

---

# Dual-K Versus Dual-T Technique for Gate Leakage Reduction : A Comparative Perspective

S. P. Mohanty, R. Velagapudi and E. Kougianos

Dept of Computer Science and Engineering

University of North Texas

email-ID: [smohanty@cs.unt.edu](mailto:smohanty@cs.unt.edu)

Homepage: <http://www.cs.unt.edu/~smohanty/>

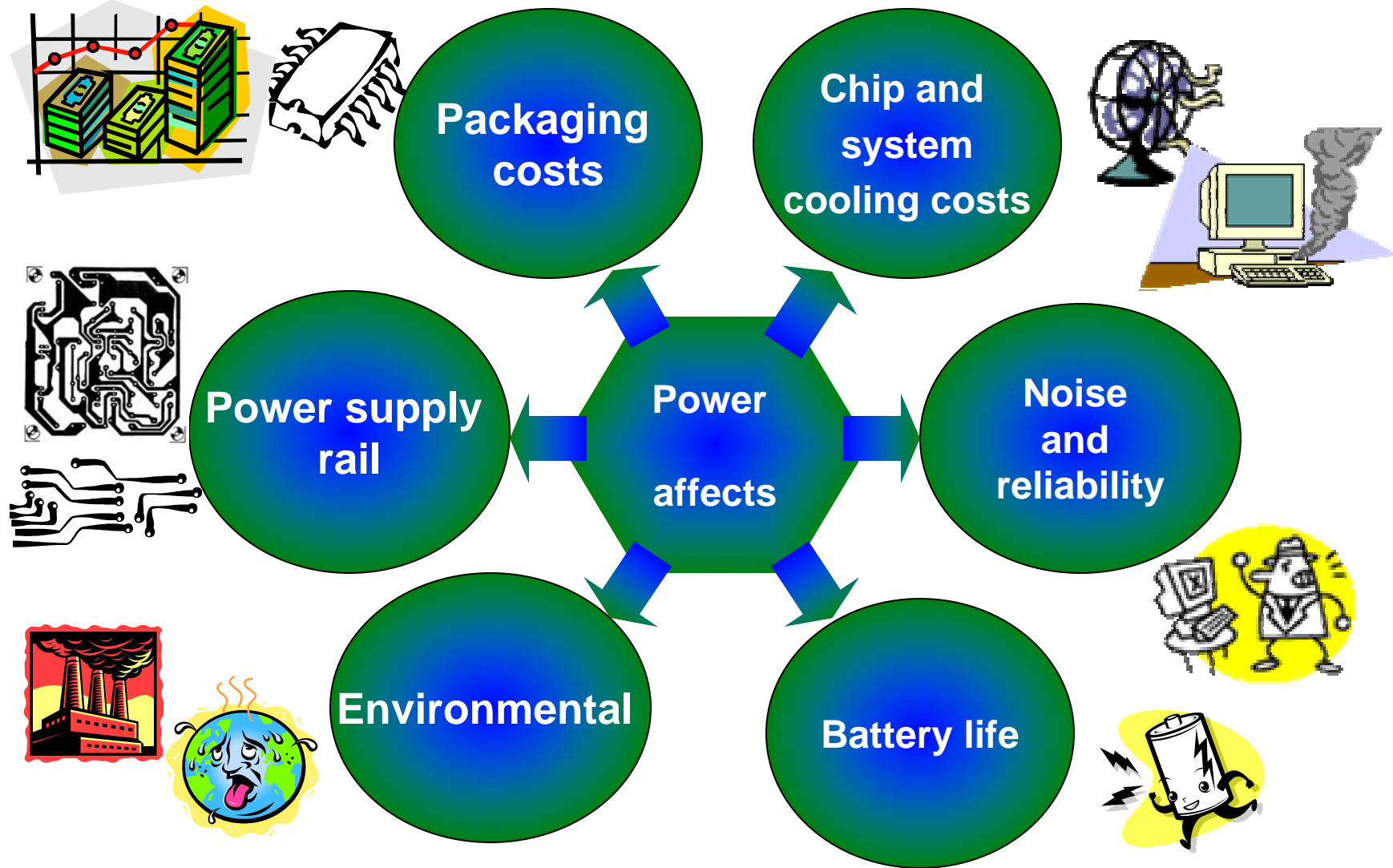
---

---

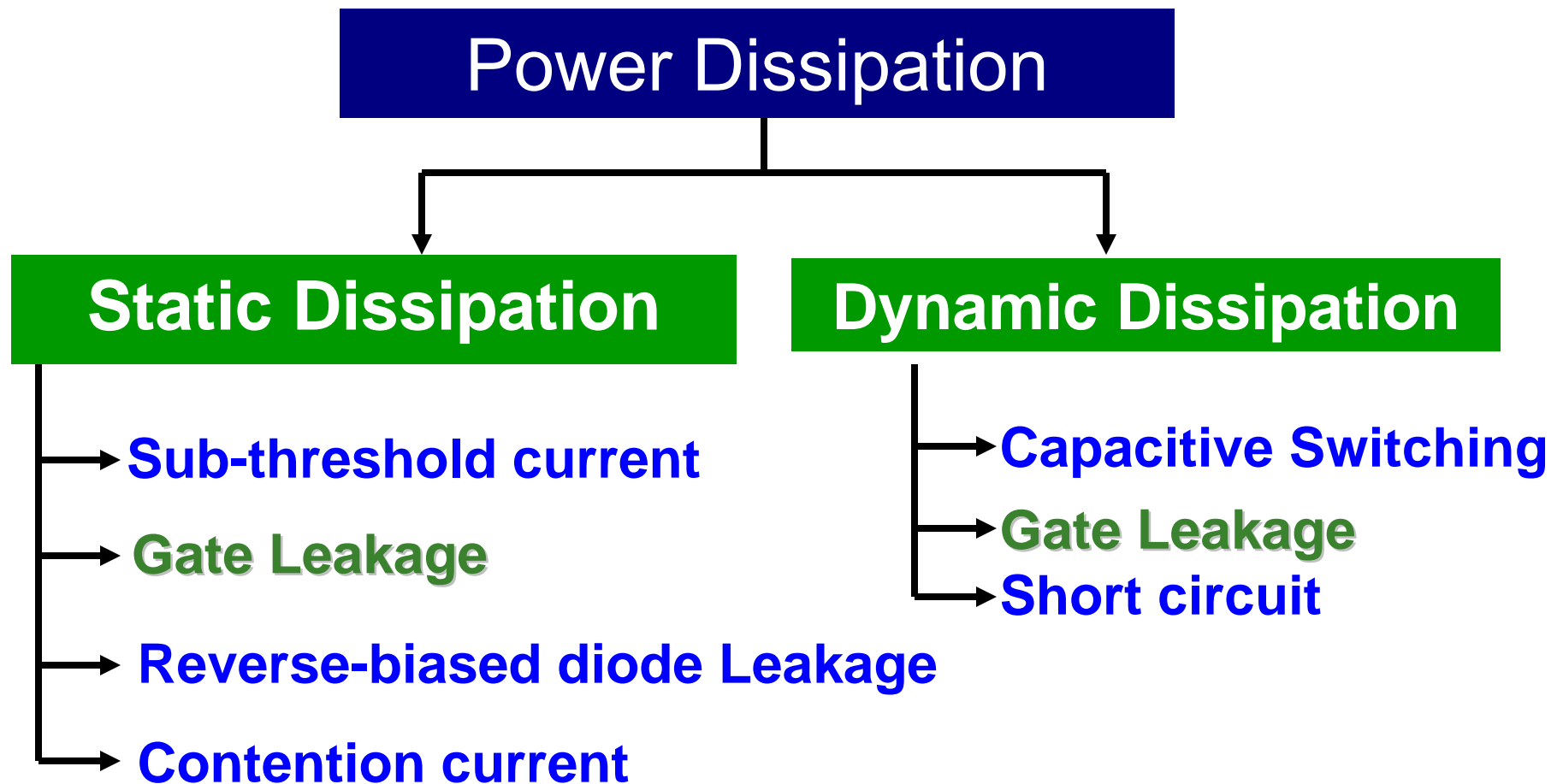
# Outline of the Talk

- Introduction
- Related Work
- Methods for Gate Leakage Reduction
- Datapath Component Library
- Gate Leakage Optimization
- Experimental Results
- Conclusions

# Why Low Power?



# Power Dissipation in CMOS



Source: Weste and Harris 2005

# Leakages in CMOS

$I_1$  : reverse bias pn junction (both ON & OFF)

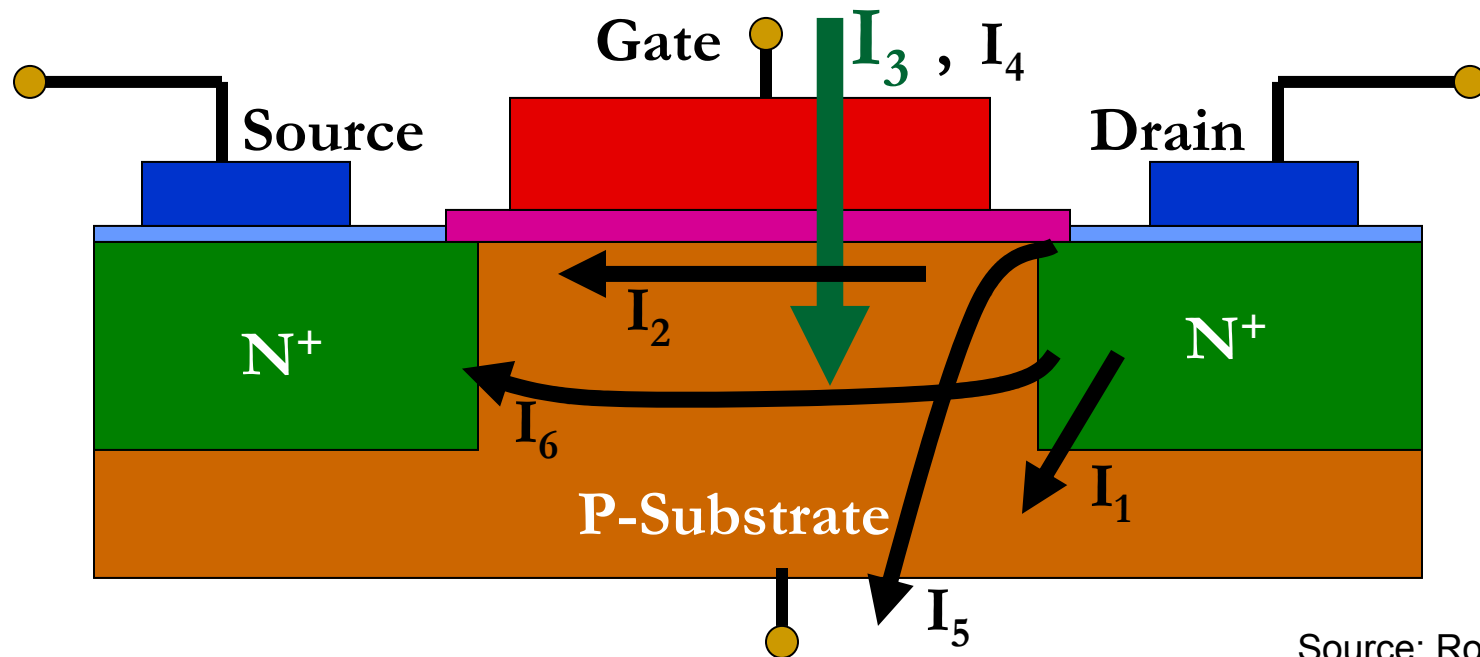
$I_2$  : subthreshold leakage (OFF )

$I_3$  : **Gate Leakage current (both ON & OFF)**

$I_4$  : gate current due to hot carrier injection (both ON & OFF)

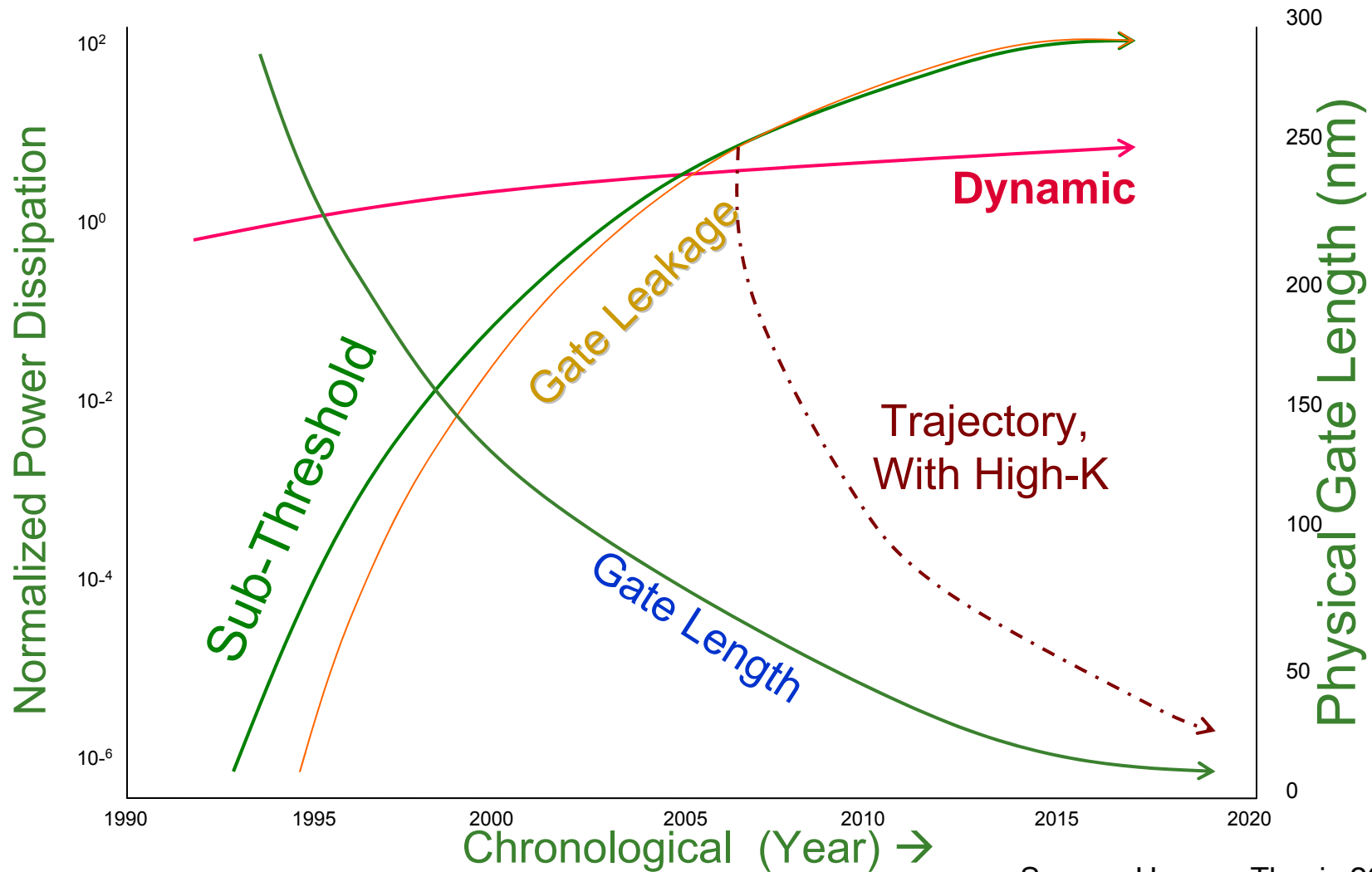
$I_5$  : gate induced drain leakage (OFF)

$I_6$  : channel punch through current (OFF)



Source: Roy 2003

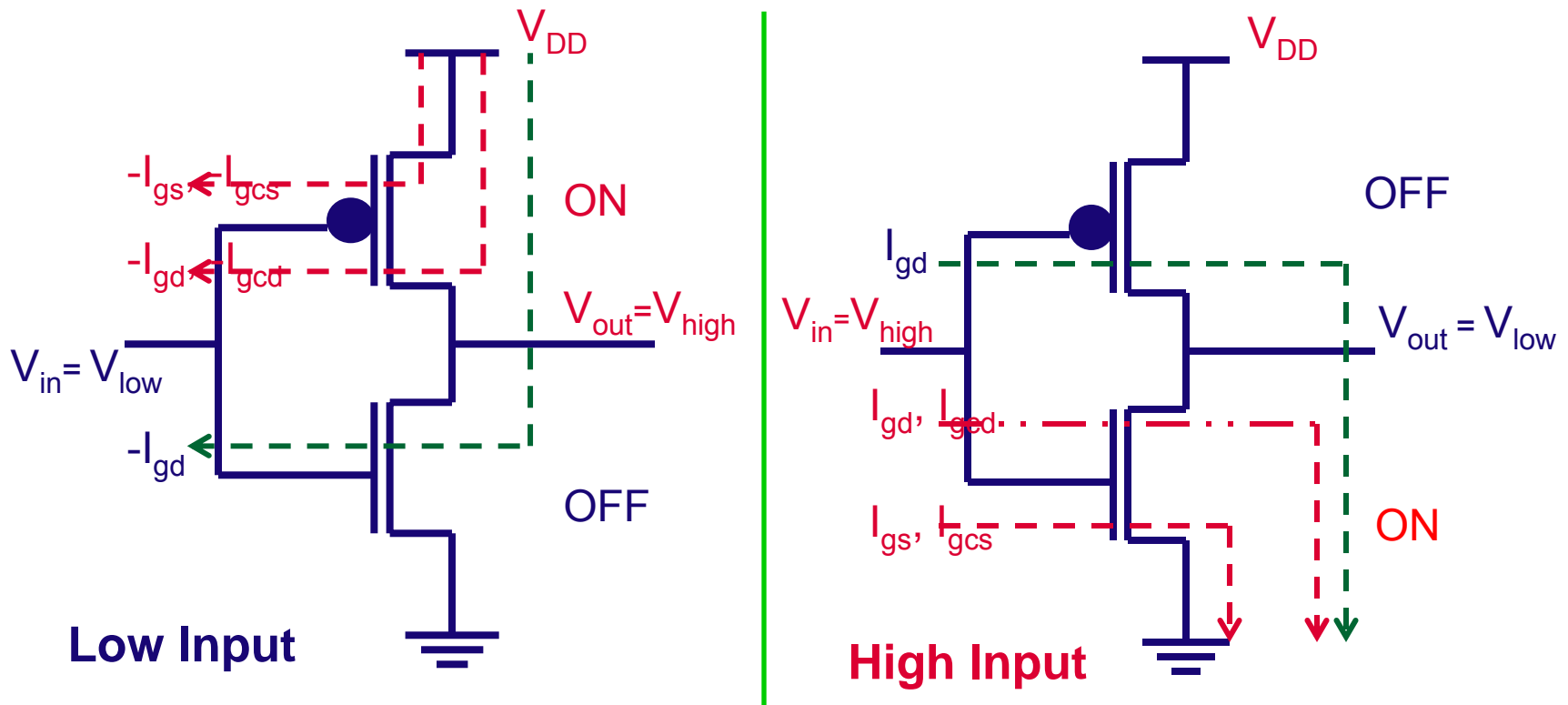
# Power Dissipation Redistribution



Source: Hansen Thesis 2004

# Gate Leakage Paths in an Inverter

- **Low Input:** Input supply feeds tunneling current.
- **High Input:** Gate supply feeds tunneling current.



NOTE: Gate to body component found to be negligible.

# Gate Leakage Reduction Techniques

Research in Gate Leakage reduction is in full swing, but is not as mature as that of dynamic power or subthreshold leakage.

Few methods:

- Dual  $T_{ox}$  (Sultania – TVLSI Dec 2005, Sultania - DAC 2004, Sirisantana - IEEE DTC Jan-Feb 2004, Mohanty – VLSI Design 2006)
- Dual K (Mukherjee - ICCD 2005)
- Pin and Transistor Reordering (Sultania - ICCD 2004, Lee - DAC 2003)



---

# Related Works : Behavioral Level

## Subthreshold Leakage:

- **Khouri - TVLSI 2002** : Algorithms for subthreshold leakage power analysis and reduction using dual- $V_{Th}$  approach.
- **Gopalakrishnan - ICCD2003**: Dual- $V_{Th}$  approach for reduction of subthreshold current through binding.

## Gate Leakage:

- **Mohanty - VLSI Design 2006**: Dual- $T_{ox}$  approach for reduction of gate leakage current.

# Related Works : Logic / Transistor Level Gate Leakage Reduction

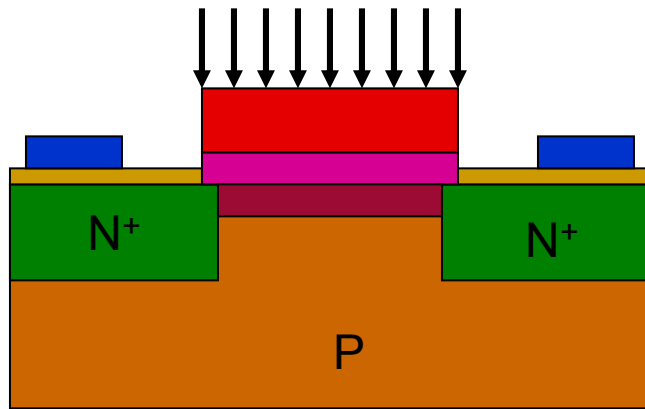
- ❑ Lee - TVLSI2004 : Pin reordering to minimize gate leakage during standby positions of logic gates.
- ❑ Sultania – TVLSI Dec 2005 and Sultania - DAC2004 : Heuristic for dual- $T_{ox}$  assignment for gate leakage and delay tradeoff.
- ❑ Sirisantana - IEEE DTC Jan-Feb 2004: Use multiple channel lengths and multiple gate oxide thickness for reduction of leakage.
- ❑ Mukherjee - ICCD 2005: Introduced dual-K approach for reduction of gate leakage.

---

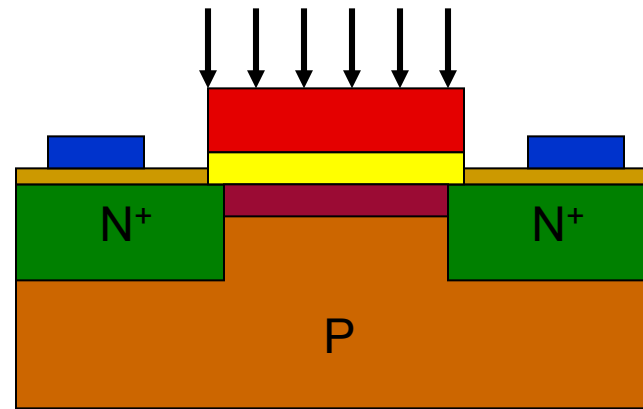
# Key Contributions of this Paper

- Introduces dual dielectric assignment approach for architectural level gate leakage reduction.
- Presents a Simulated Annealing based optimization for gate leakage current reduction during behavioral synthesis.
- Compares the two approaches (Dual-Dielectric Vs Dual-Thickness i. e. Dual-K Vs Dual-T).

# Dual-K : Low $K_{gate}$ and High $K_{gate}$



Low  $K_{gate}$  → Larger  $I_{gate}$ ,  
Smaller delay

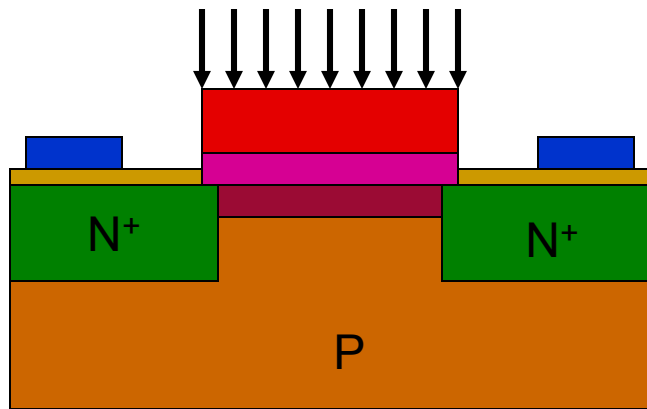


High  $K_{gate}$  → Smaller  $I_{gate}$ ,  
Larger delay

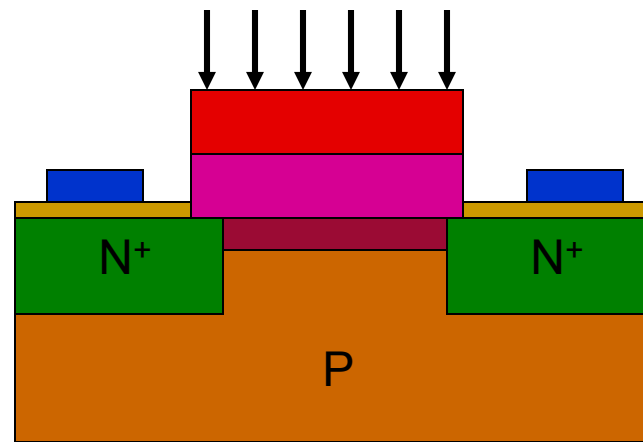
# Dielectrics for Replacement of $\text{SiO}_2$

- Silicon Oxynitride ( $\text{SiO}_x\text{N}_y$ ) ( $K=5.7$  for SiON)
- Silicon Nitride ( $\text{Si}_3\text{N}_4$ ) ( $K=7$ )
- Oxides of :
  - Aluminum (Al), Titanium (Ti), Zirconium (Zr), Hafnium (Hf), Lanthanum (La), Yttrium (Y), Praseodymium (Pr),
  - their mixed oxides with  $\text{SiO}_2$  and  $\text{Al}_2\text{O}_3$

# Dual-T : Low $T_{\text{gate}}$ and High $T_{\text{gate}}$



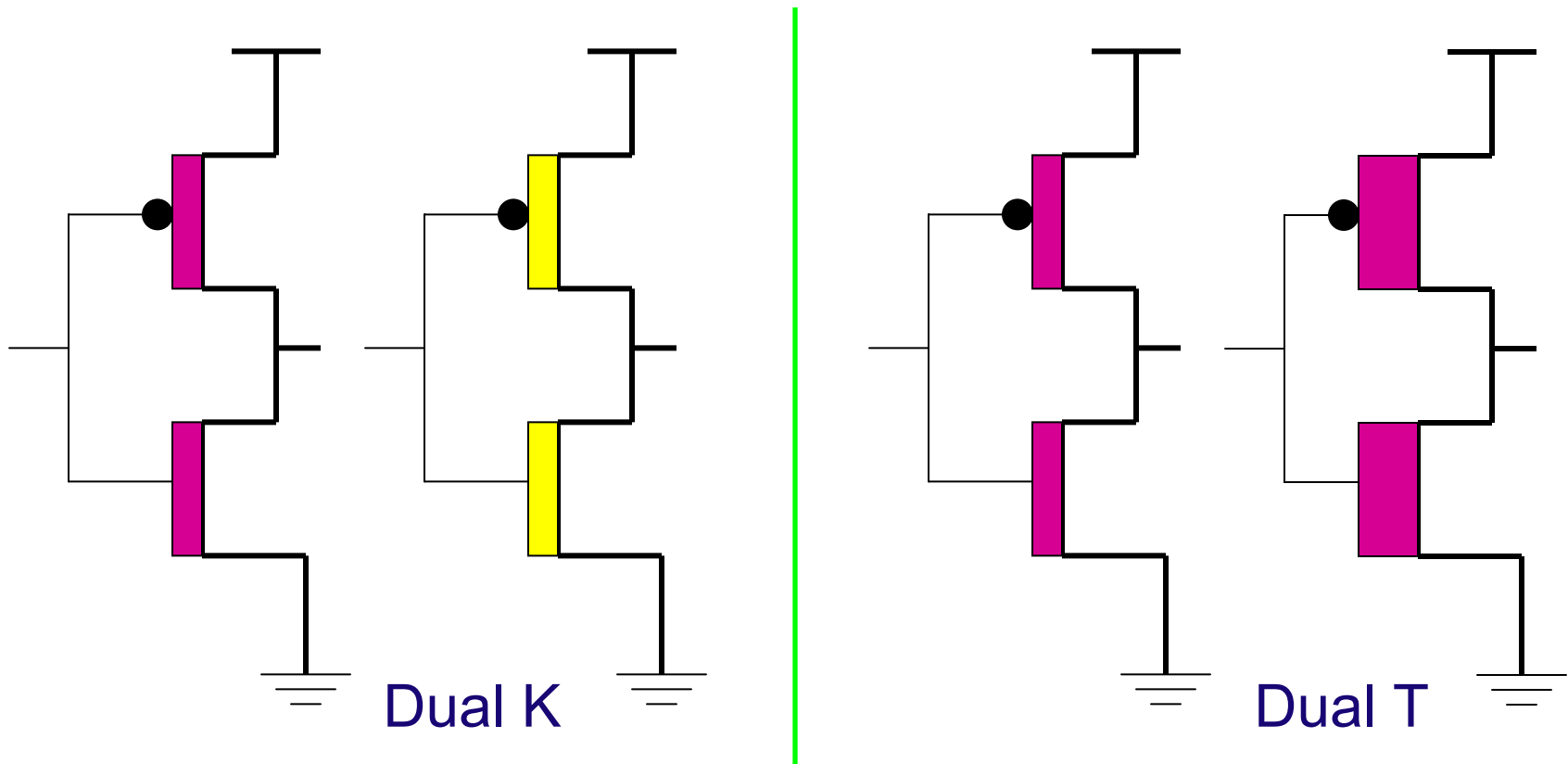
Low  $T_{\text{gate}}$  → Larger  $I_{\text{gate}}$ ,  
Smaller delay



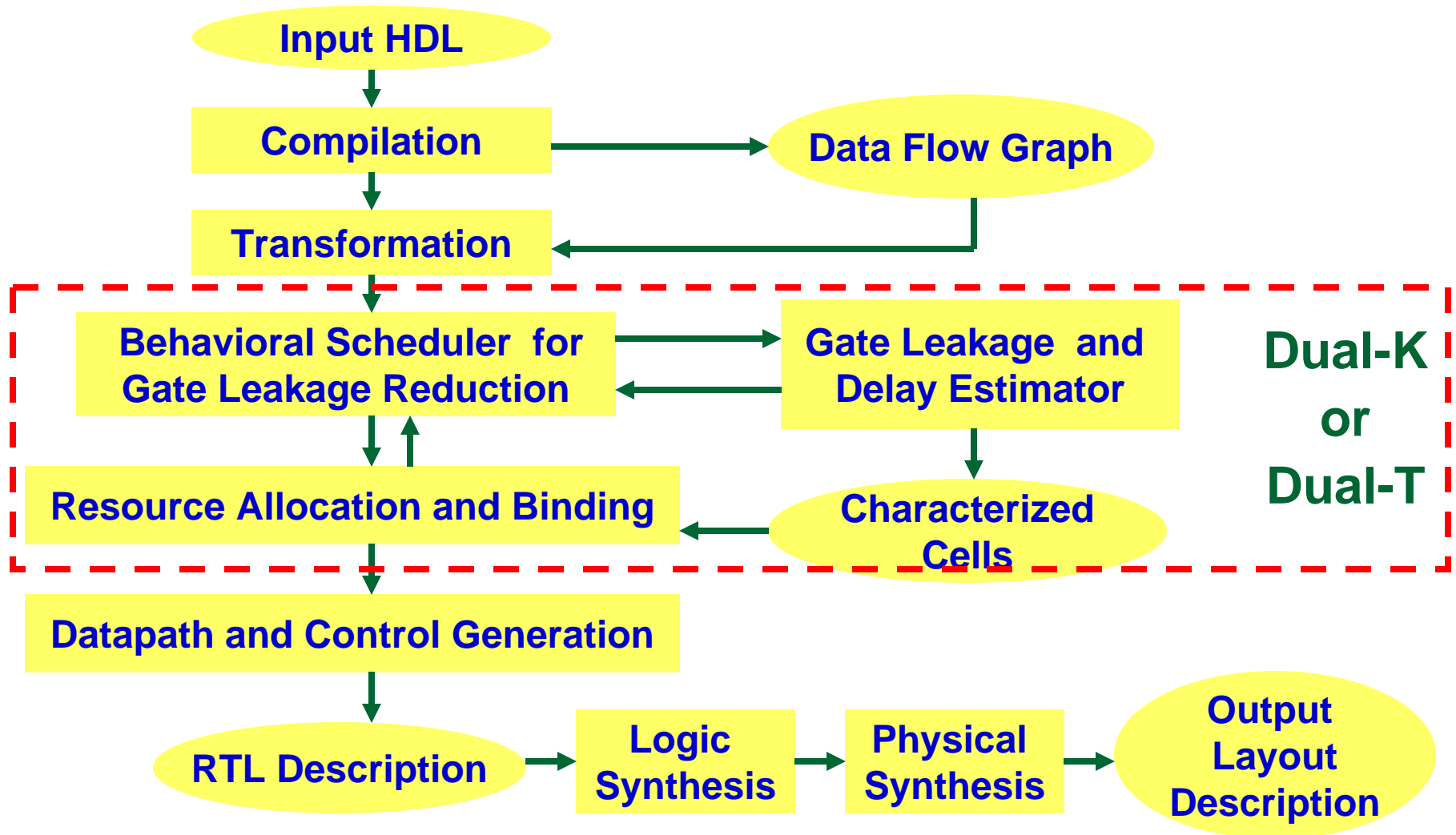
High  $T_{\text{gate}}$  → Smaller  $I_{\text{gate}}$ ,  
Larger delay

# Dual-K Vs Dual-T Approach

**Assumption:** All functional units have transistors of same  $K_{gate}$  or  $T_{gate}$ .

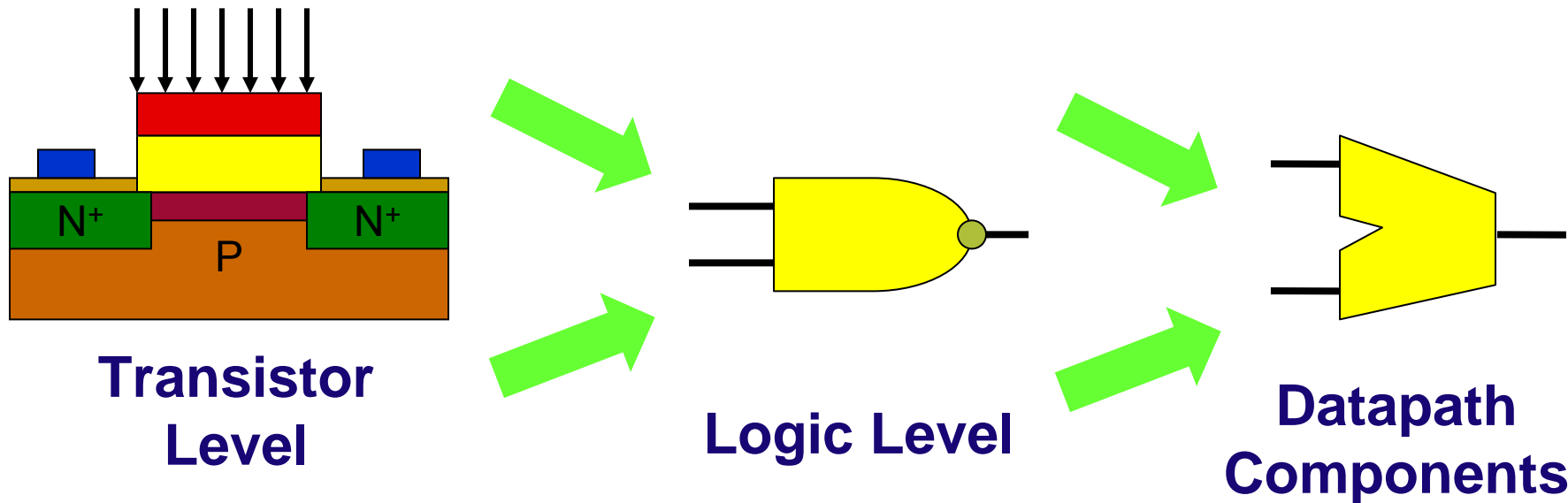


# Synthesis for Low Gate Leakage





# Datapath Component Library : 3 Level Bottom-up Hierarchical Approach



We observed that a NAND gate has least gate leakage compared to all other basic logic gates. Therefore we constructed datapath components using NAND gates.

# Datapath Component Library

- First we characterize the NAND gate using analog simulations and then characterize functional units.
- We assume that there are total  $n_{total}$  NAND gates in the network of NAND gates constituting an  $n$ -bit functional unit out of which  $n_{cp}$  are in the critical path.
- We do not consider the effect of interconnect wires and focus on the gate leakage current dissipation and propagation delay of the active units only.

# Datapath Component Library: Logic

- BSIM4 model based simulations used to calculate gate leakage  $I_{ox}$  and  $T_{pd}$ .
- Due to the unavailability of silicon data we used an analytical estimate for area calculations.

$$A_{NAND} = K_{inv} \left( 1 + 4(n_{in} - 1) \sqrt{\frac{AR_{NAND}}{K_{inv}}} \right) * \left( 1 + \frac{\left( \frac{W_{NMOS}}{f} - 1 \right) (1 + \beta_{NAND})}{\sqrt{K_{inv} AR_{NAND}}} \right)$$

$W_{NMOS}$  = NMOS width

$f$  = Minimum feature size for a technology

$k_{inv}$  = Area of minimum size inverter using  $f$

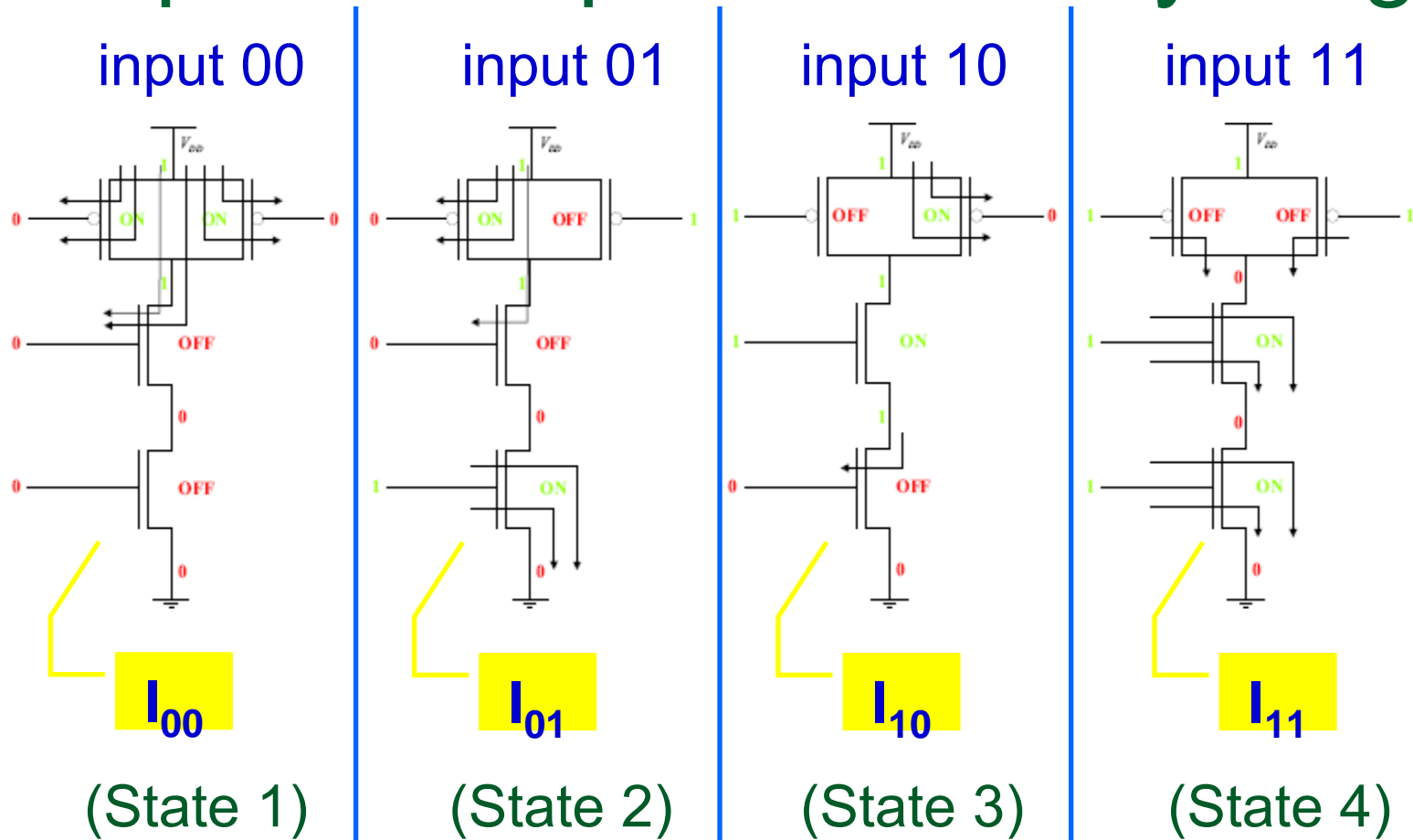
$AR_{NAND}$  = aspect ratio of NAND gate

$n_{in}$  = number of inputs

$\beta_{NAND}$  = ratio of PMOS width to NMOS width

Source: Bowman TED 2001 Aug

# Datapath Component Library: Logic



$$I_{gateNAND} = \left( \frac{I_{00} + I_{01} + I_{10} + I_{11}}{4} \right) \text{ (Assuming all states to be equiprobable.)}$$

# Datapath Component Library

- Gate leakage current ( $I_{gateFU}$ ) of an  $n$ -bit functional unit is:

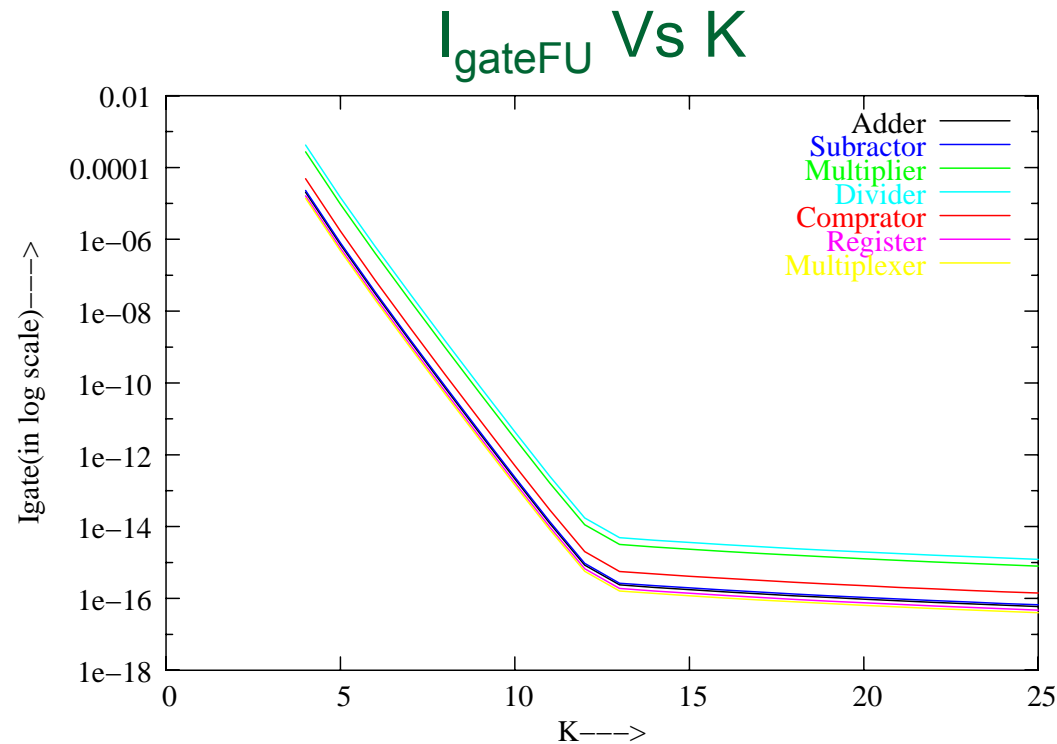
$$I_{gateFU} = \sum_{i=1}^{n_{total}} I_{gateNANDi}$$

where  $I_{gateNANDi}$  is the average gate leakage current dissipation of the  $i^{th}$  2-input NAND gate in the functional unit, assuming all states to be equiprobable.

- Similarly, the propagation delay and silicon area of an  $n$ -bit functional unit are:

$$T_{pdFU} = \sum_{i=1}^{n_{cp}} T_{pdNANDi} \quad A_{FU} = \sum_{i=1}^{n_{total}} A_{NANDi}$$

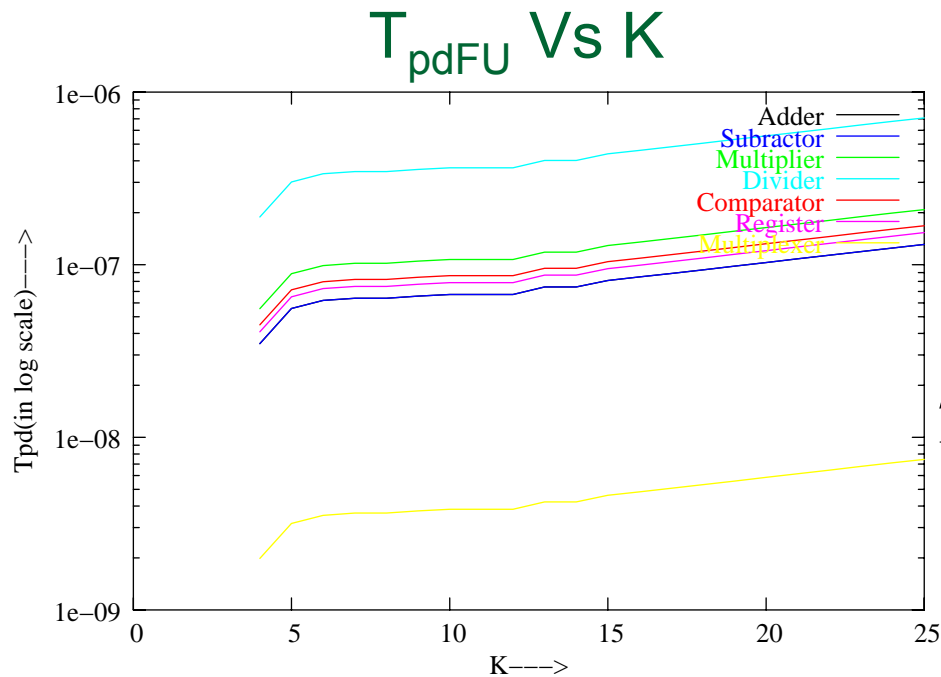
# Gate Leakage Vs Permittivity



$$I_{gateFU} (\mu A) = A \times e^{\left(\frac{-K}{\alpha}\right)} + I_{gateFU 0}$$

As the gate dielectric constant increases the gate leakage current decreases.

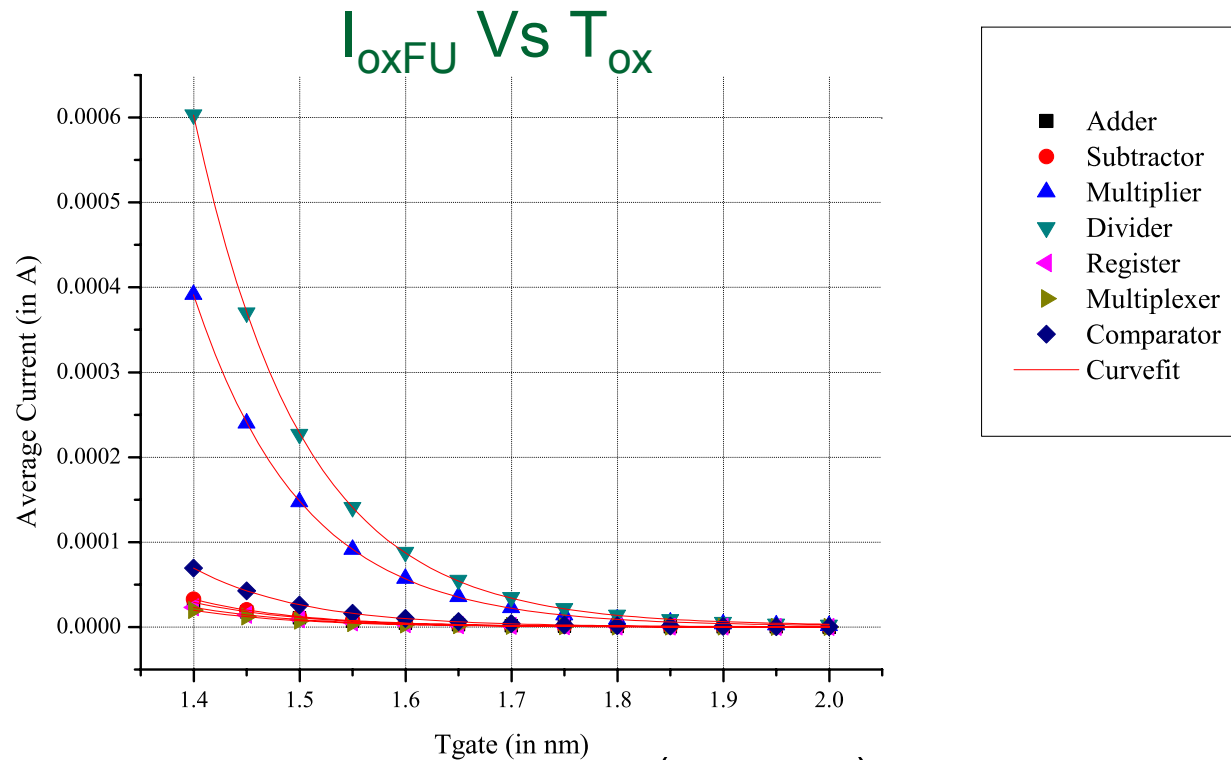
# Propagation Delay Vs Permittivity



$$T_{pdFU}(ns) = \begin{cases} A_2 + \frac{A_1 - A_2}{1 + e^{\left(\frac{K - K_0}{\beta * K}\right)}}, & 2.5 \leq K \leq 6 \\ C \times e^{\left(\frac{-K}{\alpha}\right)} + T_{pdFU0}, & 6 \leq K \leq 30 \end{cases}$$

As the gate dielectric constant increases the propagation delay increases.

# Gate Leakage Vs SiO<sub>2</sub> Thickness

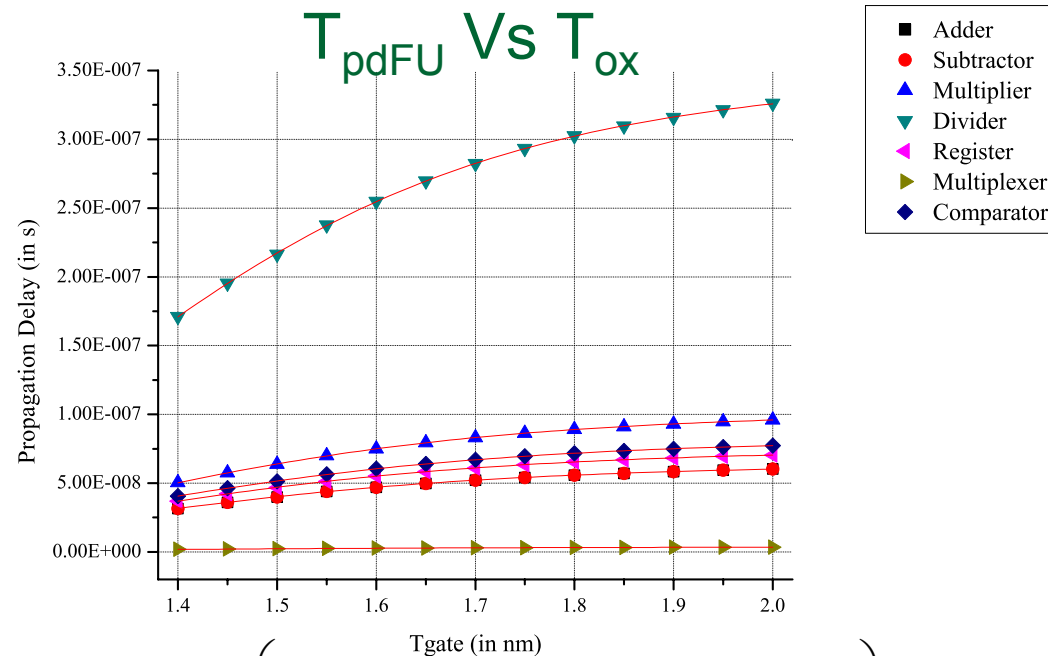


$$I_{gateFU} (\mu A) = A \exp\left(-T_{ox}/\alpha\right) + \beta$$

As the gate oxide thickness increases the gate leakage current decreases.



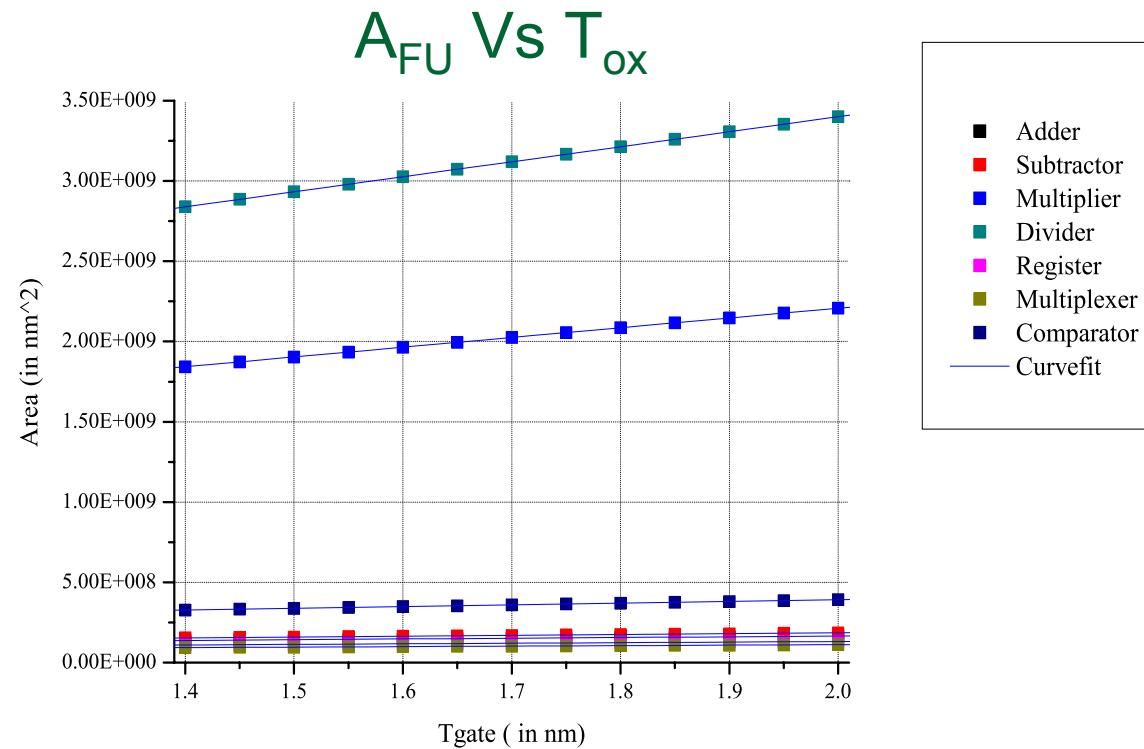
# Propagation Delay Vs SiO<sub>2</sub> Thickness



$$T_{pdFU} (ns) = \left( \frac{(A_1 - A_2)}{\left( 1 + \left( \frac{T_{ox}}{\beta} \right)^v \right)} \right) + A_2$$

As the gate oxide thickness increases the propagation delay increases.

# Silicon Area Vs SiO<sub>2</sub> Thickness



$$A_{FU} (nm^2) = \alpha T_{ox} + \beta$$

As the gate oxide thickness increases the area increases.

---

# Simulated Annealing for Optimization

- Analogous to the annealing process, the mobility of nodes in a DFG is dependent on the total available resources.
- Nodes of a DFG are analogous to the atoms and temperature is analogous to the total number of available resources.
- To maximize the leakage reduction we need to ensure that a node can be scheduled in such a way that a higher thickness (or dielectric) resource can be assigned.
- The chance of assigning a higher thickness (or dielectric) resource is higher if the total number of available higher thickness resources is higher.

# Optimization Algorithm

Simulated Annealing Algorithm (UDFG, DTF, LRM)

- (01) Available Resources  $\leftarrow \infty$
- (02) While there exists a schedule with available resources.
- (03)      $i$  = Number of iterations
- (04)     Perform resource constrained ASAP and ALAP
- (05)     Initial Solution  $\leftarrow$  ASAP Schedule
- (06)      $S \leftarrow$  Allocate Bind()
- (07)     Initial gate leakage  $\leftarrow$  gate leakage( $S$ )
- (08)     While ( $i > 0$ )
- (09)         Generate a random  $T_{ox}$  in range ( $T_{ox} - \Delta T_{ox}$ ,  $T_{ox} + \Delta T_{ox}$ )
- (10)         Generate random transition from  $S$  to  $S^*$
- (11)          $\Delta I \leftarrow$  gate leakage( $S$ ) – gate leakage( $S^*$ )
- (12)         if(  $\Delta I > 0$  ) then  $S \leftarrow S^*$
- (13)          $i \leftarrow i - 1$
- (14)     end While
- (15)     Decrement available resources
- (16) end While
- (17) return  $S$

---

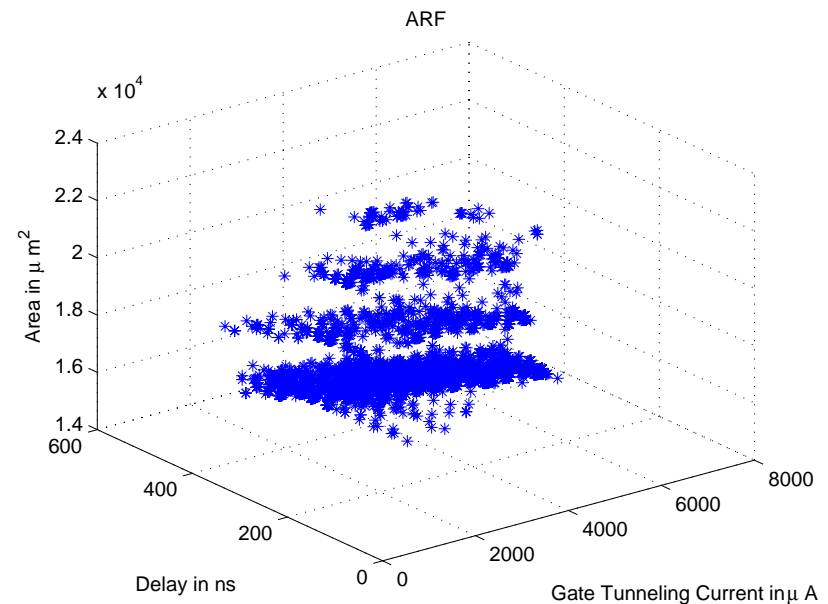
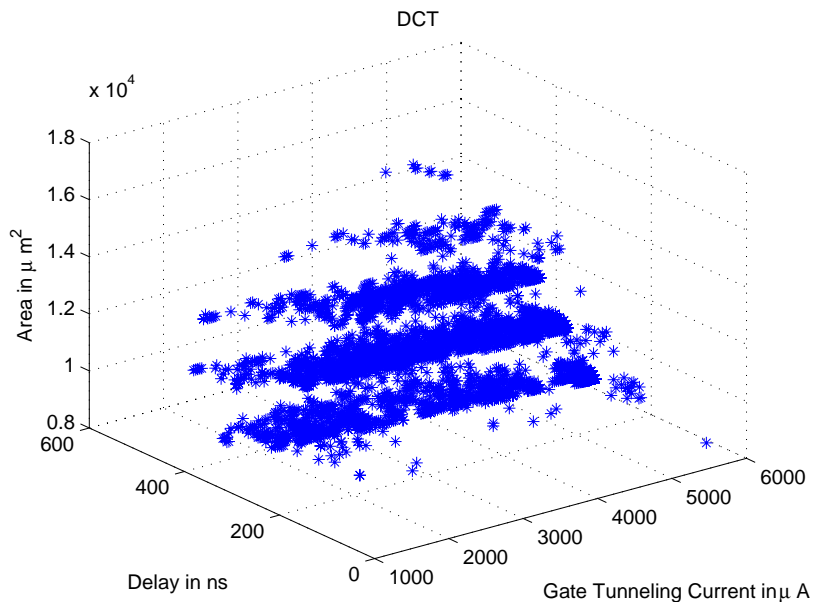
# Experimental Results

- Critical path delay of the circuit is the sum of the delays of the vertices in the longest path of the DFG for single cycle case and number of control steps times slowest delay resource for multicycling or chaining case.
- The delay trade-off factor (DTF) is used to provide various time constraints for our experiments.

# Experimental Results

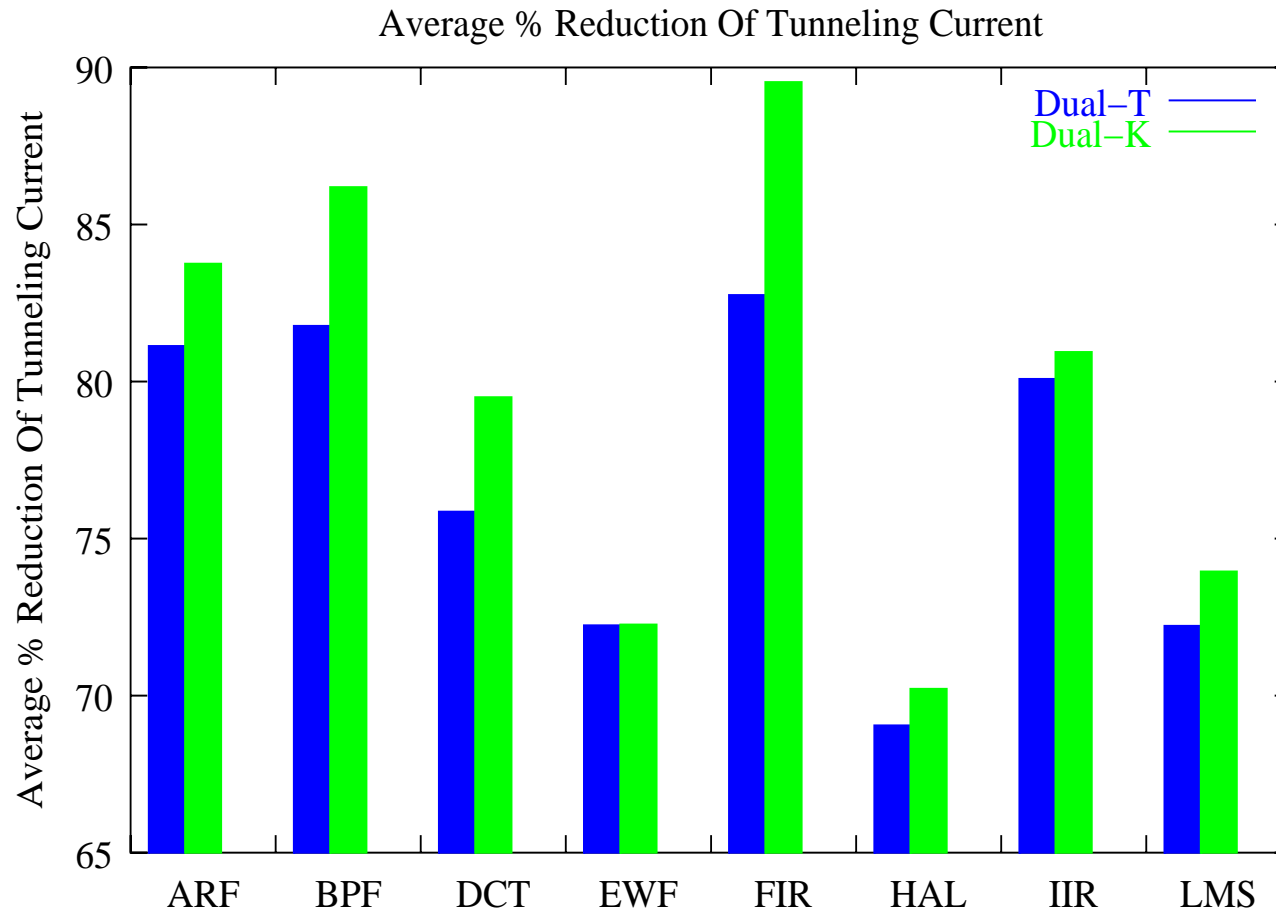
- While calculating the gate leakage current for single thickness, we used a nominal  $1.4nm$  thickness and  $SiO_2$  ( $K=3.9$ ) is used as a nominal dielectric value from BSIM4.4.0 model.
- For dual thickness approach the following pair is considered:  $1.4nm - 1.7nm$ .
- For dual dielectric approach the following pair is considered:  $SiO_2$  ( $K=3.9$ ) –  $Si_3N_4$  ( $K=7$ ).
- The results take into account the gate leakage current, area and propagation delay of functional units, interconnect units, and storage units present in the datapath circuit.

# Experimental Results



Each layer corresponds to a different resource constraint, each time the number of  $T_{oxH}$  multipliers are decreased a new layer is formed. We observed that the number of design corners reduces when we use more multipliers of  $T_{oxH}$  thickness, since delay increases and mobility of the nodes is restricted in order to satisfy the time constraint.

# Experimental Results





---

# Conclusions and Future Works

- A comparison of dual thickness and dual dielectric approaches for reduction of gate leakage during behavioral synthesis is presented.
- A simulated annealing based algorithm for simultaneous scheduling and binding of functional units is introduced.
- Tradeoff between gate leakage, area and performance is explored.
- Both approaches for gate leakage reduction account for the ON as well as OFF state.

---

# Conclusions and Future Works

- Experiments prove significant reductions in gate leakage current without performance penalty.
- The method of using dual dielectric is proven to be more productive than the dual thickness approach.
- This work on gate leakage will be extended to provide a broader solution to the problem of power dissipation in all its forms at the behavioral level.
- Dual-K or Dual-T based design may need more masks for the lithographic process during fabrication compared to single-K or single-T.