leakages

- Leakage is a big problem in the recent CMOS technology nodes
- A variety of leakage mechanisms exist in the DSM transistor
- Actual leakage levels vary depending on biasing and physical parameters at the technology node (doping, tox, $V_T$, W, L, etc.)
Relative Importance of Leakage Currents

- Relative contributions of OFF-state leakage (but magnitude of total leakage getting exponentially worse for deeper submicron nodes)

![Diagram showing current paths: \(I_{SUB}\), \(I_{GIDL}\), \(I_J\), \(I_G\)]

- Subthreshold leakage from source
- Gate-induced drain leakage (GIDL)
- Junction reverse-bias leakage
- Gate leakage (direct tunneling)


But is this really true? Need to examine each one and their trends...
Hot carriers

- Assume gate and drain are connected to $V_{DD}$
- Carriers pick up high energy from electric field as they move across channel – become “hot” carriers which are attracted to gate node
  - These “hot” carriers may be injected into the gate oxide where they become trapped – cause a shift in the $V_T$
  - Accumulation of charge in oxide causes shift in $V_T$ over time
- The higher the $V_{DD}$, the hotter the carriers (more current)
- Since we have scaled $V_{DD}$, the problem was under control for years
- However, the $V_{DD}$ value may not scale in the future so this problem may again be an issue

Use lightly-doped drain to reduce hot-electrons
Source/Drain Leakage

- Source and drain junctions are normally reverse-biased so they will leak current.
- Typically very small but may increase with scaling since doping levels are very high in future technologies (breakdown voltage decreases as doping increases – use LDD to reduce BV).

![Cross-section Diagram]

Look at cross-section

- nMOS: n+ to p substrate
  - Substrate must be p
- pMOS: p+ to n substrate
  - Substrate must be n

![Graph]

$I_{S/D}$ (μA/um)

- 10nm
- 100nm
- 1000nm

- 1
- $1E^{-2}$
- $1E^{-4}$
- $1E^{-6}$
- $1E^{-8}$
Thin-Oxide Gate Tunneling

- \( t_{\text{ox}} \) has been scaling with each technology generation
- We have reached the point where \( t_{\text{ox}} \) is so small the direct tunneling occurs \( (t_{\text{ox}} < 2\text{nm}) \)
- Gate leakage = \( f(t_{\text{ox}}, V_G) \)

NMOS leakage is 3-10X PMOS leakage (electrons vs. holes)
- Below 20 Å, the leakage increases by 10X for every 2Å in gate thickness reduction
High-k Metal Gate

Traditional Oxide

High-k Metal Gate

Experimental Data of an NMOS
Subthreshold Leakage

- Subthreshold leakage is the most important contributor to static power in CMOS
  \[
  I_{sub} = I_s \cdot e^{-\frac{q(V_{GS} - V_T - V_{offset})}{nKT}} \left(1 - e^{-\frac{-qV_{DS}}{KT}}\right)
  \]
  \[P_{static} \approx I_{sub} V_{DD}\]
- Note that it is primarily a function of \(V_T\)
- Higher \(V_T\), exponentially less current!

- But gate overdrive \((V_{GS} - V_T)\) is also a linear function of \(V_T\)
- Need to understand \(V_T\) in more detail to find ways to reduce leakage
Lecture 6

Threshold Voltage Equation

MOS Fundamentals

- $V_T = \text{FET ON voltage, i.e., gate voltage required to form inversion layer connecting source & drain by shorting out back-to-back pn-junctions with substrate}$

\[
V_T = V_{FB} + 2\phi_b + \frac{Q_{dep}}{C_{ox}}
\]

- depletion charge per unit area
  \[
  qN_A x_{dep} \propto \sqrt{N_A} \quad (x_{dep} \propto 1/\sqrt{N_A})
  \]

- bulk potential
  \[
  \phi_b = \frac{k_B T}{q} \ln \frac{N_A}{n_i}
  \]

- oxide capacitance per unit area
  \[
  \varepsilon_{ox} / t_{ox}
  \]

- flatband (offset) voltage due to oxide charge & work function difference

- silicon surface
- poly gate
- n\textsuperscript{-} inversion layer
Drain-Induced Barrier Lowering

- Since the channel must be depleted of charge before inversion takes place, any help on depletion process will reduce $V_T$
- Large $V_{ds}$ => large depletion layer around drain
- Part of channel surface already depleted
- Lowered barrier => $V_T$ reduced => increased leakage current
DIBL

- For long-channel device, the depletion layer width is small around junctions so $V_T$ does not change noticeably.
- For short-channel devices, as we increase $V_{DS}$, the depletion layer will continue to increase and help to reduce the $V_T$.
- $V_T$ will continue to decrease as depletion layer thickness grows.

- If source and drain depletion regions merge -- Punch-through occurs!

![Diagram showing $V_T$ vs $V_{DS}$ for long-channel and short-channel devices.](image)
Effect of DIBL

- As $V_{DS}$ is increased, the current goes up (shift of graph up)
Gate-Induced Drain Leakage (GIDL)

- Drain-to-substrate leakage due to band-to-band tunneling current in very high field depletion region in gate-drain overlap region
- Caused by thinner oxides, lightly-doped drains and high $V_{DD}$
Short Channel Effect (SCE)

- $V_T$ rolloff at shorter $L$ since less charge must be depleted to achieve surface inversion

- But $V_T \uparrow$ as $L \downarrow$ is observed… why?

Reverse Short Channel Effect (RSCE)
Reverse Short-Channel Effect (RSCE)

- p-type impurities gather at edges of source and drain and accumulate at point defects in this region during oxidation.
- As $L$ decreases, $V_T$ rises before conventional SCE kicks in.
- p-type impurities gather at edges of source and drain and accumulate at point defects in this region during oxidation.
- Cause $V_T$ to increase to compensate for extra charge.
- Actual curve depends on whether SCE or RSCE is dominant.
I_{on} vs. I_{off} for 90nm CMOS

- "No free lunch" principle prevails again: high $I_{ON} \rightarrow$ high $I_{OFF}$
- $V_T$'s not scaling as aggressively as $V_{DD}$ (1.0V in 90nm & 65nm)
- Technology providers offer variety of $V_T$'s on same die to concurrently meet high-speed vs. low-leakage needs
Leakage Reduction – Long $L_{\text{eff}}$

- Allow two different minimum sizes: nom-$L_{\text{eff}}$ and long-$L_{\text{eff}}$
- Most library cells are available in both flavors
- long-$L_{\text{eff}}$ are $\sim 10\%$ slower but have 3X less leakage
- Use nom-$L_{\text{eff}}$ for critical paths and long-$L_{\text{eff}}$ for non-critical paths
- Do not increase area of source/drain regions, just increase $L$

• Why does this work? effect of DIBL is reduced.
Leakage Reduction – Body-bias

- Called VTCMOS (variable threshold CMOS)
- Threshold voltage of both devices are increased by adjusting the body-bias voltage in order to reduce subthreshold leakage current in standby mode

- Requires twin-tub technology so that substrates of individual devices can be adjusted
Leakage Reduction – Body-Bias

- Devices get slower when $V_T$ is higher
- Set bias to obtain low-$V_T$ devices on critical path and high-$V_T$ devices on non-critical paths and SRAMs
- As substrate bias increases, pn junction breakdown will occur so this places a limit on the voltages that can be used
- Optimal value of reverse bias continues to decrease since doping levels continue to increase and breakdown voltage of pn junctions decrease (especially for NMOS device)
- This may not be as useful in future technologies
Leakage Reduction - Multiple $V_T$ Libraries

- Deep submicron libraries provide three types of transistor $V_T$’s for NMOS and PMOS devices
  - LVT = low threshold voltage (high speed)
  - SVT = standard threshold voltage (compromise)
  - HVT = high threshold voltage (low leakage)
- Place LVT cells along critical path
- Place SVT or HVT cells along non-critical paths and SRAM arrays
- Typical distribution in microprocessors (IBM P5)
  - SVT (65%), HVT(33%), LVT(2%)
User Higher $V_{DD}$ and $V_T$ for Memory

- $V_{DD}$: Low (0.5V)
- $V_{T}$: Low (0V)
- High-perf.

- $V_{DD}$: High (1V)
- $V_{T}$: High (0.3V)
- Low-leakage

- 0.5V 400MHz 16bit processor
- 3.5mW
Reducing Leakage Power – Power Gating

Called “MTCMOS”

Logic circuit

Thin $t_{OX}$
Low $V_T$
Low $V_{DD}$

Leakage cut-off switch

High $V_T$
High $V_{GS}$
Thick $t_{OX}$

$V_{DD}$

$V_{SSV}$

$V_{GS}$

<Stand-by> <Active>
1.5V
0.5V=$V_{DD}$
0V
Issues in MTCMOS

- Virtual ground not actual ground (lose some noise margin)
- NMOS leaks more than PMOS
- Can increase width of sleep transistor to reduce voltage at virtual ground but it will also increase subthreshold leakage and area of sleep transistor
- Sleep time must be long enough warrant its use
- Wakeup power must be offset by reduced leakage power
- Choose type, number and W of sleep transistors carefully
Virtual $V_{DD}$ drift
Use Zig-Zag Style Power Gating

- Allows gate to retain value, assuming input is known
- Need to force inputs to desired values
- Difficult placement/routing problem
Use Sleep Transistor in Each Gate

- Reduces effect of DIBL on each gate
- Put such gates in critical path and SVT, HVT cells in the non-critical paths
What is the right balance?

- Optimal dynamic/leakage power ratio is 70/30, [Kuroda, ICCAD 2002]
Summary

Major concern today is leakage power.
- Subthreshold leakage will continue to be a problem
- Power gating is a widely-used approach
- Thin-oxide gate leakage may be reduced at 45nm
- Substrate bias to adjust $V_T$ may lose its impact
- Junction leakage may be a problem soon
- Hot-carriers may become important again if $V_{DD}$ does not continue to scale with technology
- Need to strike proper balance dynamic and static power in a design to minimize power

Power and Energy reduction and recycling will continue to be dominant topics for the foreseeable future