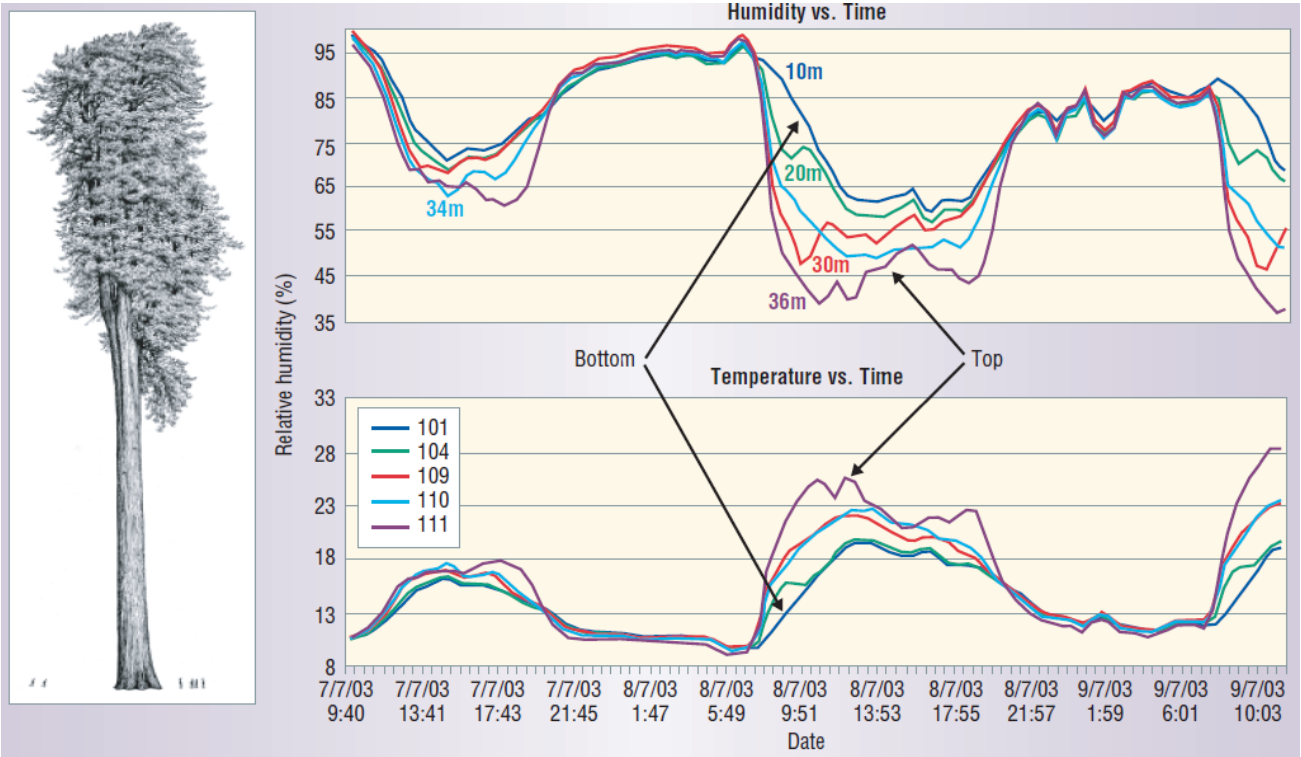
Static Power Reduction for Asynchronous Circuits

Carlos Tadeo Ortega Otero

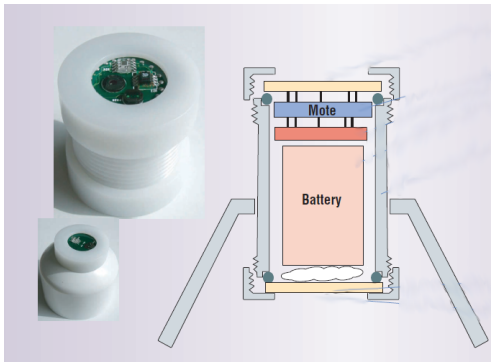
Asynchronous VLSI Group and Architecture
Computer Systems Laboratory
Cornell University
Ithaca, NY, 14853

May, 25^{th}, 2012

Low Duty-Cycle Power Constrained Designs



- Sensing every 5 minutes
- Battery life $>$ 1 year



*Culler, D . Overview of Sensor Networks

Low Duty-Cycle Power Constrained Designs

- Biological implants
- Neurological stimulators
- Cardiac rhythm management
- RFID tags
- Remote sensing devices

Low Duty-Cycle and Power Constrained

Overview

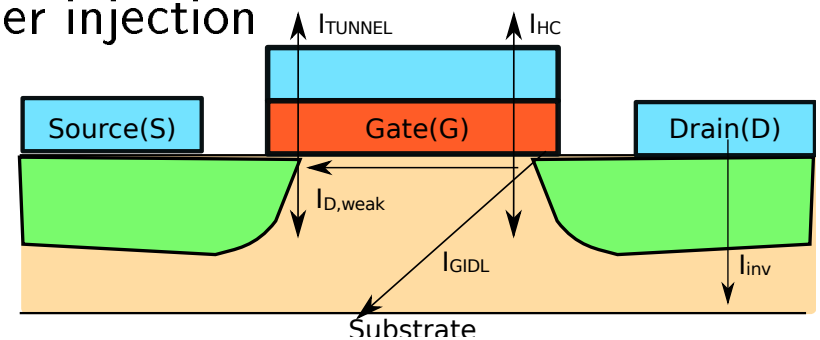
1. Review leakage mechanisms
2. Review static power reduction techniques
3. Review Power Gating techniques
   - Non-state preserving
   - State preserving

Our Contributions:

1. Power gating in the context of Async circuits
   - Non-state preserving
   - State preserving
2. Evaluate Power Gating in Async circuits
3. ZZDRTO - Pipeline Power Gating
4. Evaluate ZZDRTO

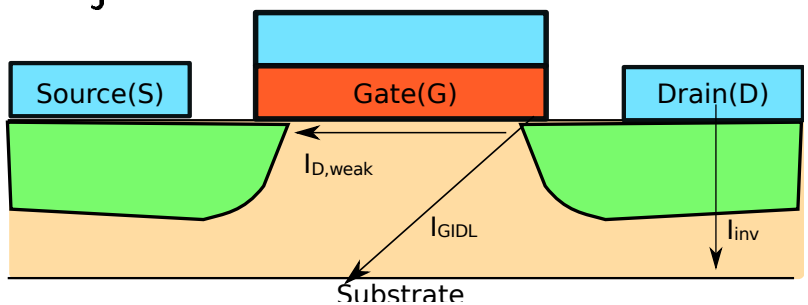
Leakage Mechanisms

- Leakage: Current that flows when the ideal current is 0A
- Source-Drain Leakage
  - Subthreshold Leakage
  - Reverse biased diode
  - Gate induced Drain leakage (GIDL)
- Gate Leakage
  - Direct Tunneling
  - Hot carrier injection



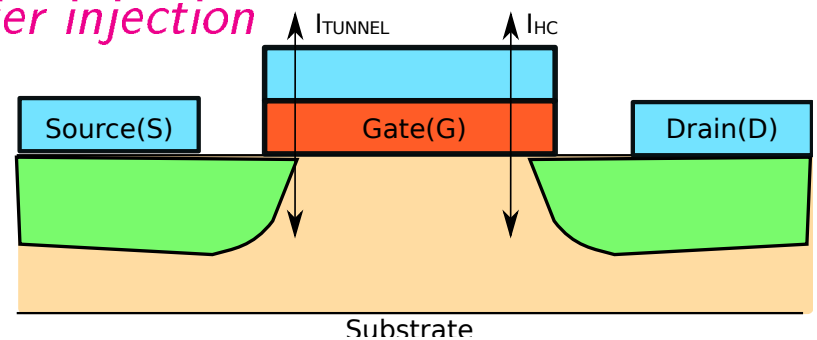
Leakage Mechanisms: Source-Drain

- Leakage: Current that flows when the ideal current is 0A
- *Source-Drain Leakage*
  - *Subthreshold Leakage*
  - *Reverse biased diode*
  - *Gate induced drain leakage (GIDL)*
- Gate Leakage
  - Direct Tunneling
  - Hot carrier injection

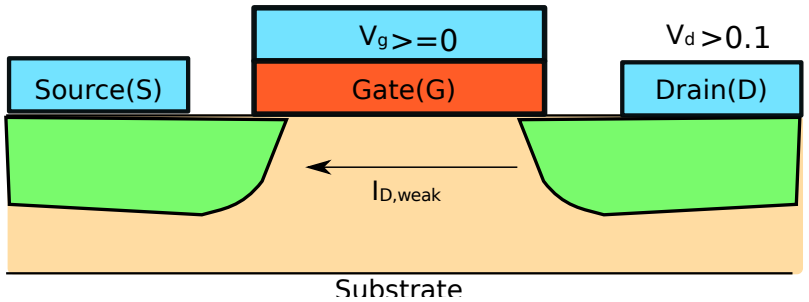


Leakage Mechanisms: Gate Leakage

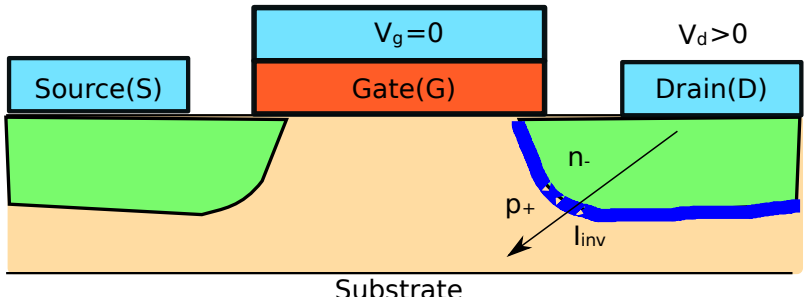
- Leakage: Current that flows when the ideal current is 0A
- Source-Drain Leakage
  - Subthreshold leakage
  - Reverse biased diode leakage
  - Gate induced Drain leakage (GIDL)
- *Gate Leakage*
  - *Direct Tunneling*
  - *Hot carrier injection*



Source-Drain Leakage



Subthreshold leakage (top)



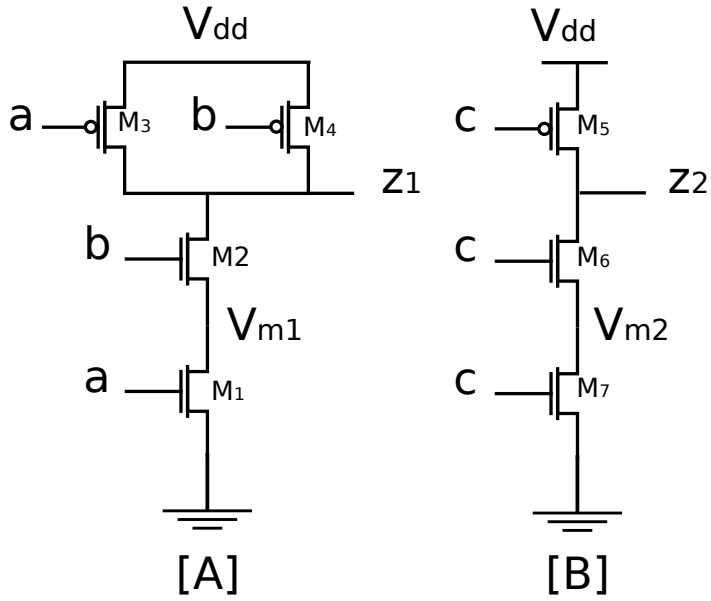
Reverse-biased diode leakage (bottom)

The Impact of CMOS Scaling on Static Power

- Leakage expected to increase as devices shrink
- Major concerns:
  - Gate Oxide Thickness (T_{ox}) scaling
  - Channel Miniaturization
  - V_{dd}, V_{th} scaling
  - Source-Drain punchthrough
  - Doping concentration
- Scientists work hard to keep static power manageable

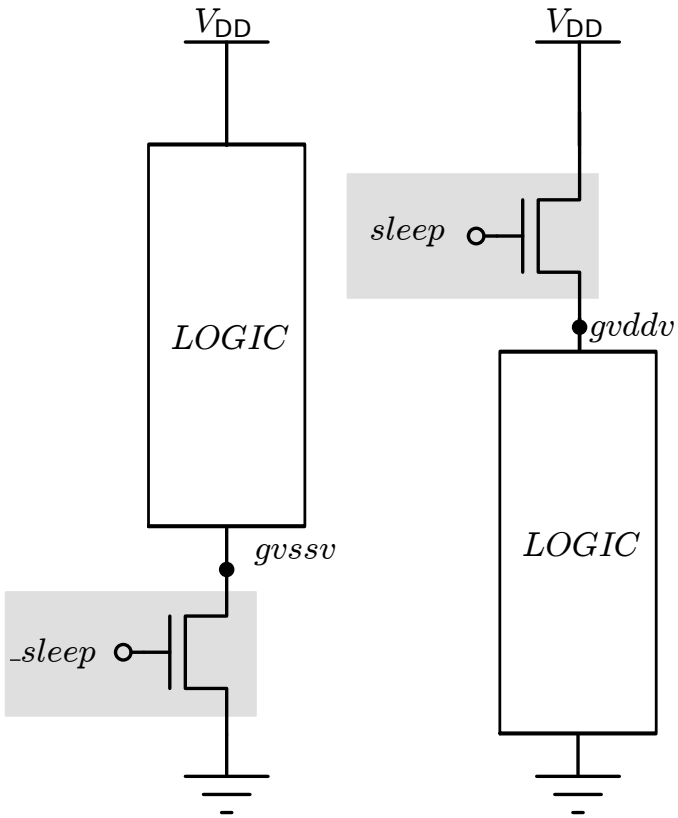
General Static Power Reduction Techniques

- Device Level: Active devices
  - Doping, materials, V_{dd} ,V_{th}
- Circuit Level. Gates
  - Natural stacks (Fig. A)
  - Forced stacks (Fig. B)
- System Level: macro blocks, datapaths and co-processors
  - Power Gating



Power Gating

- Series Resistance
  - Leakage Reduction
  - Performance Penalty
- Virtual Power Nets
  - gvddv
  - gvssv

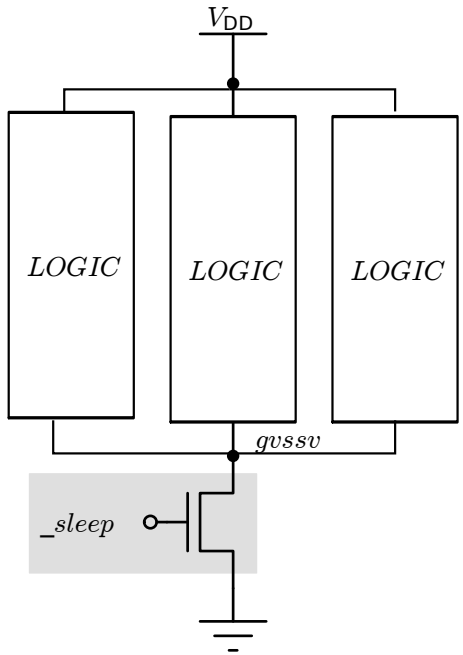


Power Gating Techniques

- Non-state preserving
  - Value of dynamic nodes drift towards power rails
- State preserving
  - Retain value of registers
  - Retain value of dynamic gates

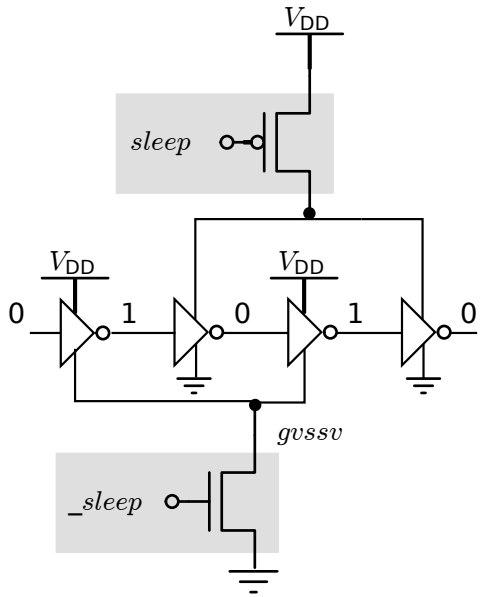
Non-State Preserving

- Cut-off (CO)
- Internal Nodes float
  - Foot transistor (Figure) N-type transistor
  - Head transistors P-type transistor
- Transient behavior
  - Long sleep settle time
  - Long wake-up time

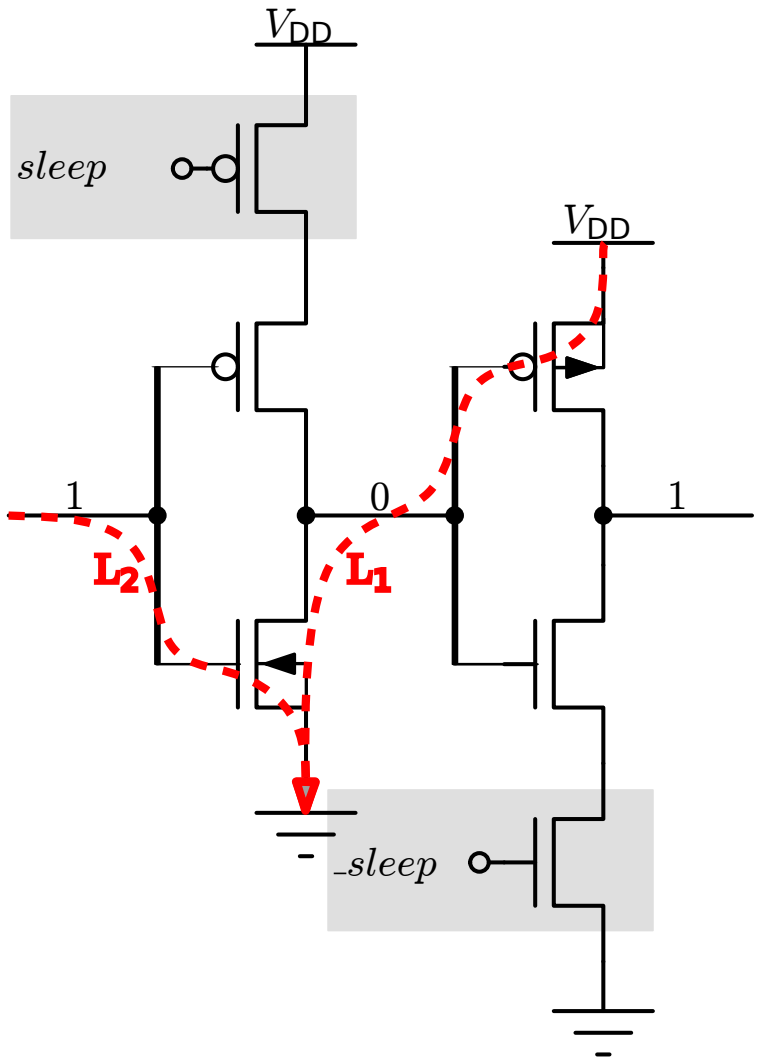


State Preserving

- Zig-Zag Cut Off (ZZCO)
- Select Head or Foot transistor
  - Head $\rightarrow$0
  - Foot $\rightarrow$1
- Better transient behavior vs cut-off
- Less effective for leakage reduction



State Preserving - Sneaky paths



Overview

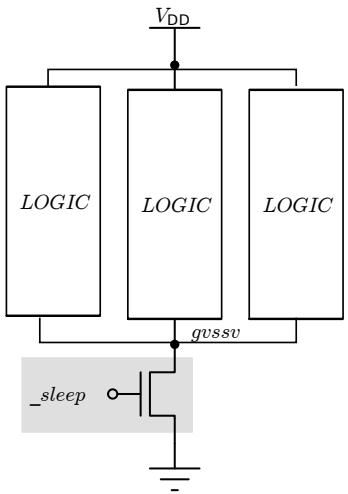
1. Review leakage mechanisms
2. Review static power reduction techniques
3. Review Power Gating techniques
   - Non-state preserving
   - State preserving

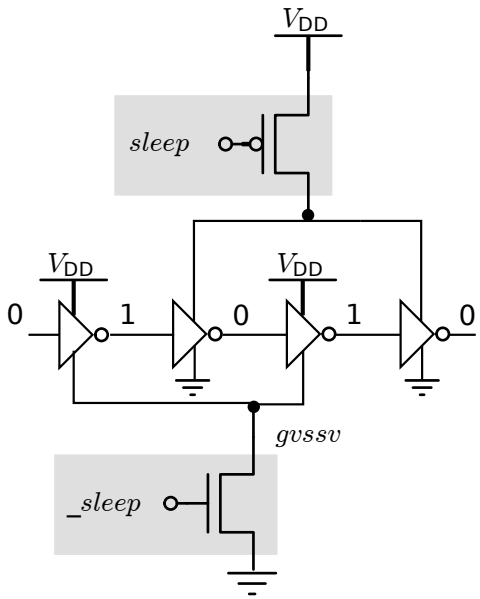
Our Contributions:

1. **Power gating in the context of Async circuits**
   - **Non-state preserving**
   - **State preserving**
2. **Evaluate Power Gating in Async circuits**
3. **ZZDRTO - Pipeline Power Gating**
4. **Evaluate ZZDRTO**

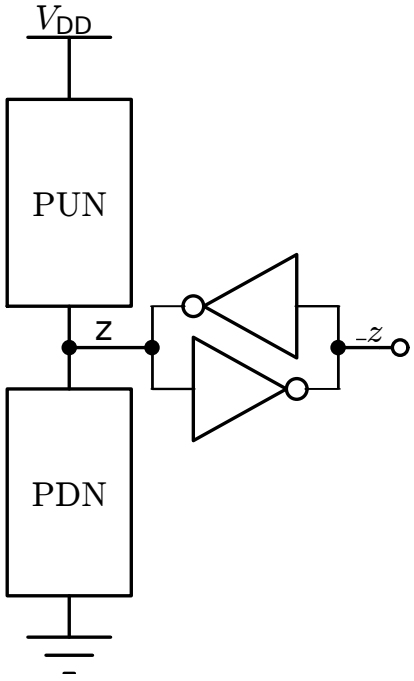
Asynchronous Power Gating Techniques

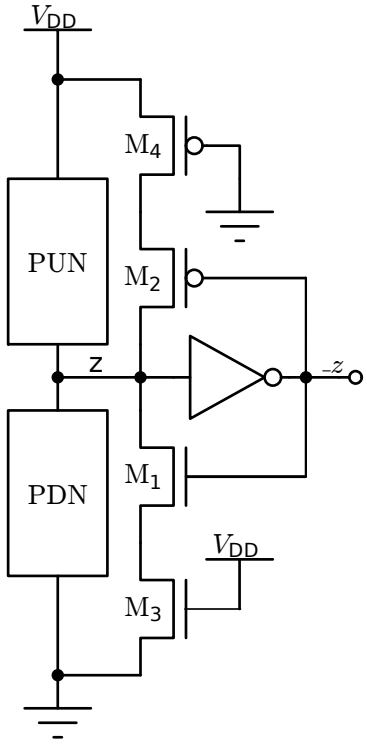
- Standard techniques work! ... mostly
- Conditions necessary for correct operation



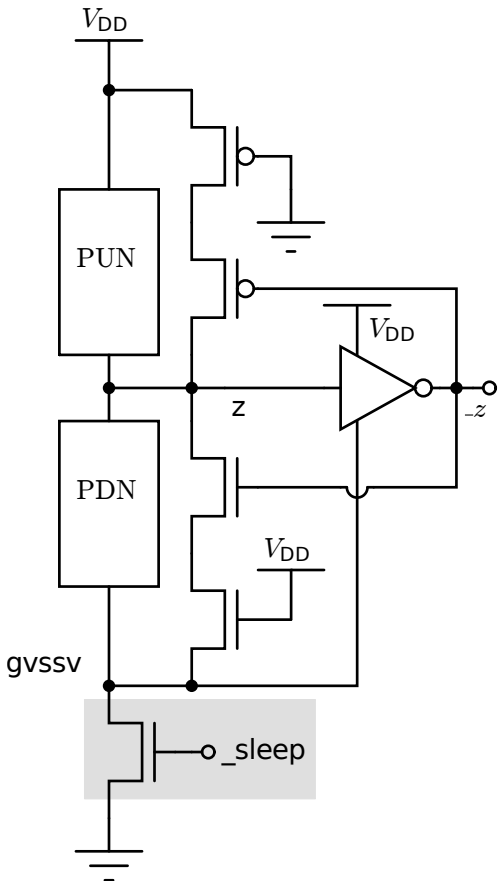


Pseudo-Static Gates





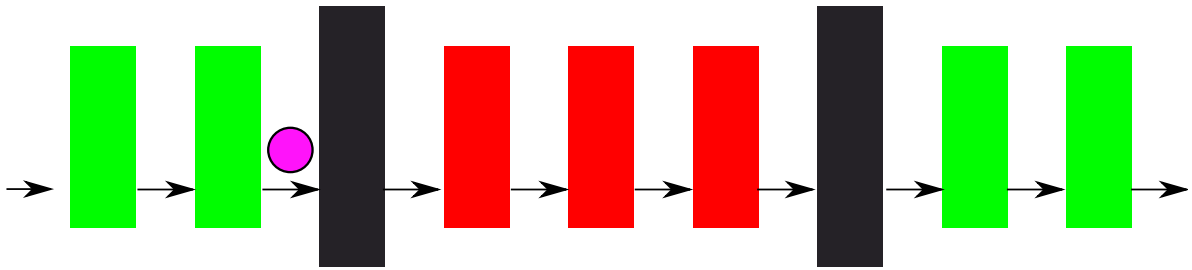
Async Non-State Preserving Power Gating



- Cut-Off (CO)
- Similar approach
  - Static CMOS gates
  - Pseudo-static gates

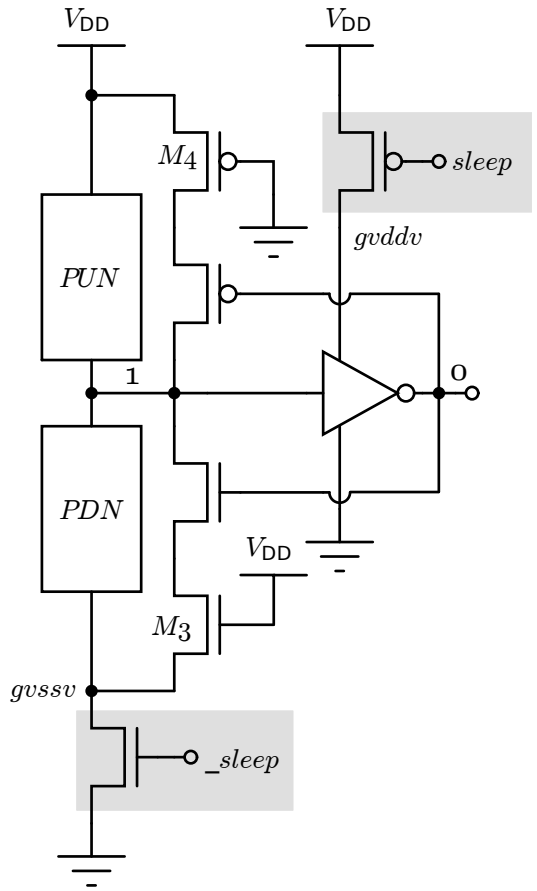
Async Non-State Preserving Power Gating

- Isolation Circuits
- Wake-up sequence
  - De-assert *sleep*
  - Exercise *reset* sequence
  - Assert *safe*



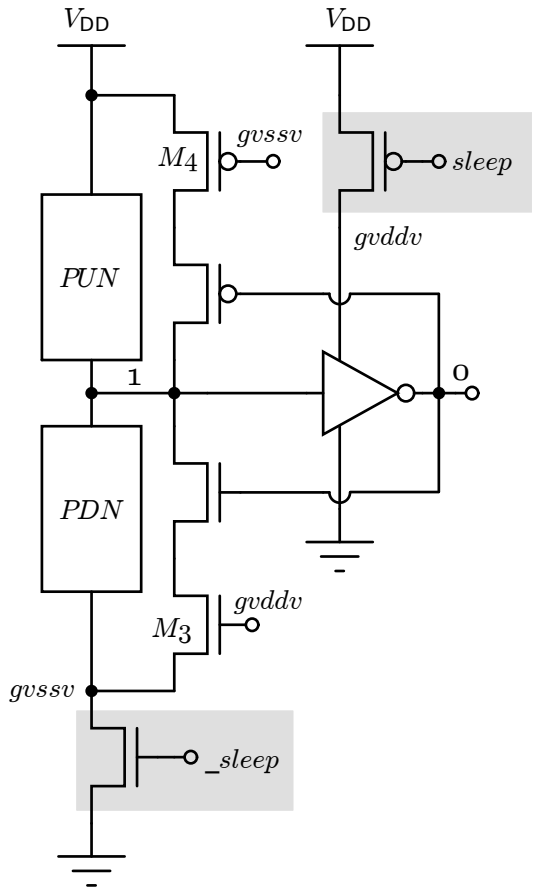
Async State Preserving Power Gating

- Zig-Zag Cut-Off (ZZCO)
- Same approach
  - Static CMOS Gates
  - Pseudo-static gates
- Note forward inverter

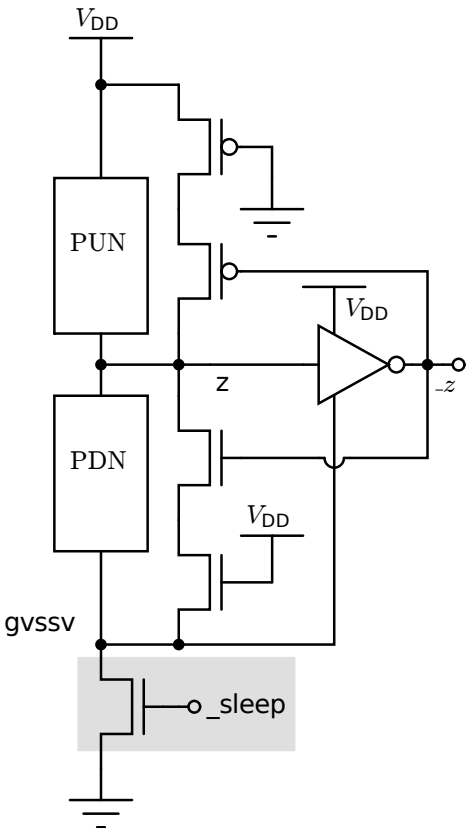


Async State Preserving Power Gating

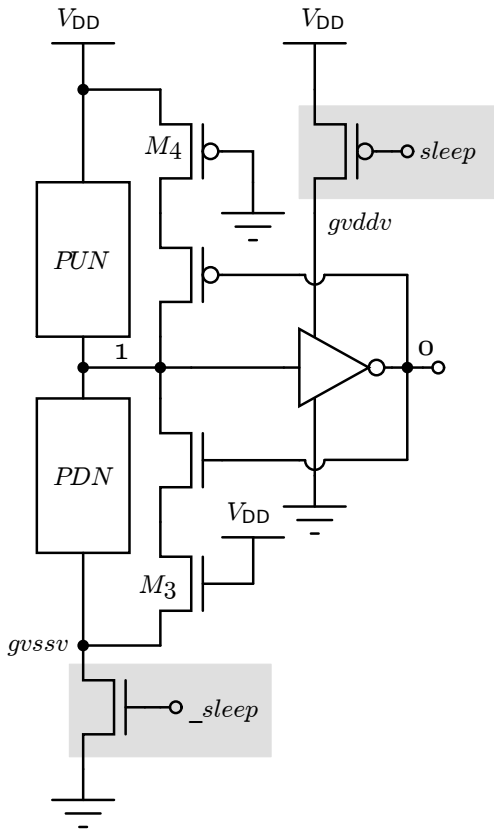
- Zig-Zag Cut-Off Weakened Statizicer (ZZCO-WS)
- Weakened Staticizer
  - gvddv instead of Vdd
  - gvssv instead of GND
  - Better power savings
  - Better performance
- Low-cost
  - No topology change
  - Similar wiring costs



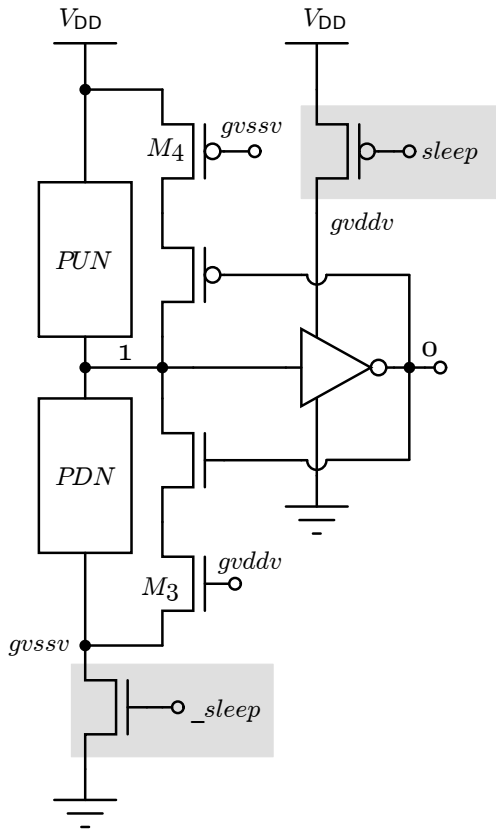
Power Gating Circuit Techniques: Evaluation



CO



ZZCO



ZZCO-WS

Power Gating Circuit Techniques: Steady State Evaluation

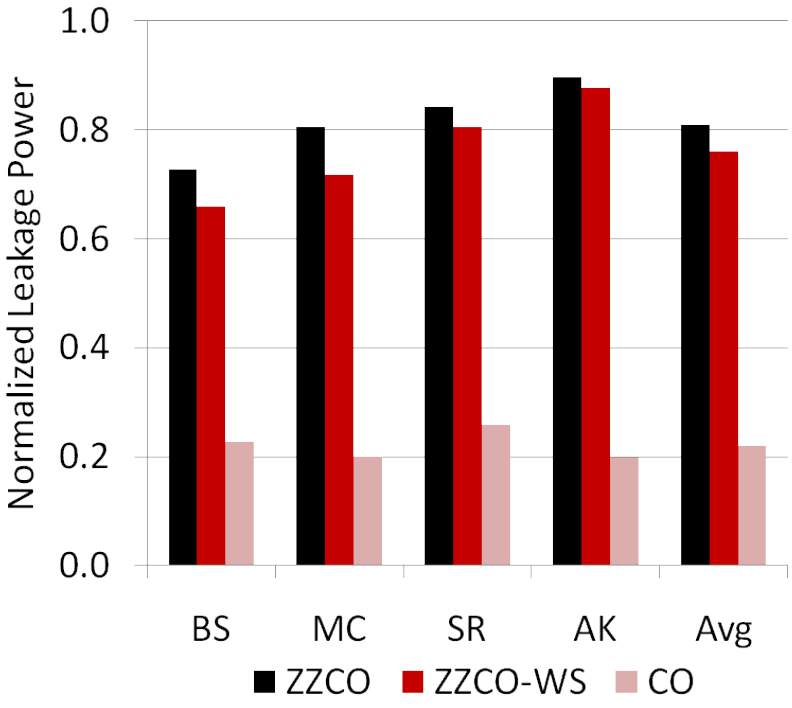
Pipeline Cluster	Transistors	FO4
Add Round Key (AK)	8400	2.4
Shift Rows (SR)	7567	2.6
Byte Substitute (BS)	84144	20.4
Mix Column (MC)	30000	16.8

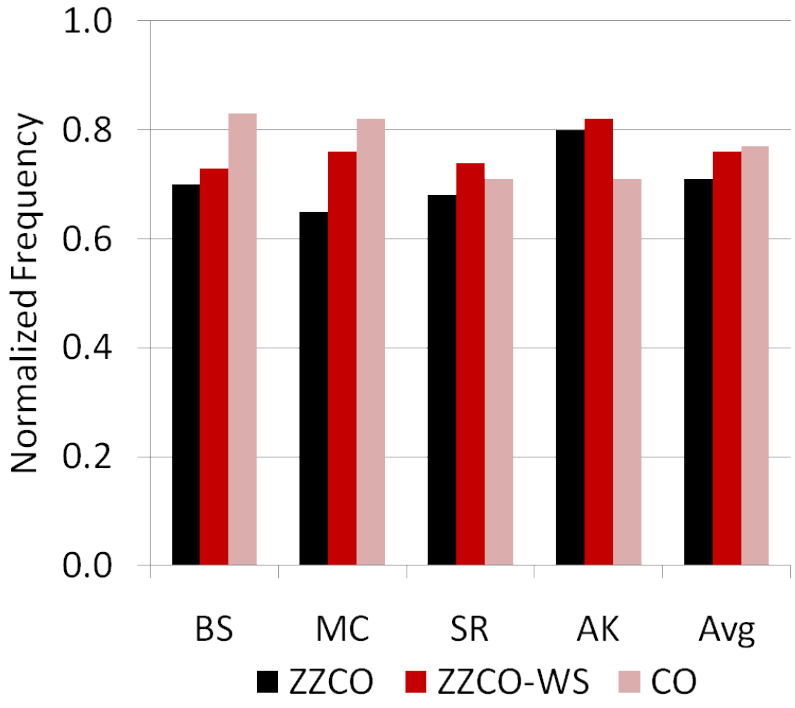
Circuit	Transistor Count
Control Circuitry	18000
Counter overhead	4300

Total Number of transistors: 153000

Power Gating Circuit Techniques: Steady State Evaluation

- BSIM, T-T, conservative wire cap, 90nm, 298K
- AES pipeline functional blocks





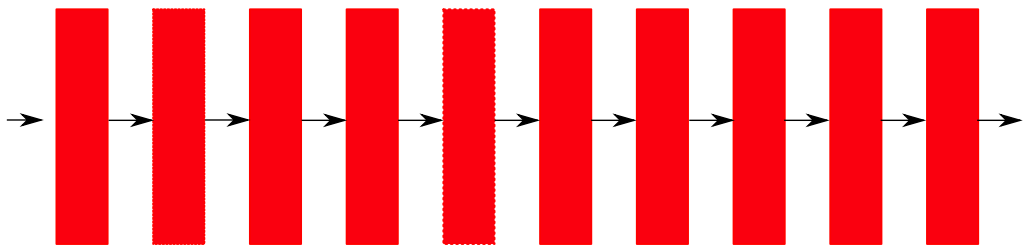
Power Gating Circuit Techniques: Evaluation

- CO offers the best steady state
  - Power savings
  - Performance
- ZZCO-WS is better in steady state than ZZCO
  - Power savings
  - Performance
  - Negligible implementation costs

Pipeline Power Gating

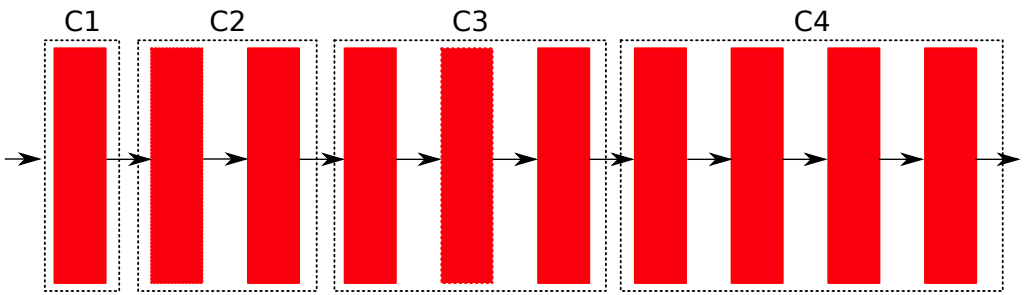
- Async: Self throttling circuits
- Control circuitry
  - Safe turn off (Empty pipeline detection)
  - Correct dynamic operation (isolation circuits)
  - Quick wake-up

Zero-Delay Ripple Turn On (ZDRTO)



Zero-Delay Ripple Turn On (ZDRTO)

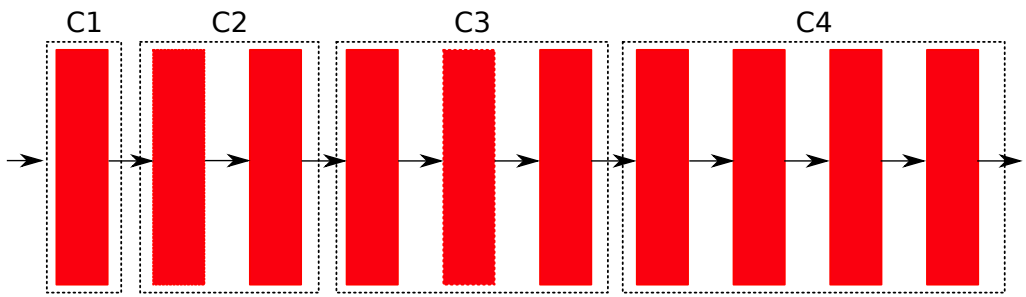
- Pipeline Cluster = Power Gating Domain



Pipeline Clusters

Zero-Delay Ripple Turn On (ZDRTO)

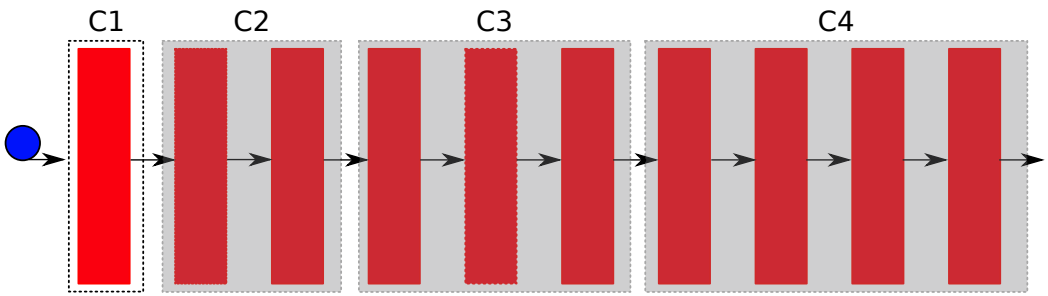
- Pipeline Cluster
  - Power Gating Domain
- Choose
  - Cluster size depends on the application
  - Power Gating technique for each cluster



Pipeline Clusters

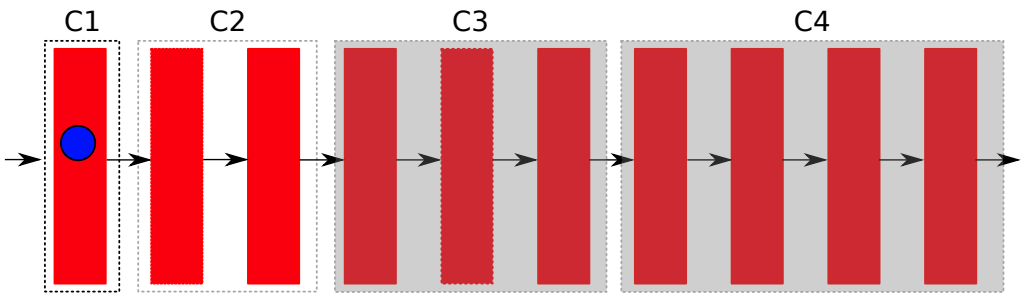
Zero-Delay Ripple Turn On (ZDRTO)

- Leverage pipeline stage computation latency
  - Hide latency of powering up downstream stages
- Leverage asynchronous circuits robustness
  - Do computation during power up



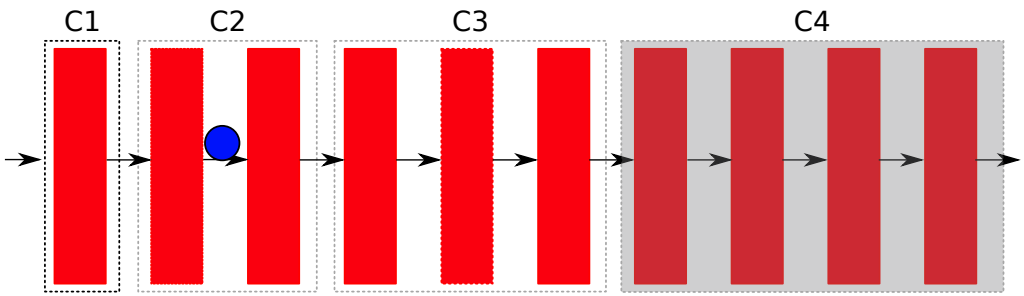
Zero-Delay Ripple Turn On (ZDRTO)

- Leverage pipeline stage computation latency
  - Hide latency of powering up downstream stages
- Leverage asynchronous circuits robustness
  - Do computation during power up - Domino effect turn on



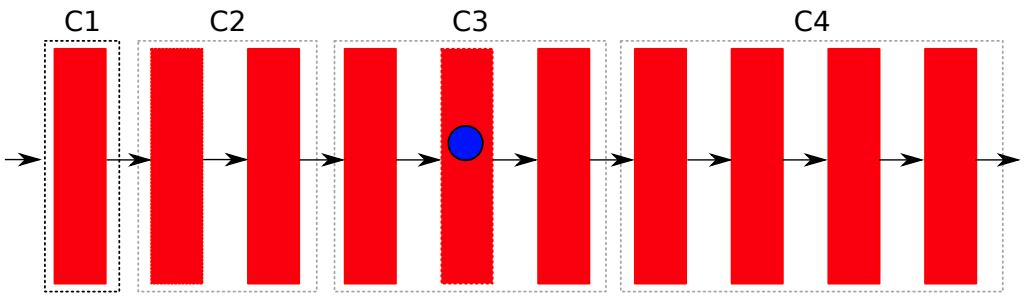
Zero-Delay Ripple Turn On (ZDRTO)

- Leverage pipeline stage computation latency
  - Hide latency of powering up downstream stages
- Leverage asynchronous circuits robustness
  - Do computation during power up - Domino effect turn on



Zero-Delay Ripple Turn On (ZDRTO)

- Leverage pipeline stage computation latency
  - Hide latency of powering up downstream stages
- Leverage asynchronous circuits robustness
  - Do computation during power up - Domino effect turn on



ZDRTO Evaluation

- Result trade-offs between
  - Wake-up latency
  - Power savings
  - Operating frequency
- BSIM4, T-T conservative Wire cap, 90nm, 298K
- Example pipeline: 4-cluster AES pipeline
- Different Power Gating techniques

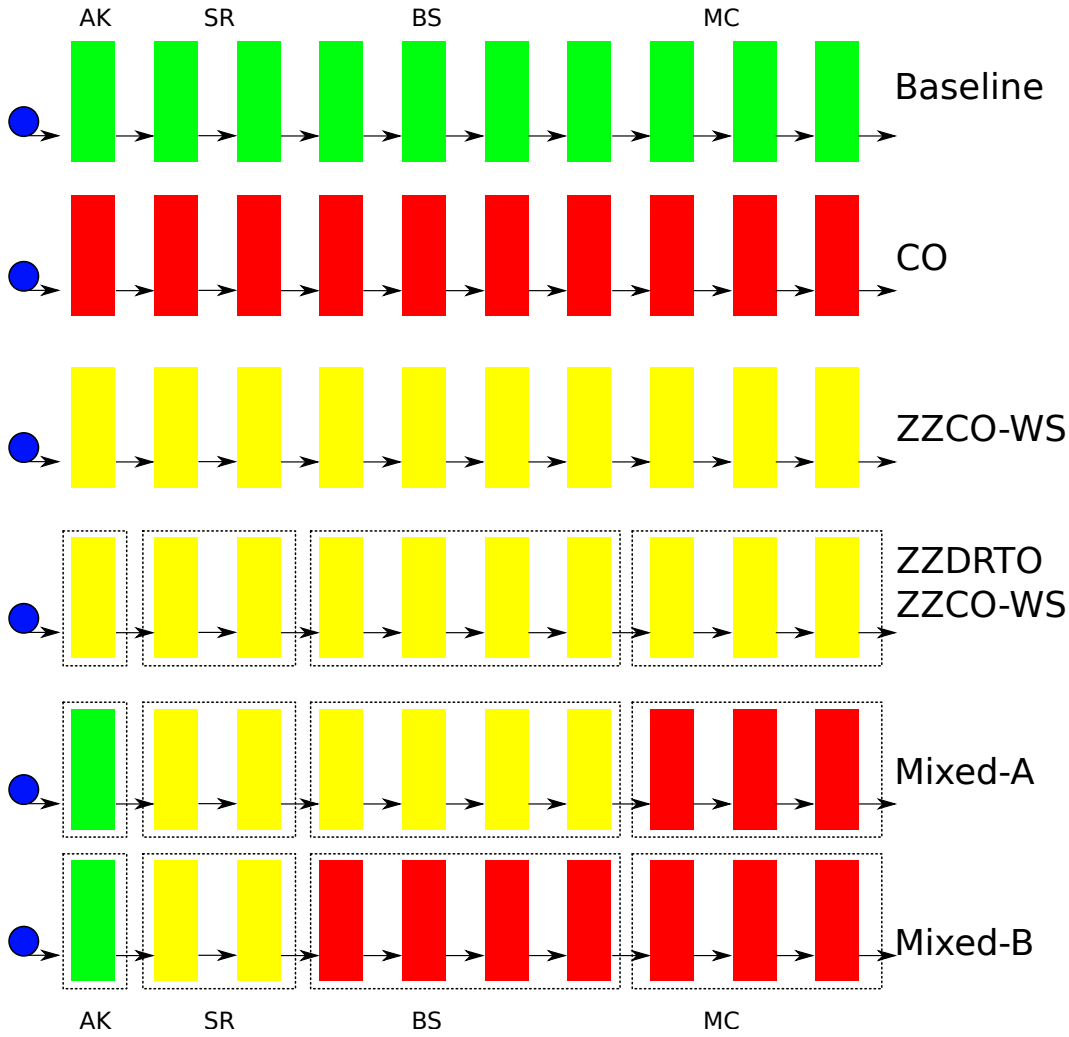
ZDRTO Evaluation

Pipeline Cluster	Transistors	FO4
Add Round Key (AK)	8400	2.4
Shift Rows (SR)	7567	2.6
Byte Substitute (BS)	84144	20.4
Mix Column (MC)	30000	16.8

Circuit	Transistor Count
Control Circuitry	18000
Counter overhead	4300

Total Number of transistors: 153000

ZDRTO Evaluation

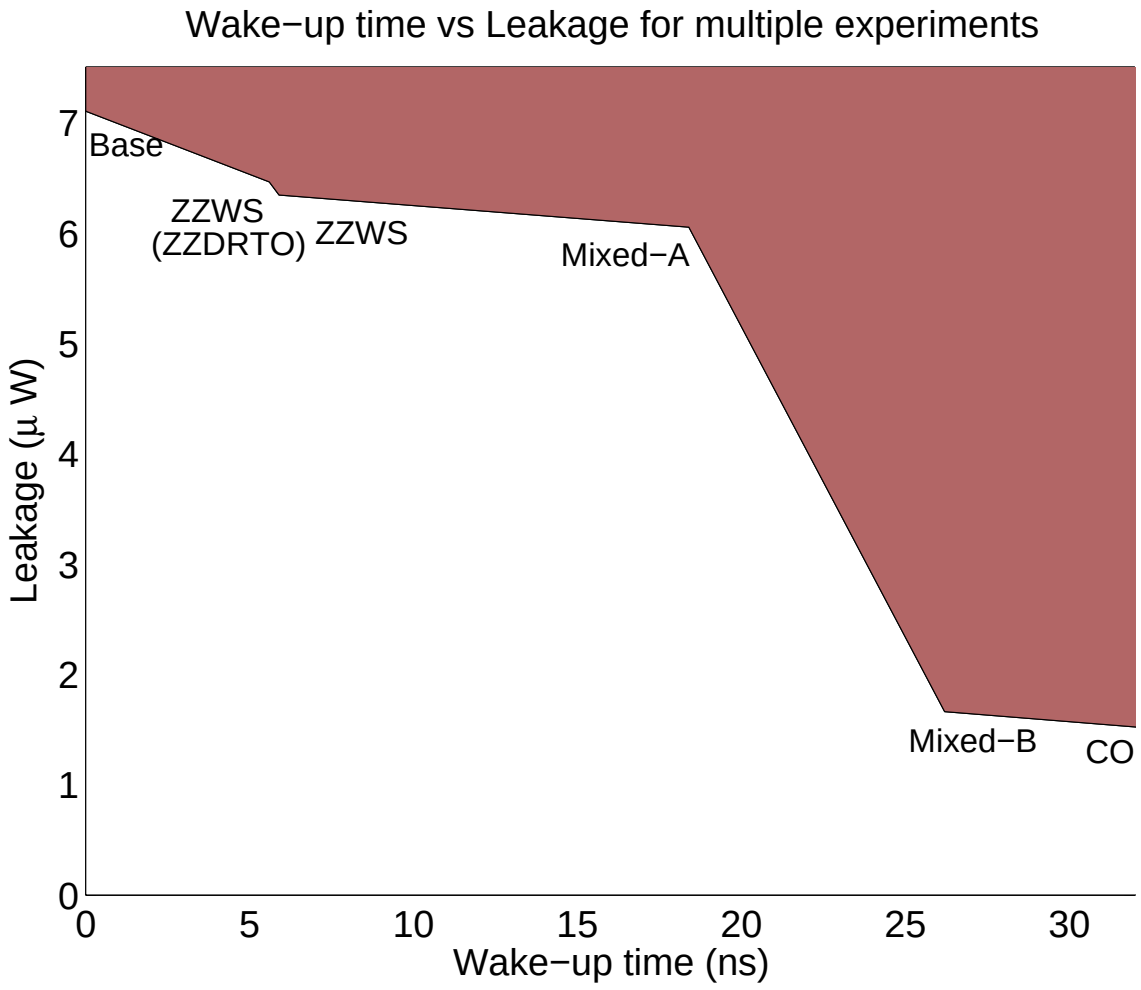


ZDRTO Evaluation

Non-ZDRTO	Wake-up(ns)	Leakage(μW)	Frequency (Mhz)
Baseline	0	7.1	285
CO	32.9	1.5	262
ZZWS	5.9	6.34	180

ZDRTO	Wake-up (ns)	Leakage(μW)	Frequency(Mhz)
ZZWS	5.6	6.46	182
Mixed-A	18.4	6.05	226
Mixed-B	26.2	1.62	260

ZDRTO Evaluation - Wake-up vs Leakage



Conclusions

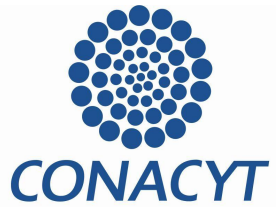
- Overview static power leakage mechanisms
  - Source-Drain leakage
  - Gate leakage
  - Analyzed the impact of miniaturization on several currents
- Overview methods to reduce static power
  - Device level
  - Circuit level
  - System level: Maximal gains with Power Gating

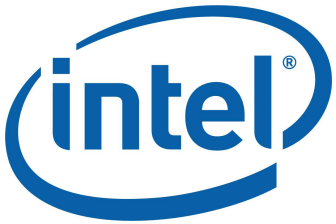
Conclusions

- Asynchronous Power Gating techniques
  - State preserving (Avg *25%* savings)
  - Non-state preserving (Avg *80%* savings)
- Pipeline Power Gating techniques
  - Empty pipeline detection
  - Zero-Delay Ripple Turn On (ZDRTO)
- Demonstrated trade-offs
  - Wake-up latency
  - Power savings
  - Operating frequency

Acknowledgements

- Prof Manohar, Prof Myers and Prof Suh
- Jonathan Tse
- λ-team: Rob Karmazin and Benjamin Hill
- Async Members: Nabil, Fang, Filipp, Basit, Ilya, Stephen
- CSL faculty and students
- Funding agencies:





Static Power Reduction for Asynchronous Circuits

Carlos Tadeo Ortega Otero

Asynchronous VLSI Group and Architecture
Computer Systems Laboratory
Cornell University
Ithaca, NY, 14853

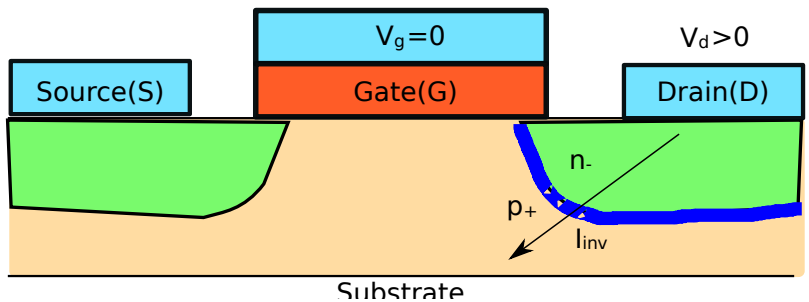
May, 25^{th}, 2012

Reverse-Biased Diode Leakage

$$I_{INV} = I_s \times (e^{(Vd/Ut)} - 1) \quad (1)$$

- Reverse biased diode current
- I_s =Reverse Saturation Current
- U_t= Temperature voltage

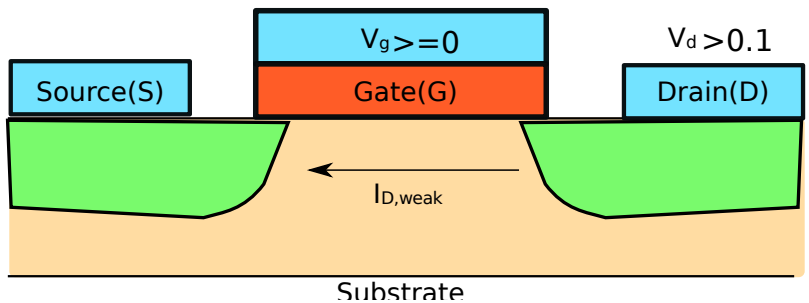
$$I_{INV} = A_d \times J_{INV} \quad (2)$$



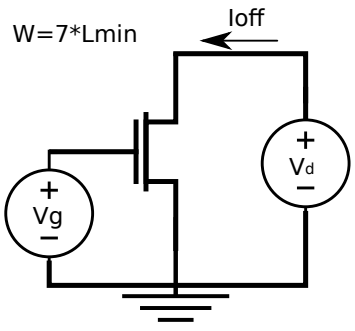
Source-Drain Leakage

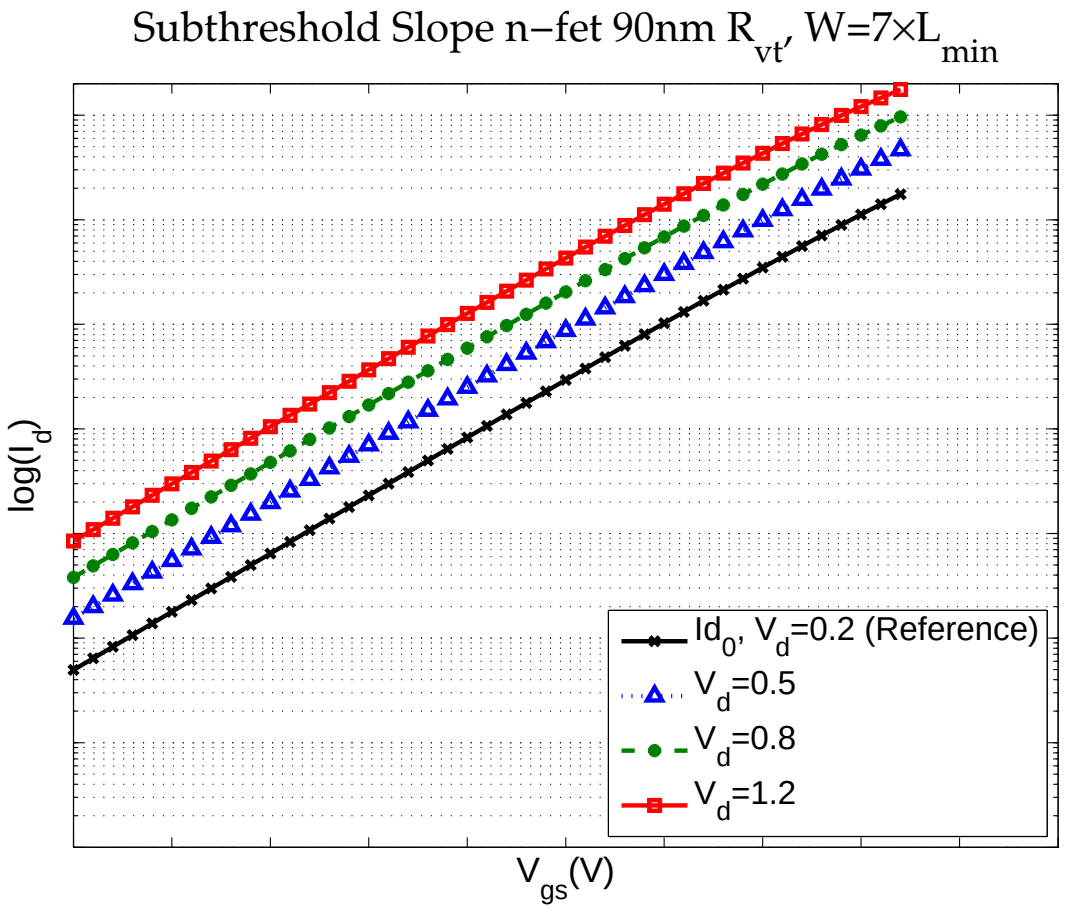
$$I_{d,weak} = \frac{W}{L} \times I_o \times e^{(V_{gs}-V_{th})} \times (1 - e^{V_{ds}(mU_T)-1}) \qquad (3)$$

- Condition $V_g < V_{th,}|V_d| \geq 0.1$ *and* $Vs = Vb = 0$
- V_{gs} =Voltage Gate-Source - exponential dependence
- V_{ds} =Voltage Drain-Source - linear dependence
- U_T =Thermal Voltage

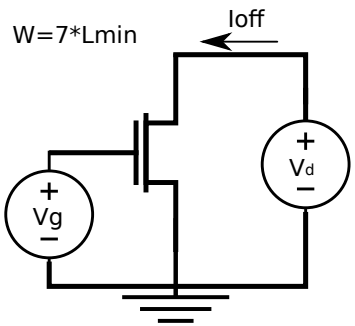


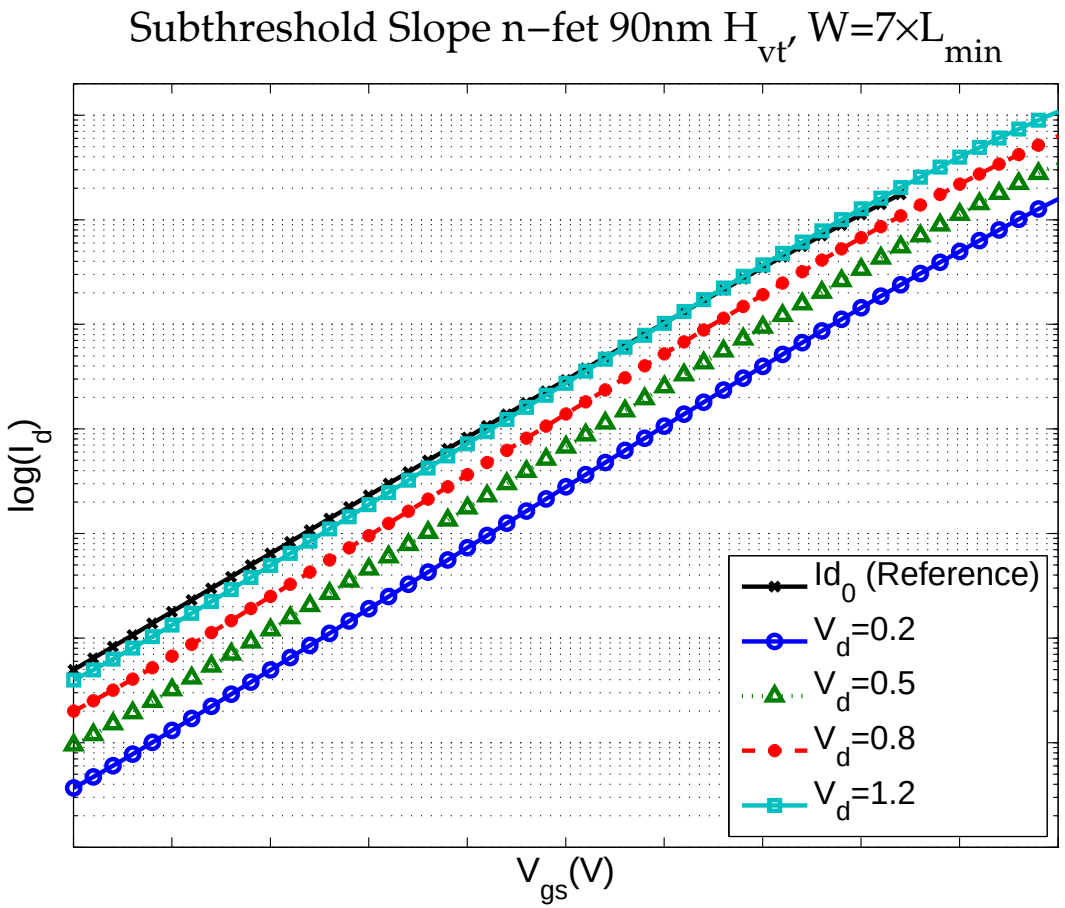
Source-Drain Leakage: Subthreshold Slope-RVT - 90nm



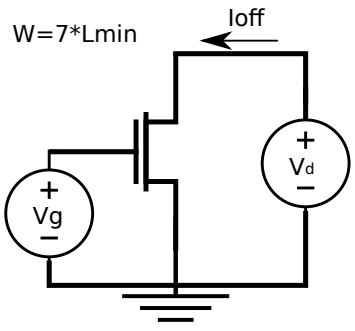


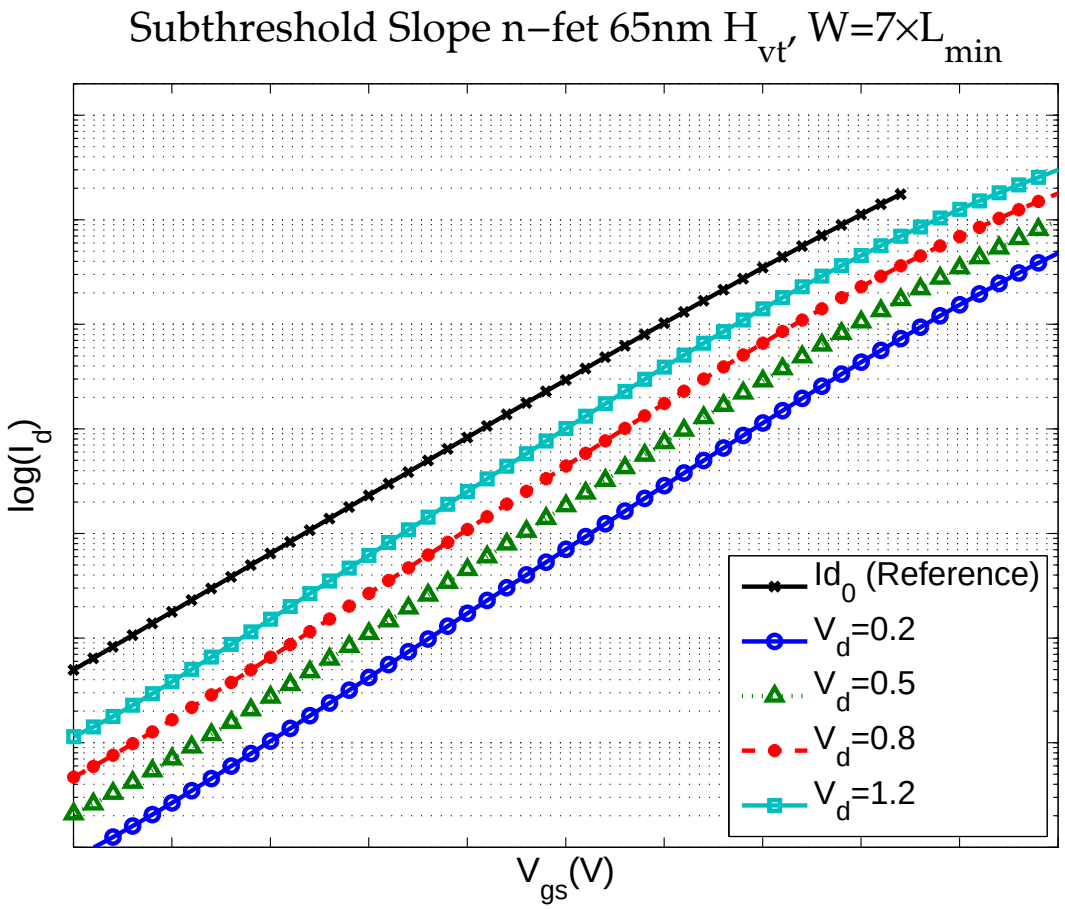
Source-Drain Leakage: Subthreshold Slope-HVT - 90nm





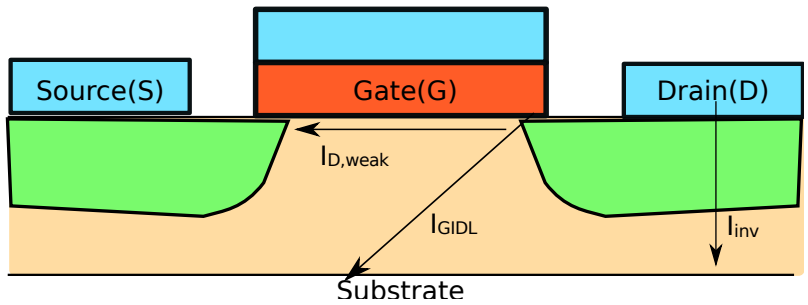
Source-Drain Leakage: Subthreshold Slope-HVT - 65nm





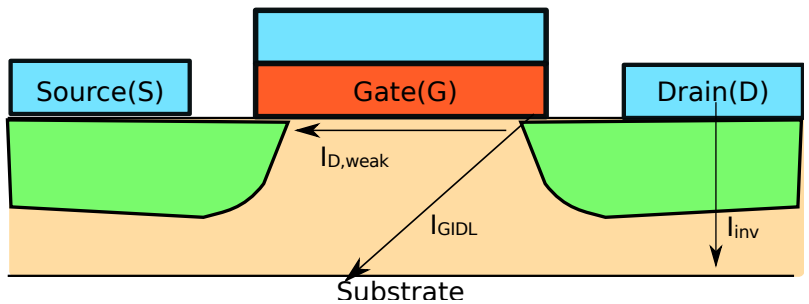
Direct Tunneling

- Dependent on large voltages V_{gd}
- Gate to Source-Drain-channel-body
  - Electron tunneling
  - Hole tunneling



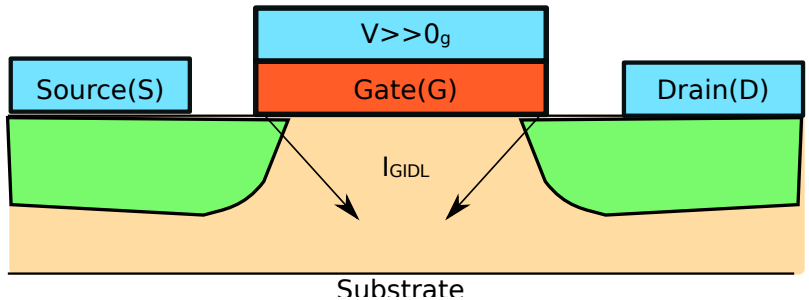
Hot Carrier Injection

- Dependent on large voltages V_{gd}
- Gate-to-Drain tunneling



Gate-Induced Drain current

- Dependent on large voltages V_{gd}
- Gate-to-Drain tunneling



CMOS Scaling: Gate Oxide Thickness Scaling

- $L_{eff} = 45 \times T_{ox}$(Intel)
- What happens at $100nm$?
- $T_{OX} \sim$ 12Amstrongs to16Amstrongs
  - GIDL. Imposes a limit on T_{OX} as Electric field increases significantly
  - GIDL. Less relevant as voltages reduce below the energy band gap of the Silicon
  - Gate Direct Tunneling.

Channel Miniaturization

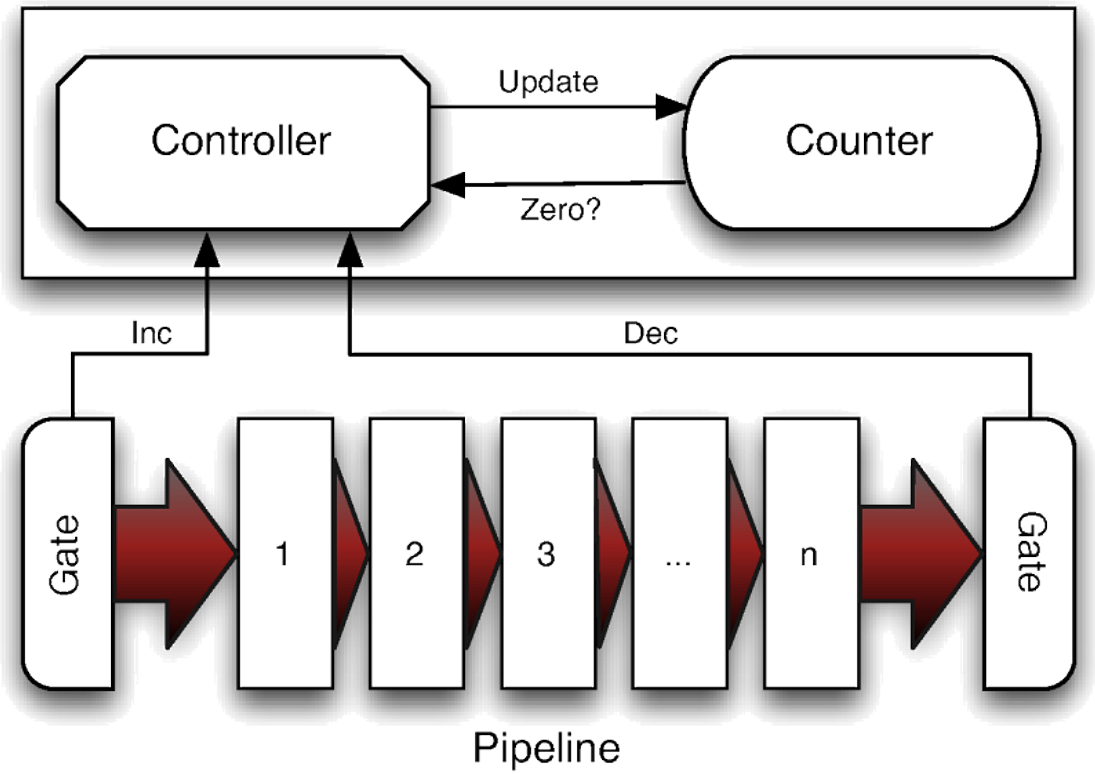
- Add here

Source-Drain *punchthourgh*

- Not a concern f

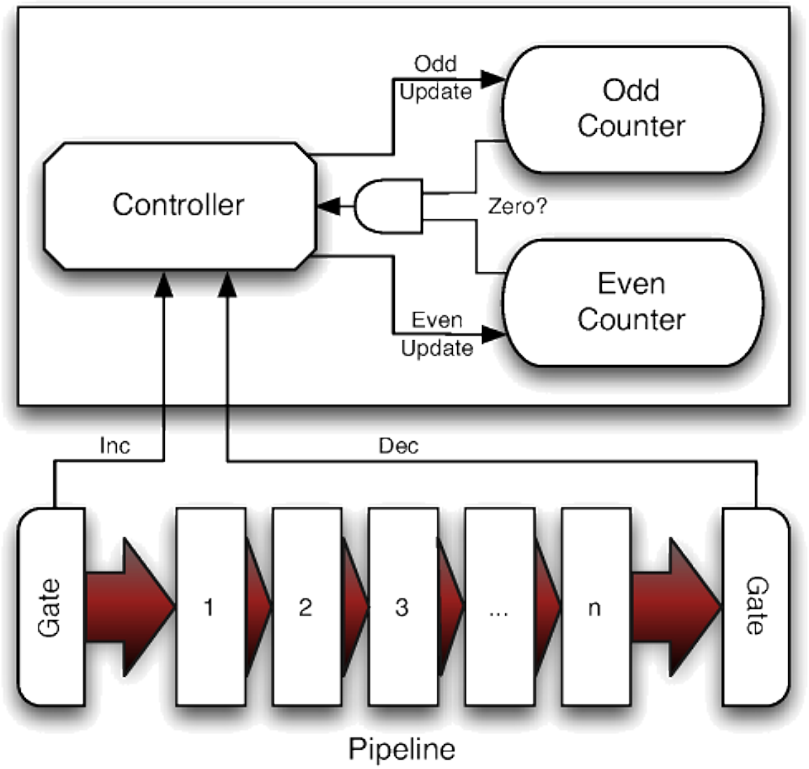
Empty Pipeline Detection

- Detect when it is safe to power gate
- Constant reponse time counter



Empty Pipeline Detection

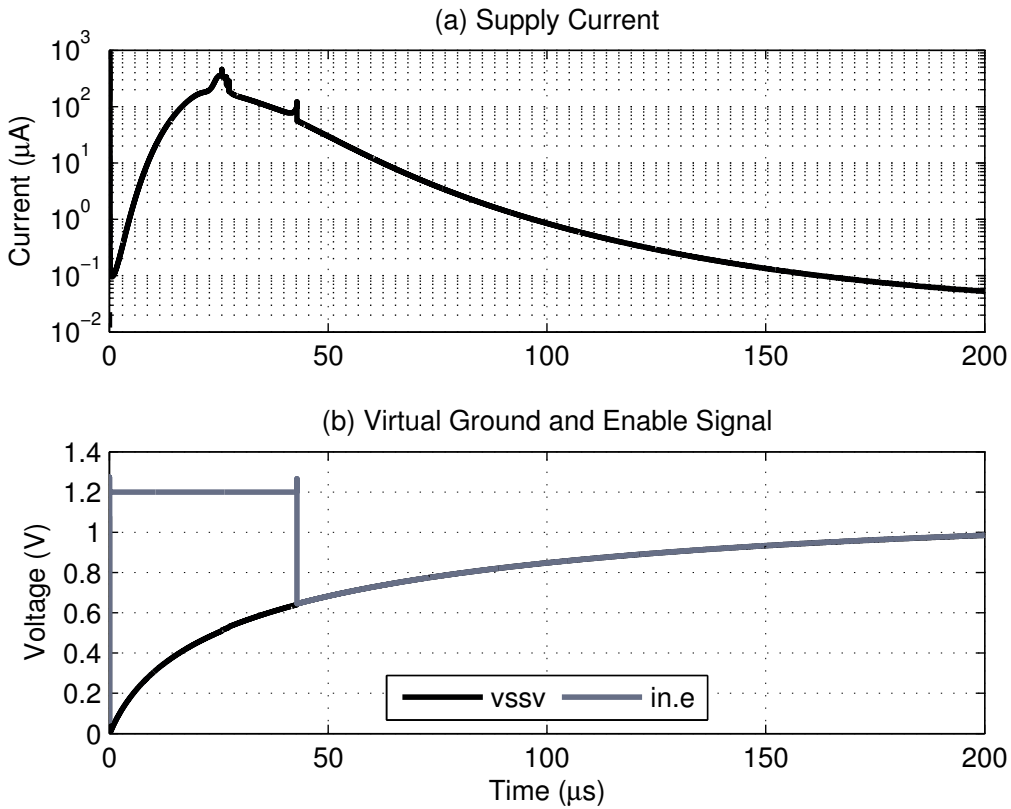
- Interleaved counter allows full throughput operation
- Minimum overhead



Cut-off techniques

- Cut-off. Both logic and sleep use RVT devices
- MTCMOS: Logic implemented using low/regular VT while sleep transistors -> high-VT devices
- BGCMOS: Boosted Gate: high-Vt thick oxide sleep transistors -> hurt performance -> overdrive Vdd during active mode
- Super Cut-Off: Gate of sleep transistors driven past supply voltages when idle
- Problems: Foundry support and biasing

Cut-off transient behaviour



Self-throttling Asynchronous Circuits

- Stability: G->t is stable when G can change from true to false only in states where R(t) holds.
  - State wont change unless it is acknowledged
- Problems: Noise margins are reduced.
  - You should wait to assert "safe" until enough noise margin exist
- Assumption: A monotonic change on input will create a monotonic change on output
  - Noise margins again. Charge Sharing and Capacitive Coupling can break this assumption

Problems and Future Research

- Stability: G->t is stable when G can change from true to false only in states where R(t) holds.
  - State wont change unless it is acknowledged
- Problems: Noise margins get diminished
- If there is not enough "sleep" time, then you can end-up using more power
  - We should compute break-even point
- We need some tools to power gate in a systematic and provably correct way

Gate Leakage

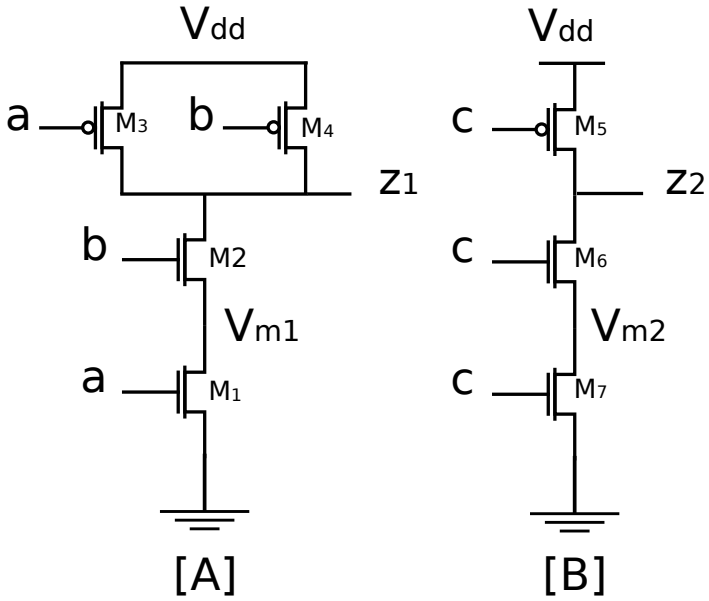
- Current that dribbles through the gate of the transistors
  - Direct Tunneling
  - Hot carrier injection
- Increasing component of the absolute leakage currents

Device Level

- Device Engineering. Transistors and devices
- L_{eff}, T_{ox}, Substrate depth, nominal values of V_{dd} and V_{th}
- Choice of materials (semiconductor, metal, dielectric)
- Doping profile and doping halo
- Requires expertise on device physics
- Choice of fabrication process

Circuit Level

- Natural transistor stacking
- Forced stacking



System Level

- Design specific
  - SRAM topology
  - Amount of dark silicon
- General techniques
  - Power gating