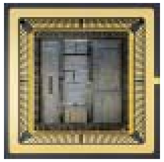


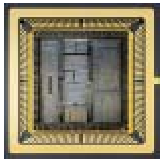
A Dual Dielectric Approach for Performance Aware Gate Tunneling Reduction in Combinational Circuits

V. Mukherjee, S. P. Mohanty, E. Kougianos

University of North Texas, Denton, TX 76203.

Email: smohanty@cs.unt.edu

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



Outline of the Talk

- Introduction
- Why Dual Dielectric
- Related Works
- DKDT Assignment Algorithm
- Cell Characterization for High-K
- Conclusions

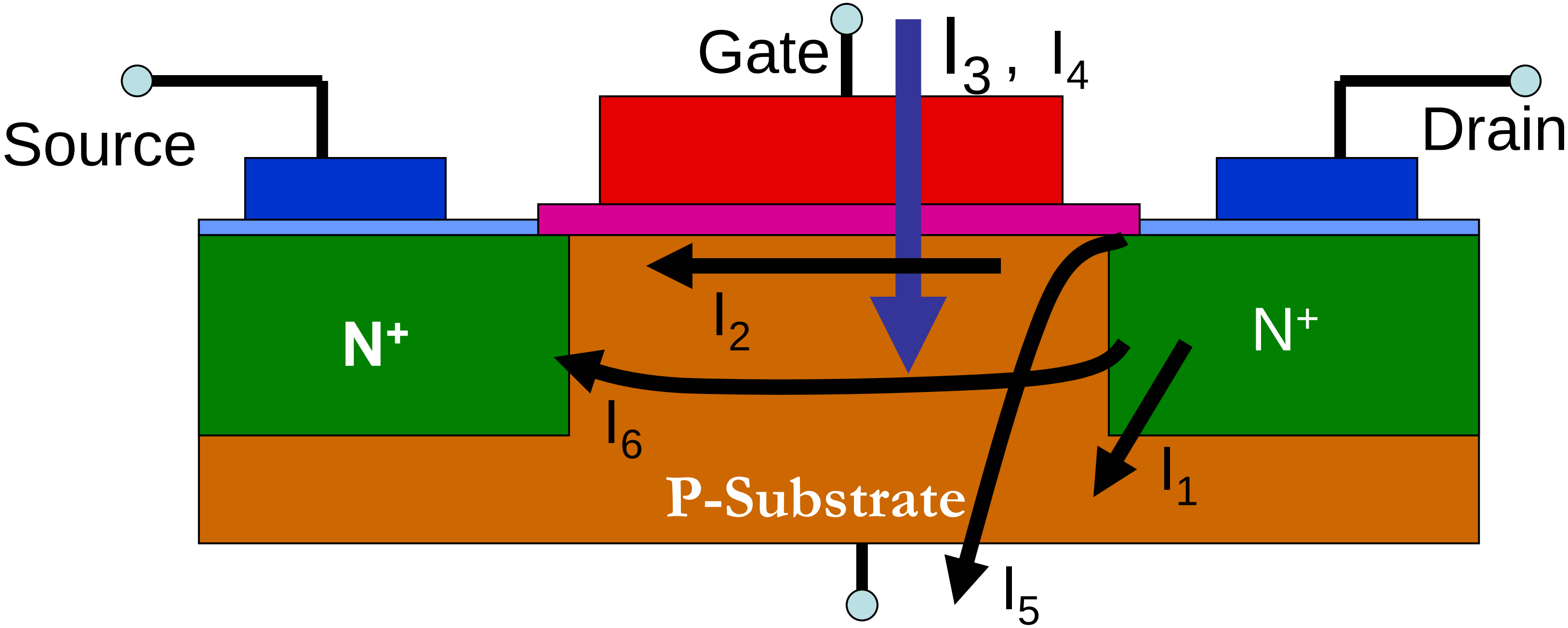
Why Low-Power ?

Major Motivation: Extending battery life

Source: Power Integrations Inc

Leakages in Nanometer CMOS

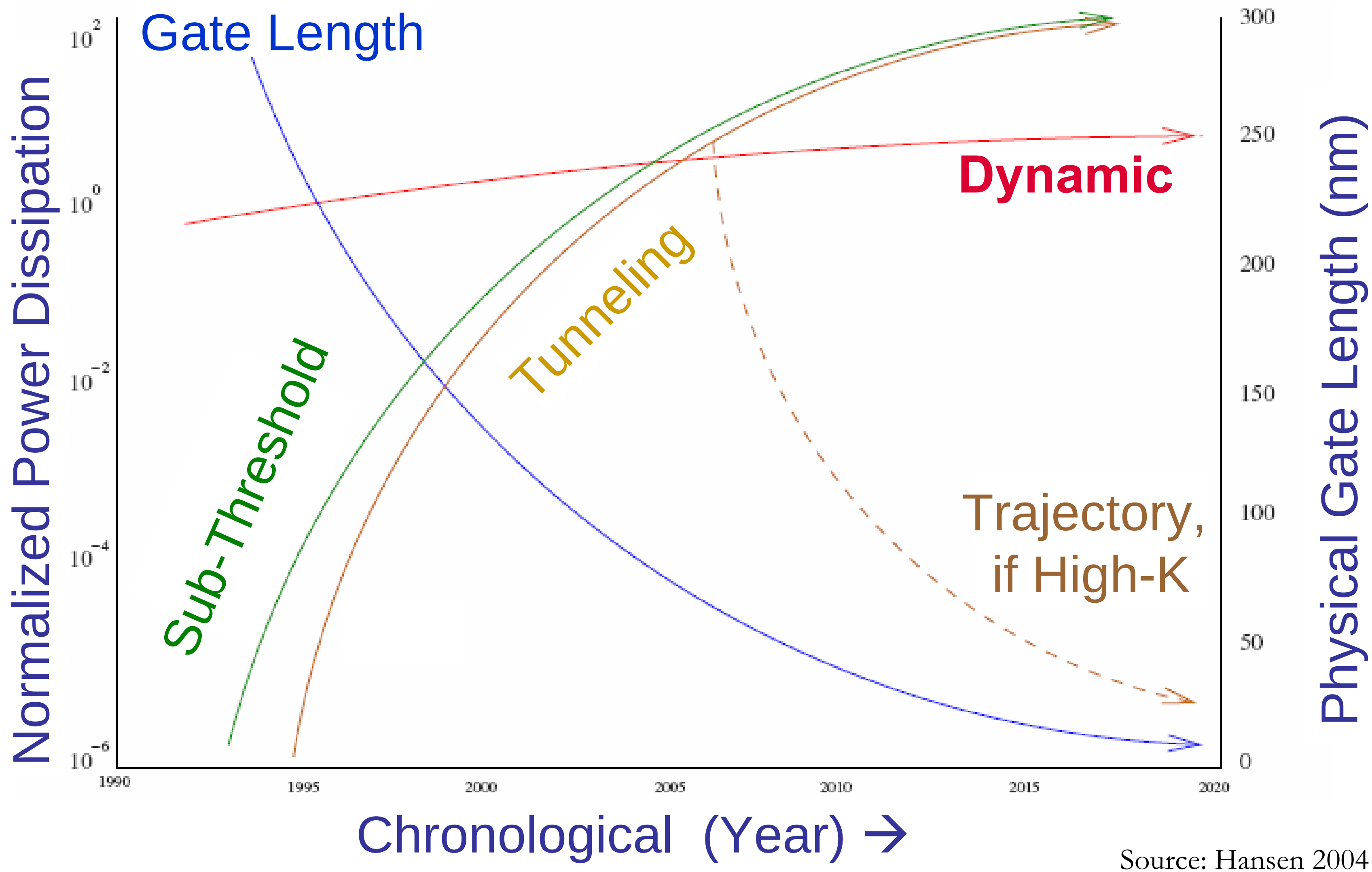
I_1 : reverse bias pn junction (both ON & OFF)
I_2 : subthreshold leakage (OFF)
I_3 : oxide tunneling current (both ON & OFF)
I_4 : gate current due to hot carrier injection (both ON & OFF)
I_5 : gate induced drain leakage (OFF)
I_6 : channel punch through current (OFF)



UNIVERSITY of NORTH TEXAS CSE

Source: Roy 2003

Power Dissipation Trend



Source: Hansen 2004

Why Dual-K and Dual-T ?

- Gate oxide tunneling current I_{gate} [Kim2003, Chandrakasan2001] (k is a experimentally derived factors):

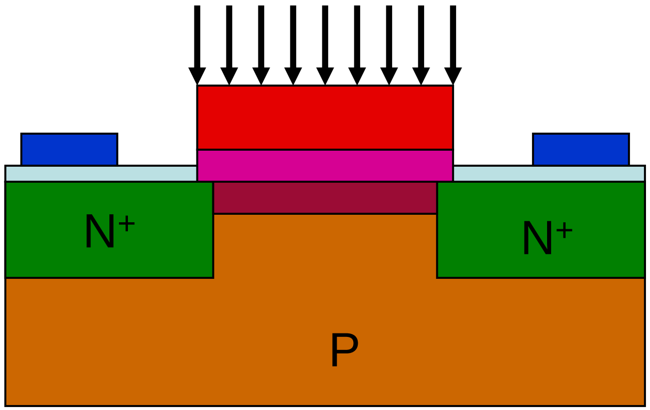
$$\mathbf{I_{gate}\ \alpha\ (V_{dd}/T_{gate})^2 \exp(-k\ T_{gate}/V_{dd})}$$

- Options for reduction of tunneling current:
  - Decreasing of supply voltage V_{dd} (*will play its role*)
  - Increasing gate SiO_2 thickness T_{gate} (*opposed to the technology trend !!*)

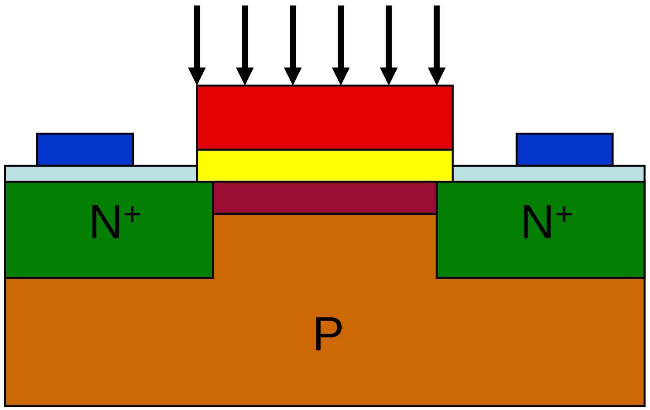
Why Dual-K and Dual-T ?

Use of multiple dielectrics (denoted as K_{gate}) of multiple thickness (denoted as T_{gate}) will reduce the gate tunneling current significantly while maintaining the performance.

Why Dual-K and Dual-T ? (Low K_{gate} Vs High K_{gate})

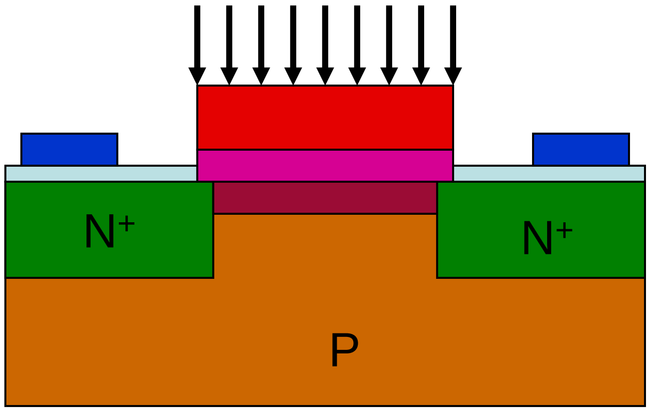


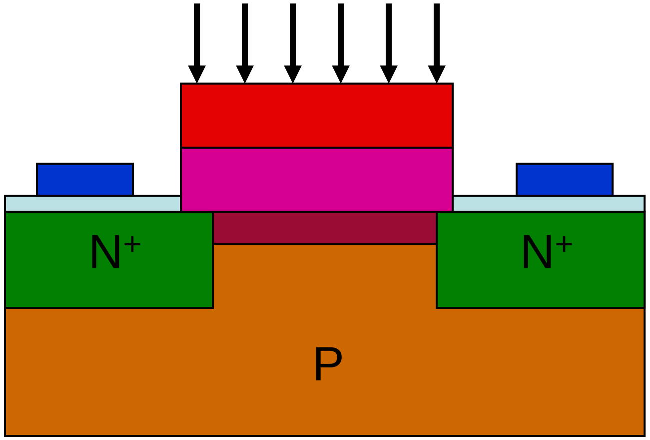
Low K_{gate} → Larger I_{gate} , Smaller delay



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

Why Dual-K and Dual-T ? (Low T_{gate} Vs High T_{gate})

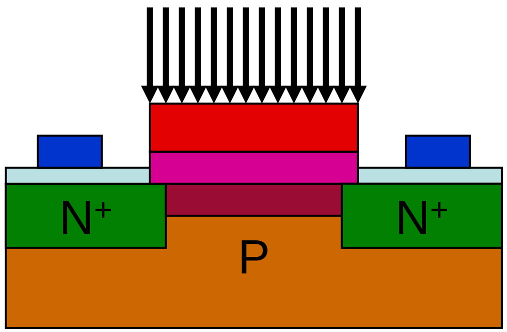




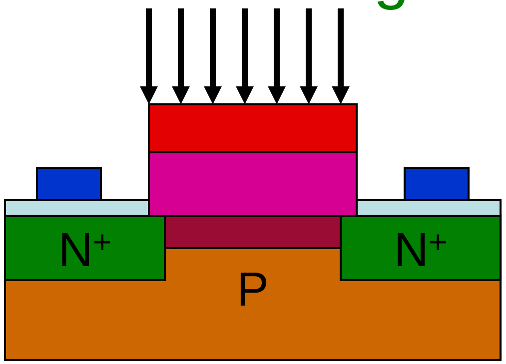
Low T_{gate} → Larger I_{gate} , Smaller delay

High T_{gate} → Smaller I_{gate} , Larger delay

Why Dual-K and Dual-T ?
(Four Combinations of K_{gate} & T_{gate})

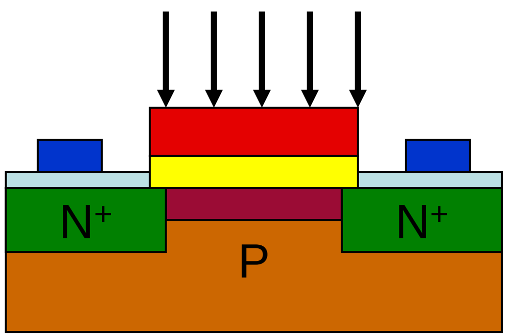


(1) K_1T_1

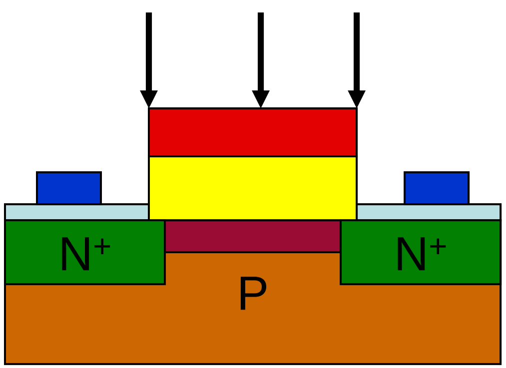


(2) K_1T_2

Tunneling Current ↓
Delay ↑



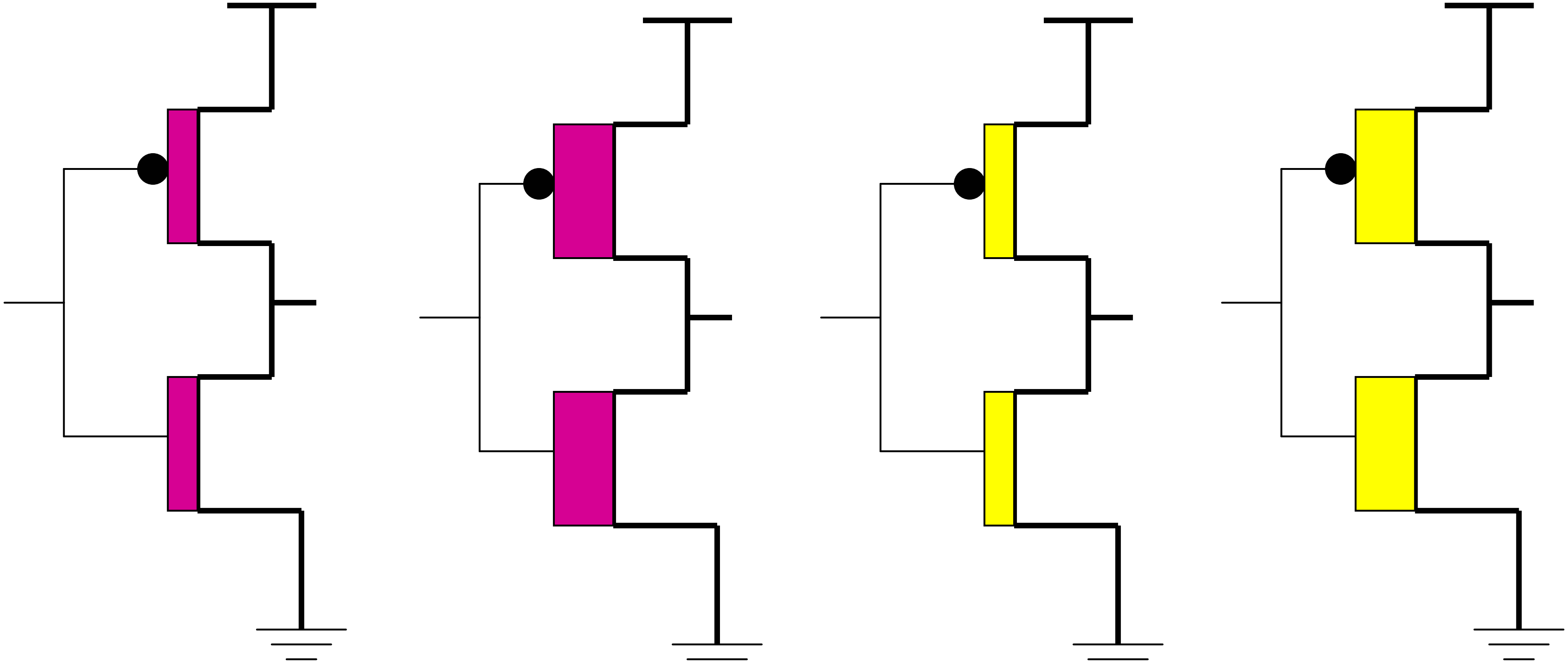
(3) K_2T_1



(4) K_2T_2

Why Dual-K and Dual-T ? (Example: Four Types of Inverter)

Assumption: all transistors of a logic gate are of same K_{gate} and equal T_{gate}.



(1) K_1T_1 (2) K_1T_2 (3) K_2T_1 (4) K_2T_2

Dielectrics for Replacement of SiO_2

- Silicon Oxynitride (SiO_xN_y) (K=5.7 for SiON)

- Silicon Nitride (Si_3N_4) (K=7)

- Oxides of :
  - Aluminum (Al), Titanium (Ti), Zirconium (Zr), Hafnium (Hf), Lanthanum (La), Yttrium (Y), Praseodymium (Pr),
  - their mixed oxides with SiO_2 and Al_2O_3

- **NOTE**: I_{gate} is still dependent on T_{gate} irrespective of dielectric material.

Related Works: Gate Leakage Reduction

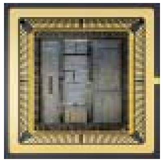
- **Inukai et. al. - CICC2000:** Boosted Gate MOS device using dual T_{ox} and dual V_{Th}
- **Rao et. al. - ESSCIRC2003:** Sleep state assignment for MTCMOS circuits
- **Lee et. al. - DAC2003 and TVLSI2004Feb:** Pin reordering to minimize gate leakage during standby positions
- **Sultania et. al. - DAC2004 and ICCD2004:** Heuristic for dual T_{ox} assignment
- **Sirisantana et. al. - IEEEDTC2004Feb and ICCD2000:** Use multiple channel lengths and multiple gate oxide thickness

Related Works : Summary

- Developed methods that use oxide of different thicknesses for tunneling reduction.
- Do not handle emerging dielectrics that will replace SiO_2 to reduce the tunneling current.
- Either consider ON or OFF state, but do not account both.
- Degradation in performance due to dual thickness approach.

Key Contributions of Our Work

- A new approach called dual dielectric assignment for tunneling current reduction.
- Considers dual thickness approach for both of the dielectrics.
- Explores a combined approach called DKDT (Dual-K of Dual Thickness) and proposes an assignment algorithm.
- Accounts both ON and OFF state gate tunneling.
- Logic cell characterization for average tunneling considering **non-SiO_2** dielectrics



DKDT Based Logic Synthesis

Input Circuit
↓
Technology Independent Optimization
↓
Intermediate Circuit
↓
Technology Mapping
↓
Mapped Circuit
↓
DKDT Assignment and Optimization
↓
Tunneling Optimized Circuit
↓
Placement and Routing
↓
Final Circuit Layout

Cell Library (4 Types) → Technology Mapping, DKDT Assignment and Optimization

- K_1T_1
- K_1T_2
- K_2T_1
- K_2T_2

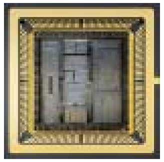
Optimization Problem Definition

Given a weighted directed acyclic graph G(V,E) it is required to find the best possible assignment of dielectric and thickness such that the total tunneling current is minimized and latency constraint (circuit performance) is satisfied.

Optimization Problem:

$$Minimize \quad \sum_{v_i \in V} I_{gate}(v_i)$$

such that, $\sum_{v_i \in V_P} D_i(v_i) \leq D_{CP}$



DKDT Assignment : Basis

- **Observation**: Tunneling current of logic gates increases and propagation delay decreases in the order K_2T_2, K_2T_1, K_1T_2, and K_1T_1 (where, $K_1 < K_2$ and $T_1 < T_2$).
- **Strategy**: Assign a higher order K and T to a logic gate under consideration
  - To reduce tunneling current
  - Provided increase in path-delay does not violate the target delay

DKDT Assignment : Algorithm

Step 1: Represent the network as a directed acyclic graph G(V, E).

Step 2: Initialize each vertex $v \in$ G(V, E) with the values of tunneling current and delay for K_1T_1 assignment.

Step 3: Find the set of all paths $P\{\Pi_{in}\}$ for all vertex in the set of primary inputs (Π_{in}), leading to the primary outputs Π_{out} .

Step 4: Compute the delay D_P for each path $p \in P\{\Pi_{in}\}$.

DKDT Assignment : Algorithm

Step 5: Find the critical path delay D_{CP} for K_1T_1 assignment.

Step 6: Mark the critical path(s) P_{CP}, where P_{CP} is subset $P\{\Pi_{in}\}$.

Step 7: Assign target delay $D_T = D_{CP}$.

Step 8: Traverse each node in the network and attempt to assign K-T in the order K_2T_2, K_2T_1, K_1T_2, and K_1T_1 to reduce tunneling while maintaining performance.

DKDT Assignment: Step8 Heuristic

(1) FOR each vertex $v \in G(V, E)$
(2) {
  (1) Determine all paths P_v to which node v belongs;
  (2) Assign K_2T_2 to v; Carry out Local Fanout Optimization (LFO); Determine timing closure and insert buffers (TCD/BI);
  (3) Calculate new critical delay D_{CP};
  (4) Calculate slack in delay as $\Delta D = D_T - D_{CP}$;
  (5) IF ($\Delta D < 0$) then
  (6) {
    (1) Assign K_2T_1 to v; LFO; TCD/BI; Calculate D_{CP}; Calculate ΔD;
    (2) IF ($\Delta D < 0$) then
    (3) {
      (1) Assign K_1T_2 to v; LFO; TCD/BI; Calculate D_{CP}; Calcu ΔD;
      (2) IF ($\Delta D < 0$) then
        (1) reassign K_1T_1 to v;
    (4) } // end IF
  (7) } // end IF
(3) // end FOR

Logic Cell Characterization : Load

- The Berkeley Predictive Technology Model (BPTM) has been used.
- The first step in the characterization was the selection of an appropriate capacitive load ($C_{Load} = 10 * C_{ggPMOS}$ used).
- The supply voltage is held at $V_{DD} = 0.7V$.
- We define the delay as the time difference between the 50% level of input and output.

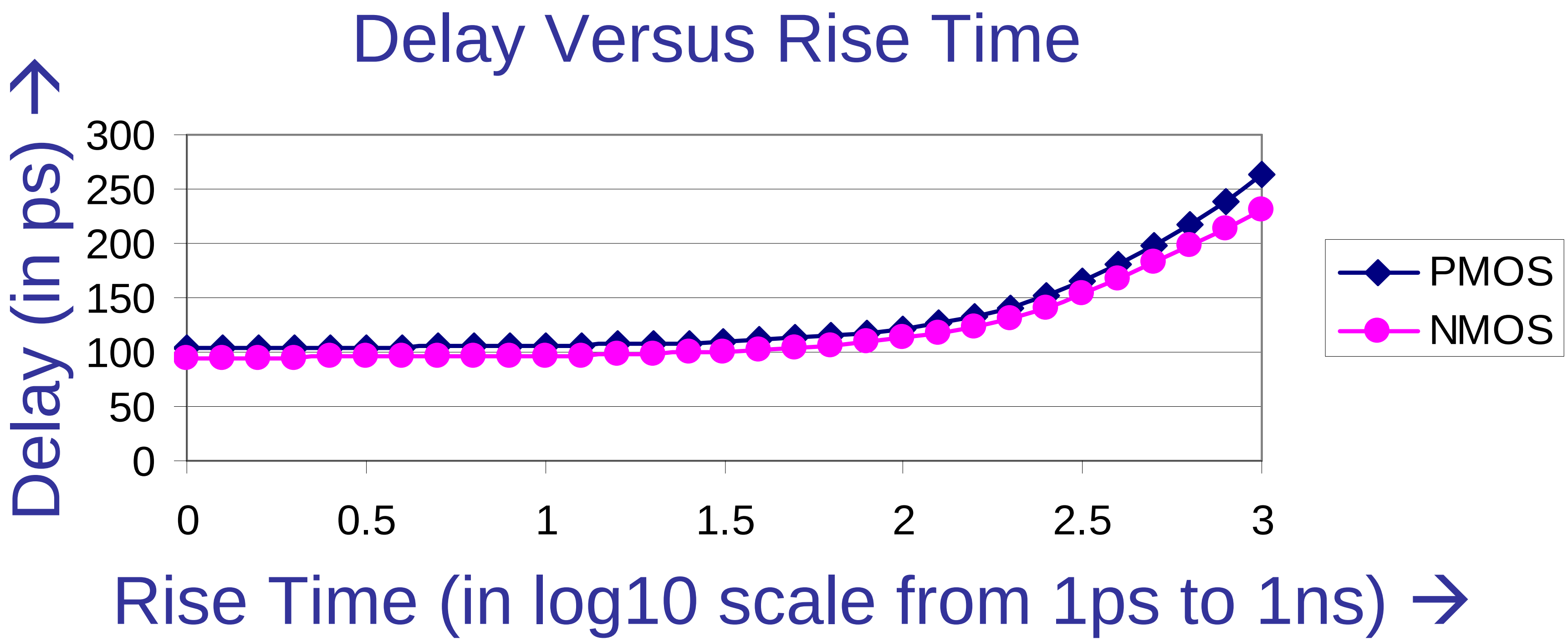
Logic Cell Characterization : t_r

- For worst-case scenarios in the development of the algorithm, we chose the maximum delay time [i. e. maximum (t_{pdr}, t_{pdf})].

- The effect of switching pulse rise time t_r was initially examined on the delay characteristics.

- To eliminate an explicit dependence of the algorithm results on t_r , we chose a value that is realistic yet does not affect the delay significantly.

Logic Cell Characterization : t_r



t_r = 10ps

Cell Characterization : K_{gate} Modeling

- The effect of varying dielectric material was modeled by calculating an equivalent oxide thickness (T^*_{ox}) according to the formula:

$$\mathbf{T^*_{ox} = (K_{gate} / K_{ox})\, T_{gate}}$$

- Here, K_{gate} is the dielectric constant of the gate dielectric material other than SiO_2, (of thickness T_{gate}), while K_{ox} is the dielectric constant of SiO_2.

Cell Characterization : T_{gate} Modeling

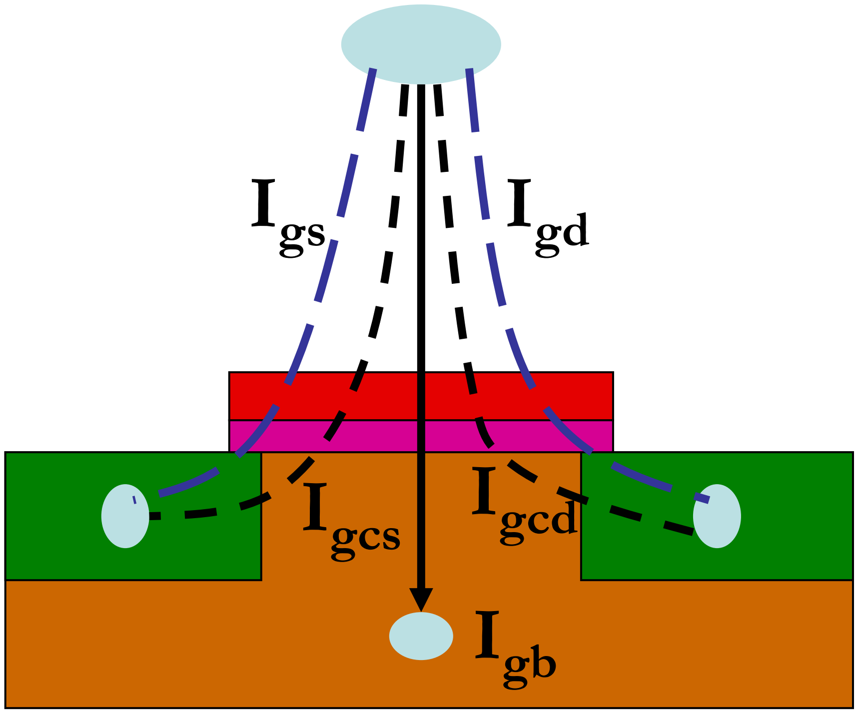
- The effect of varying oxide thickness T_{ox} was incorporated by varying TOXE in SPICE model.

- Length of the device is proportionately changed to minimize the impact of higher dielectric thickness on the device performance :

$$L^* = (T^*_{ox} / T_{ox})\, L$$

- Length and width of the transistors are chosen to maintain (W:L) ratio of (4:1) for NMOS and (8:1) for PMOS.

Logic Cell Characterization : I_{gate}

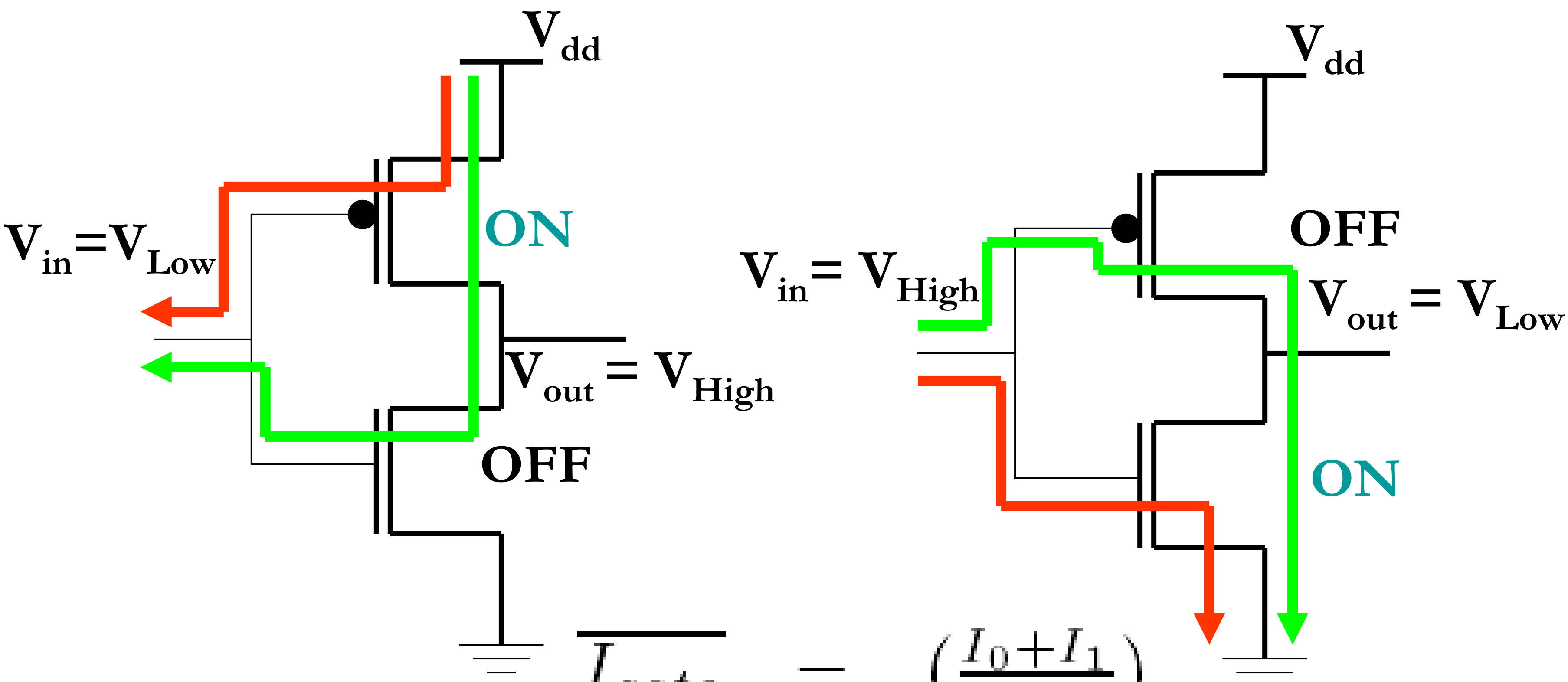


BSIM4 Model

- Calculated by evaluating both the source and drain components
- For a MOS, $I_{gate} = (|I_{gs}| + |I_{gd}| + |I_{gcs}| + |I_{gcd}| + |I_{gb}|)$
- Values of individual components depends on states, ON or OFF

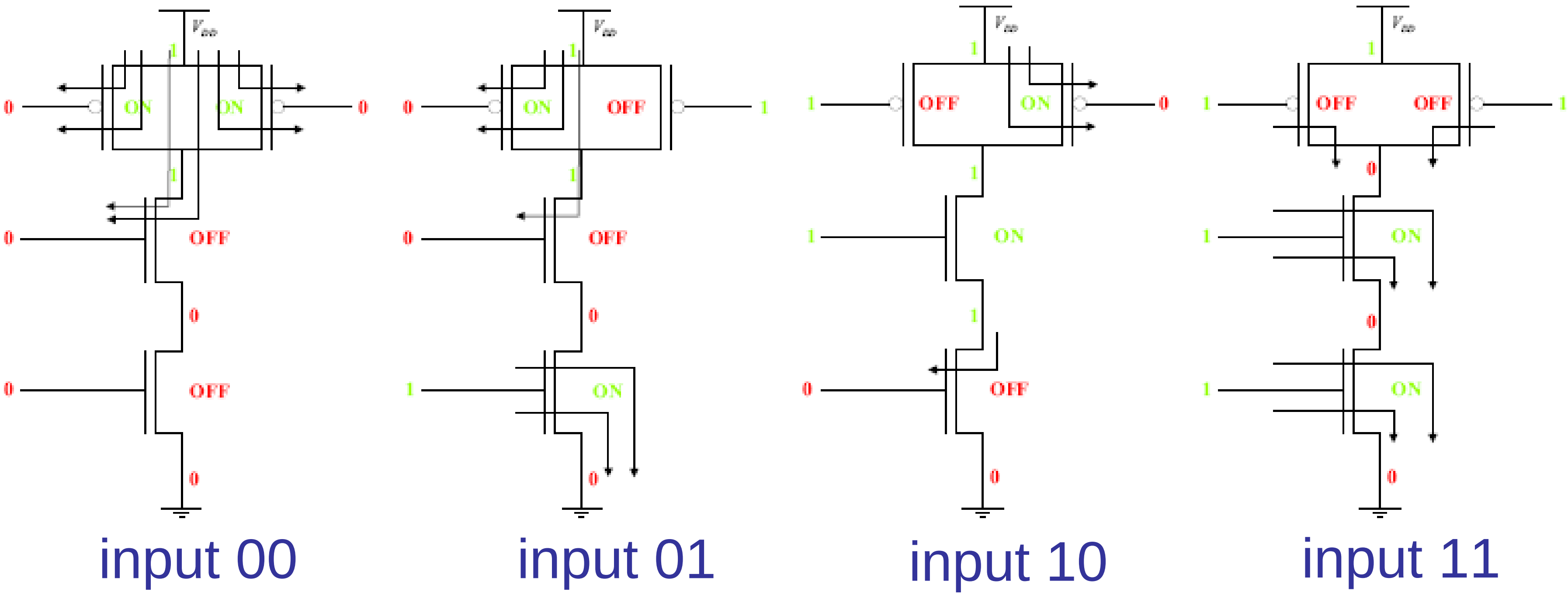
Logic Cell Characterization : An Inverter

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



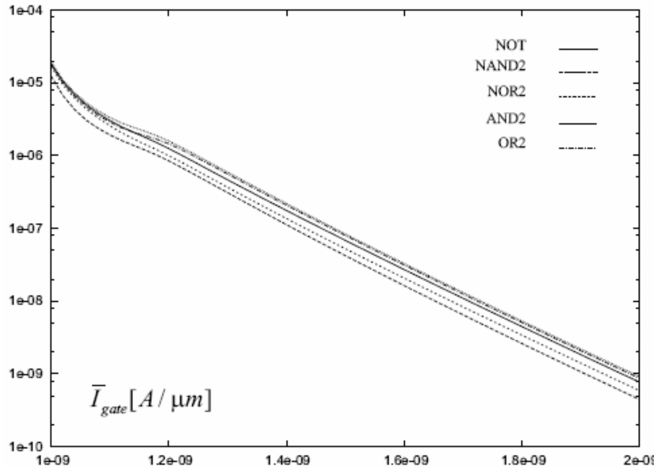
$$\overline{I_{gate}} = \left(\frac{I_0+I_1}{2}\right)$$

Logic Cell Characterization : A NAND Gate

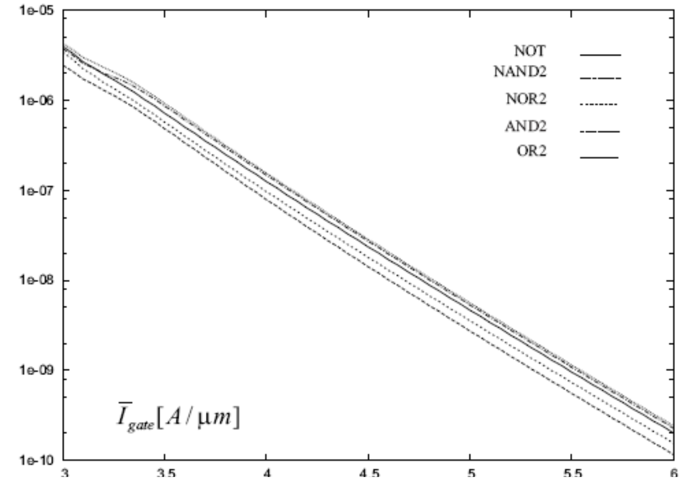


$$\overline{I_{gate}} = \left(\frac{I_{00}+I_{01}+I_{10}+I_{11}}{4}\right)$$

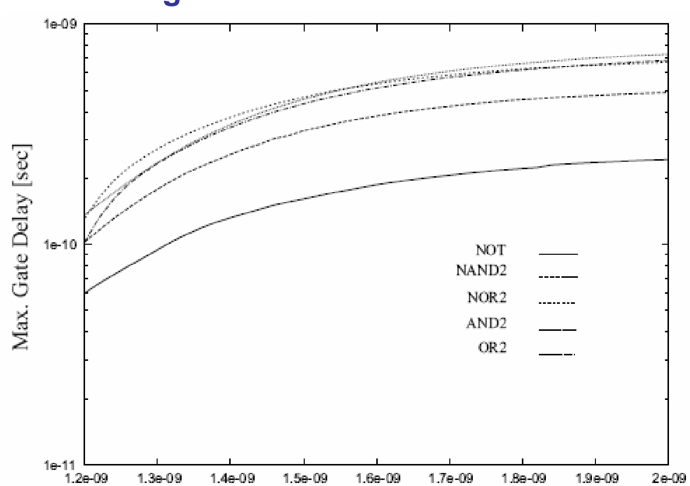
Cell Characterization: 45nm Tech Cells



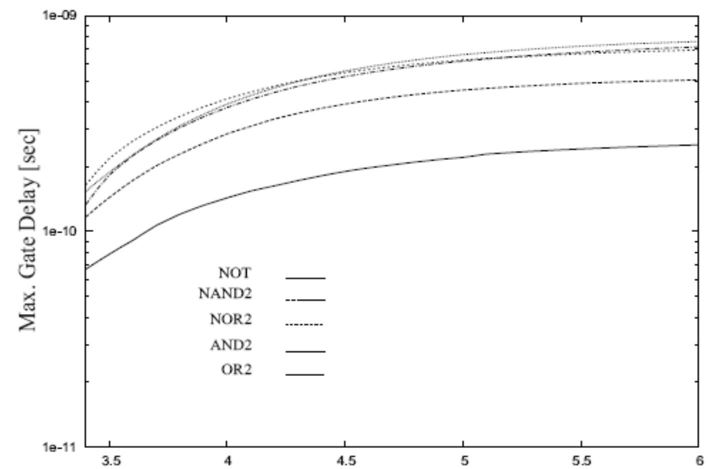
I_{gate} Vs Thickness



I_{gate} Vs Dielectric Constant



T_{pd} Vs Thickness



T_{pd} Vs Dielectric Constant

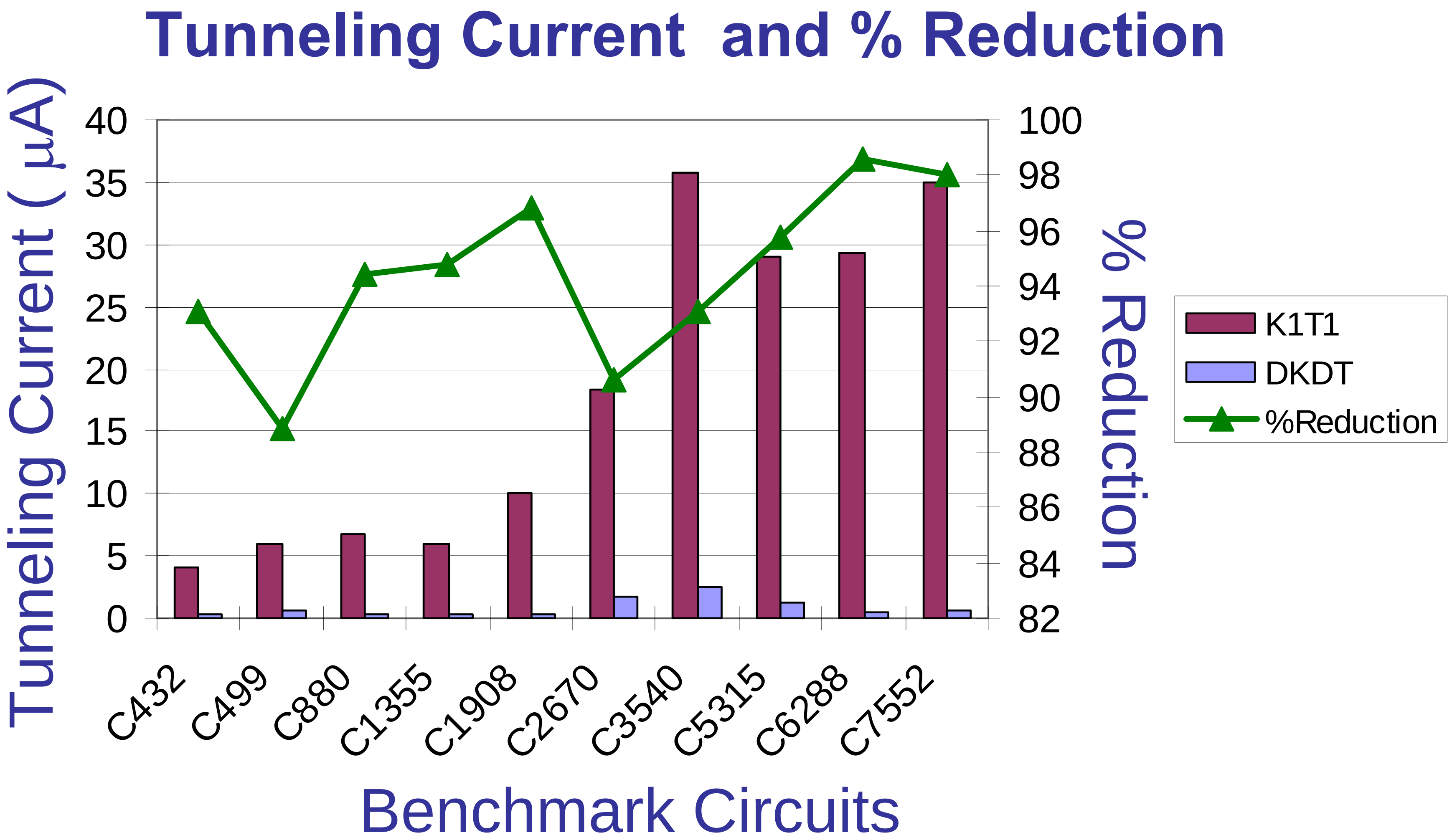
Experimental Results: Setup

- DKDT algorithm integrated with SIS, and tested on the ISCAS'85 benchmarks.
- Used K_1 = 3.9 (for SiO_2), K_2 = 5.7 (for SiON), T_1 = 1.4nm, and T_2 = 1.7nm for our experiments.
- T_1 is chosen as the default value from the BSIM4.4.0 model card and value of T_2 is intuitively chosen

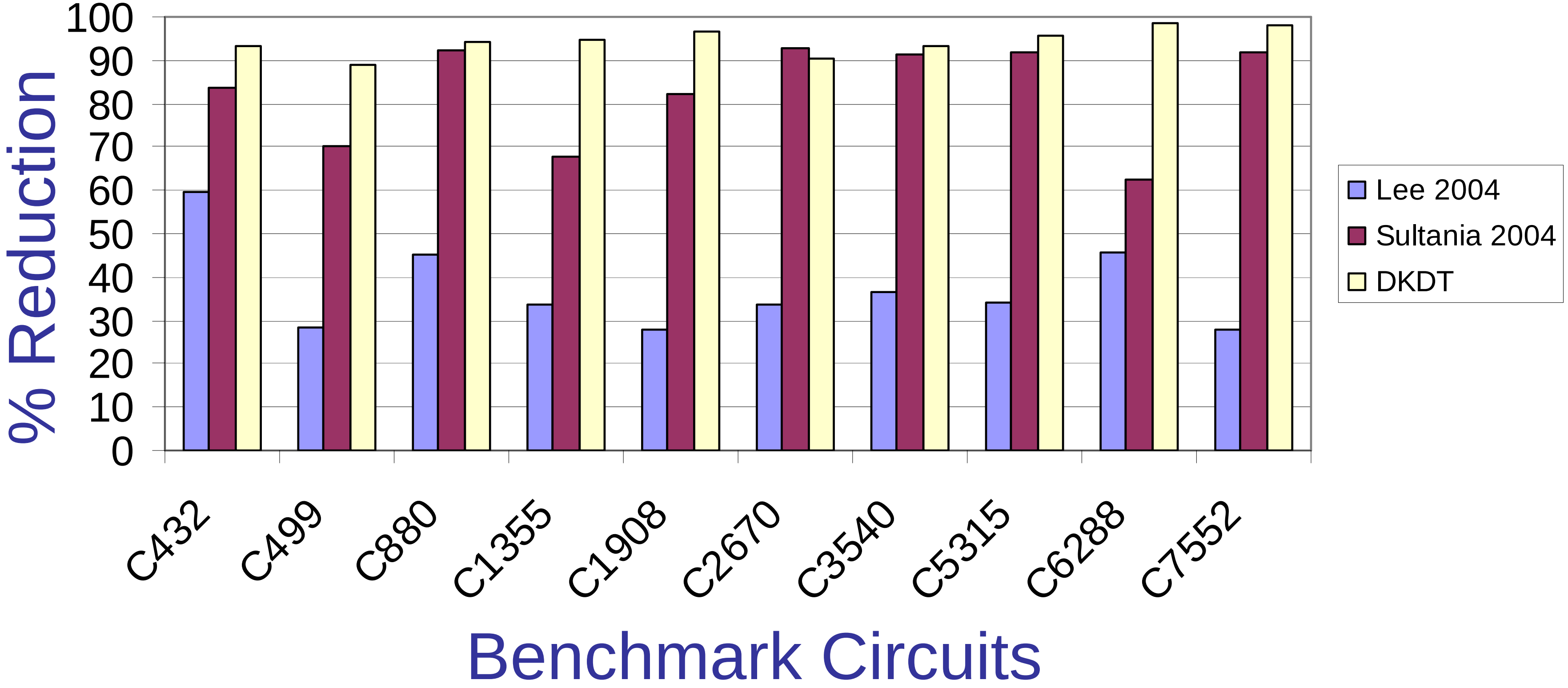
Experimental Results : Table

Bench. Circuits	Gates	Critical Delay (ps)	Current for K_1T_1 (nA)	Current for DKDT (nA)	%Reduction
C432	160	3.848	3949.45	253.26	93.58
C499	202	2.054	5708.55	590.45	89.66
C880	383	6.162	6537.02	337.84	94.83
C1355	546	2.054	5708.55	274.644	95.19
C1908	880	6.675	9714.74	287.72	97.04
C2670	1193	24.64	17863.33	1560.67	91.27
C3540	1669	18.23	34637.15	2215.74	93.60
C5315	2406	23.10	28156.87	1098.80	96.10
C6288	2406	24.89	28474.64	372.56	98.69
C7552	3512	26.44	33899.46	625.84	98.15

Experimental Results ….



Experimental Results: A Broad View



NOTE: DKDT has not time penalty.

Conclusions and Future Works

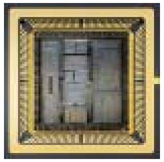
- New approach for tunneling current reduction accounting for both ON and OFF states.

- Algorithm could perform the such assignment for circuits in reasonable amount of time.

- Experiments prove significant reductions in tunneling current without performance penalty.

Conclusions and Future Works

- Modeling for other high-K dielectrics is under progress.
- Development of optimal assignment algorithm can be considered.
- Tradeoff of tunneling, area and performance needs to be explored.
- DKDT based design may need more masks for the lithographic process during fabrication.



Thank You