ILP based Gate Leakage Optimization using DKCMOS Library during RTL Synthesis

Saraju P. Mohanty

VLSI Design and CAD Laboratory (VDCL)

Dept of Computer Science and Engineering

University of North Texas.

Email: smohanty@cse.unt.edu

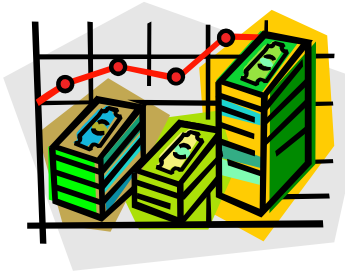
Acknowledgment: This work is partially supported by NSF award number 0702361.

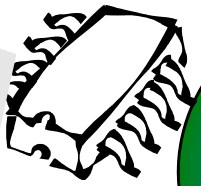
Outline of the Talk

- Introduction and Motivation
- Background Research
- DKCMOS Technology
- Gate Leakage Modeling
- Problem Formulation
- ILP-Based Gate Leakage Optimization
- Datapath Component Library
- Experimental Results
- Conclusions and Future Works

Introduction and Motivation

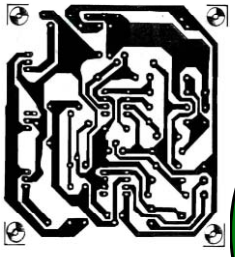
Why Low Power?



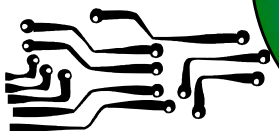


Packaging costs

Chip and system cooling costs



Power supply rail

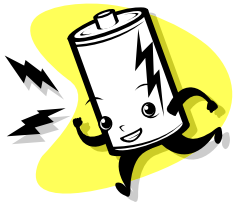


Power affects

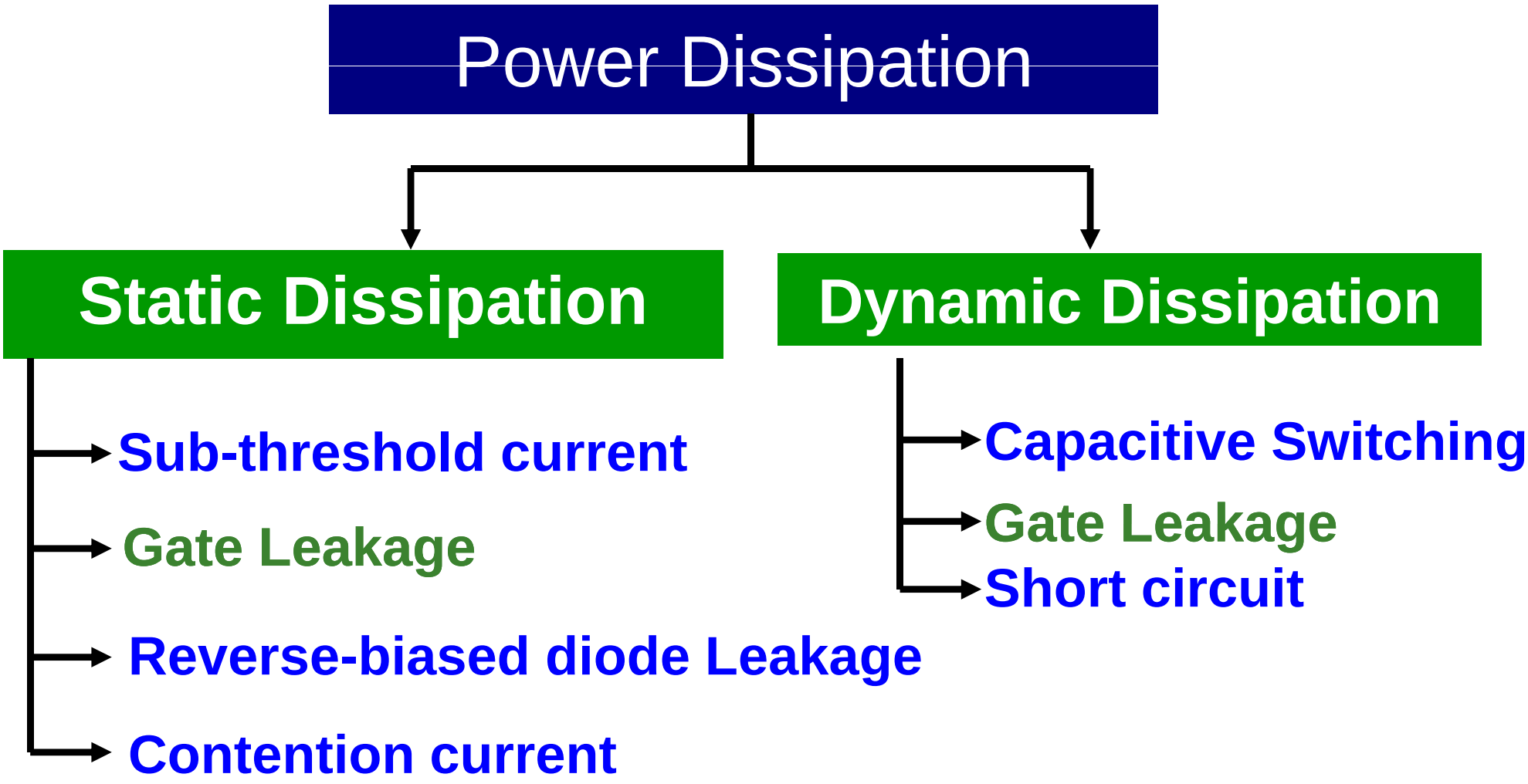
Noise and reliability

Environmental

Battery life



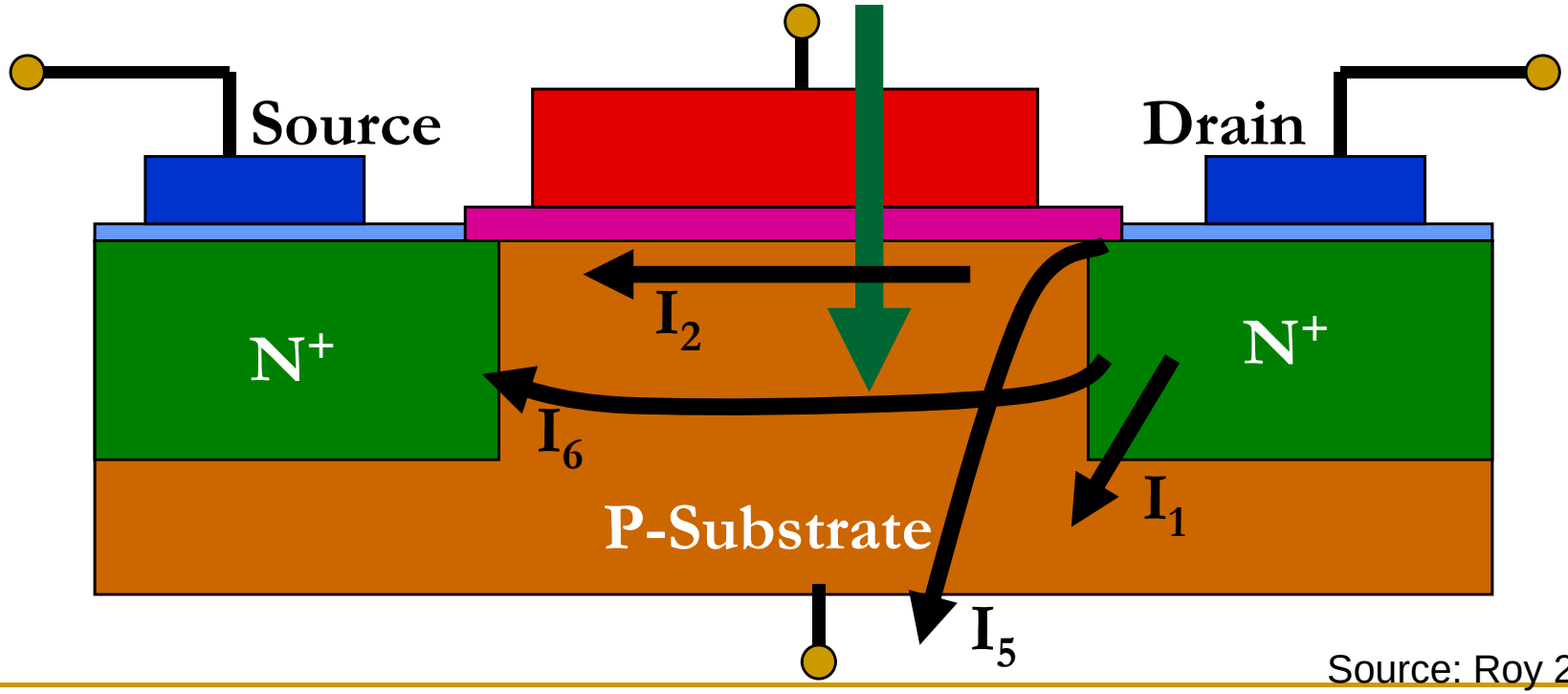
Power Dissipation in CMOS



Source: Weste and Harris 2005

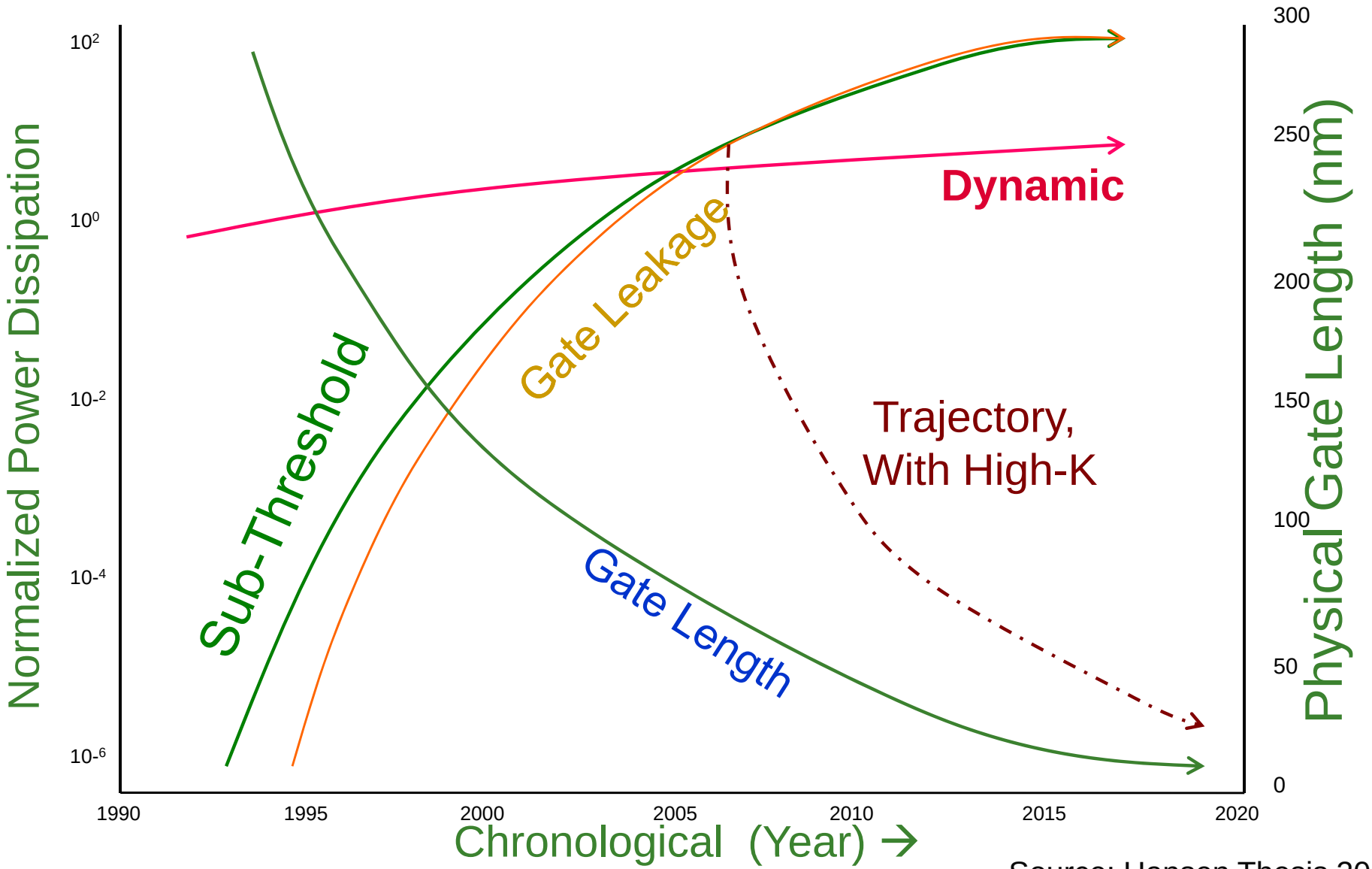
Leakages in CMOS

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

Key Contributions of this Paper

- Introduces dual dielectric (DKCMOS) technology for architectural level gate leakage reduction.
- Presents a ILP-based optimization for gate leakage reduction during behavioral synthesis.
- ILP-based optimization uses DKCMOS technology.
- The algorithm minimizes the leakage delay product (LDP) of datapath circuits for given resource constraints.

Background Research

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:

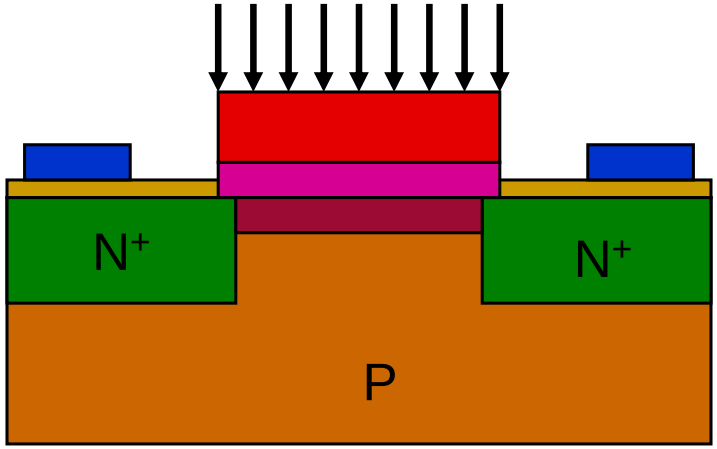
- Mohanty - VLSI Design 2006: Dual-T_{ox} approach for reduction of gate leakage current.
- Mohanty - ISQED 2006: Simulated annealing algorithms using dual-K or dual-Tox.

Related Works : Logic / Transistor Level Gate Leakage Reduction

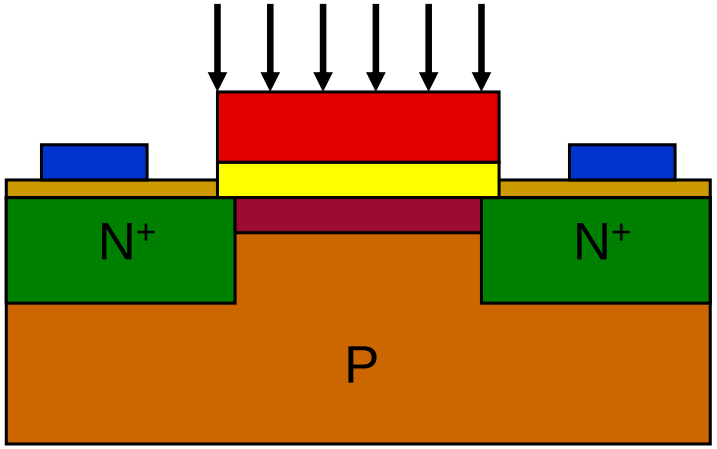
- Lee - TVLSI2004 : Pin reordering to minimize gate leakage during standby positions of logic gates.
- Sultania – TVLSI Dec 2005 and Sultania - DAC2004 : Heuristic for dual-T_{ox} assignment for gate leakage and delay tradeoff.
- Sirisantana - IEEE DTC Jan-Feb 2004: Use multiple channel lengths and multiple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

DKCMOS Technology

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



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

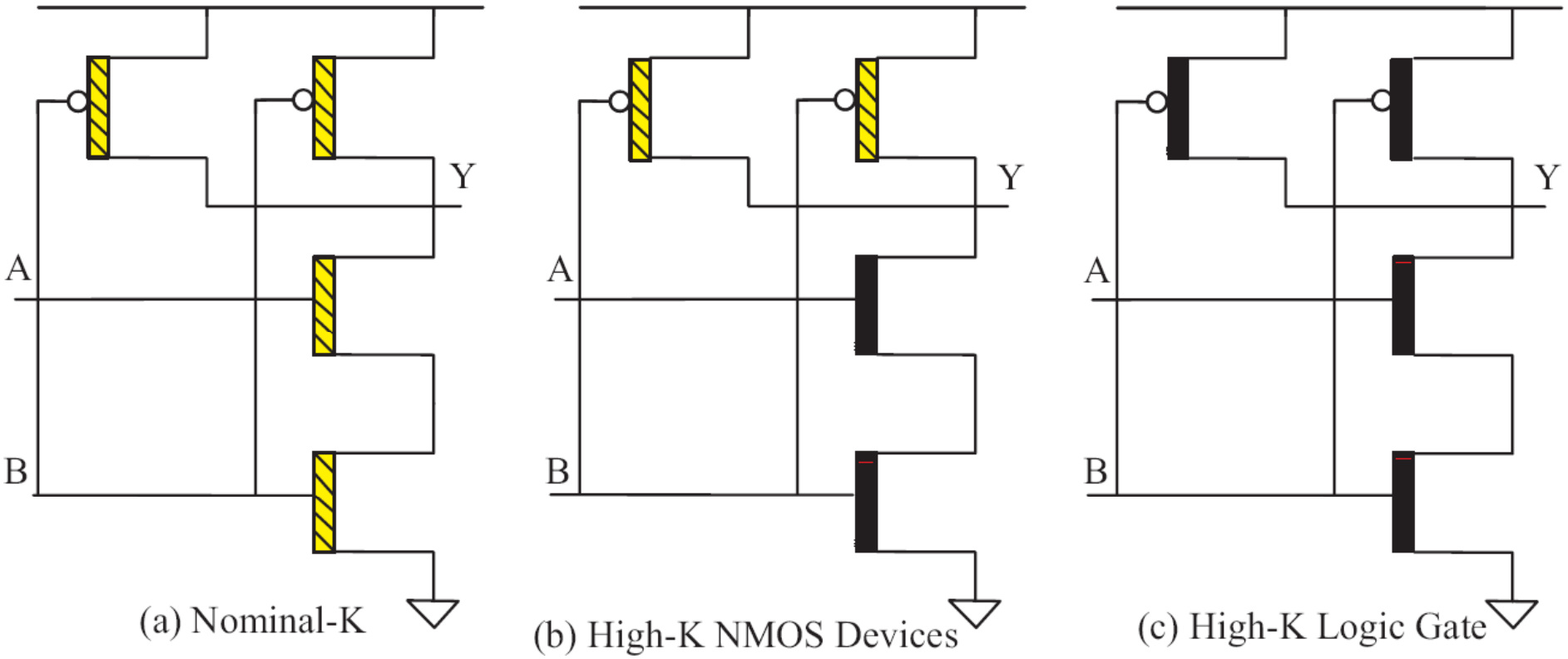


High K_{gate} → Smaller I_{gate} , Larger delay

Dielectrics for Replacement of 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

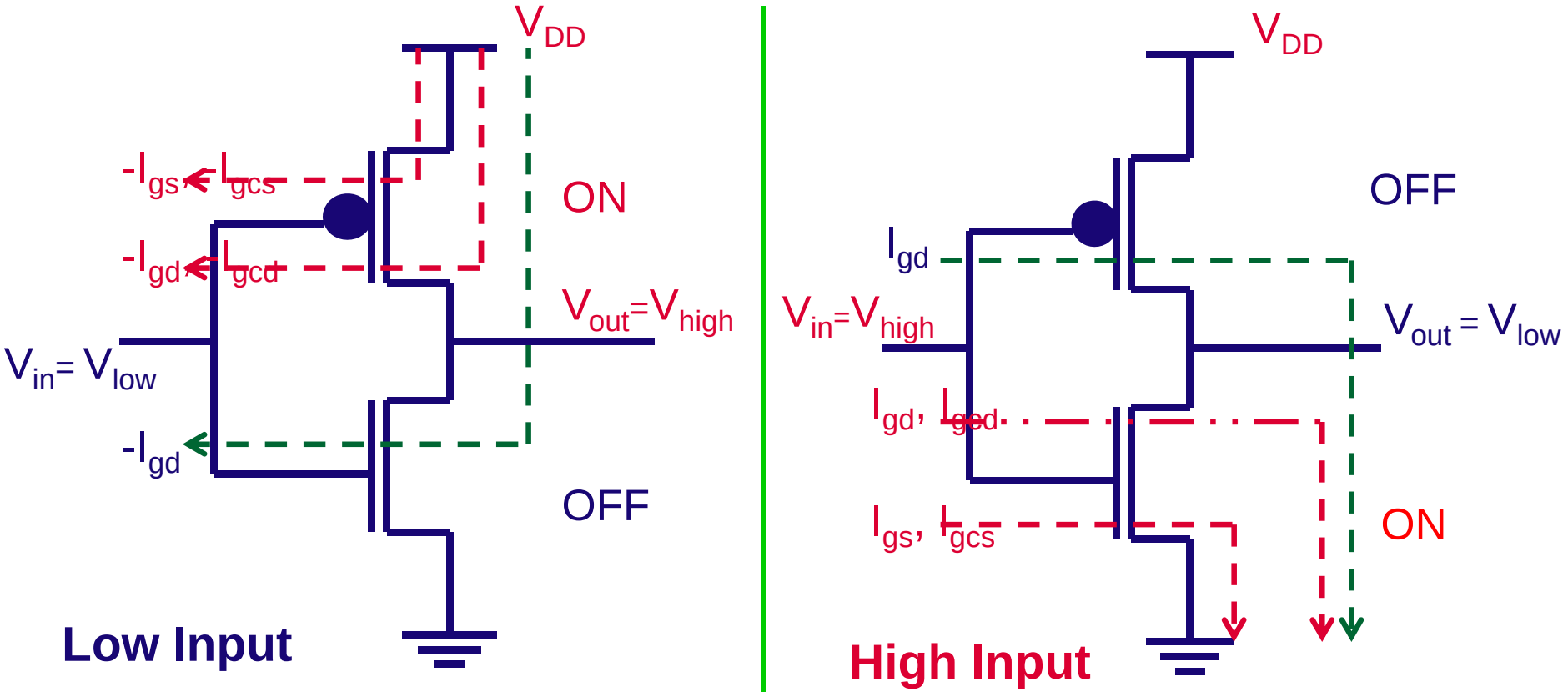
The DKCMOS Technology



Gate Leakage Modeling

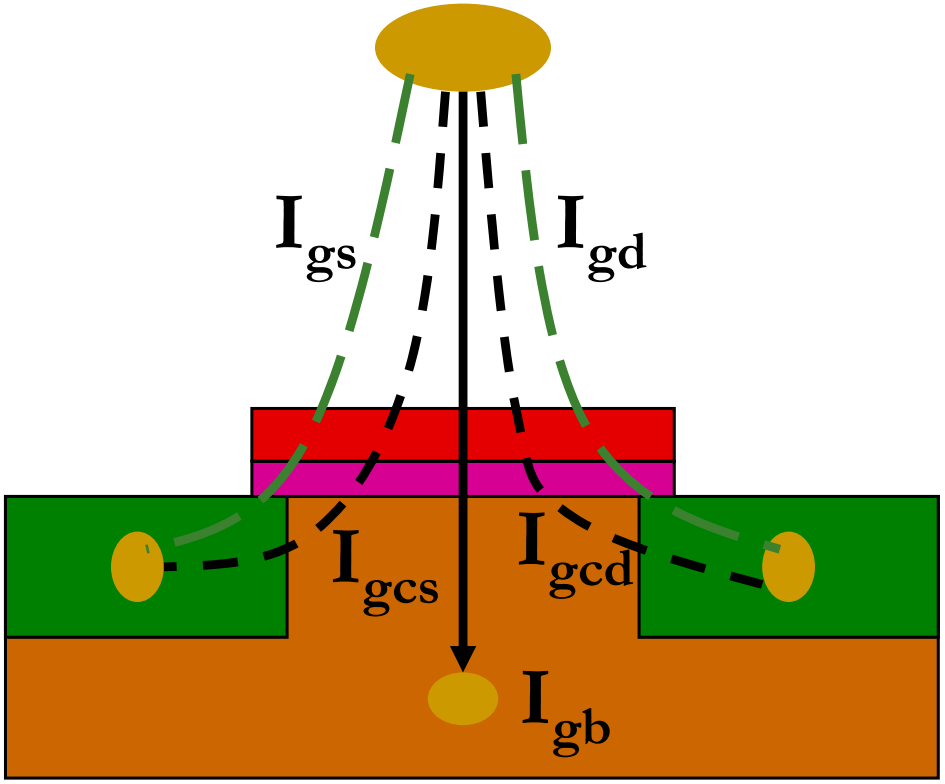
Gate Leakage Paths in an Inverter

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



NOTE: Gate to body component found to be negligible.

I_{gate} Modeling for a Device



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

K_{gate} Modeling

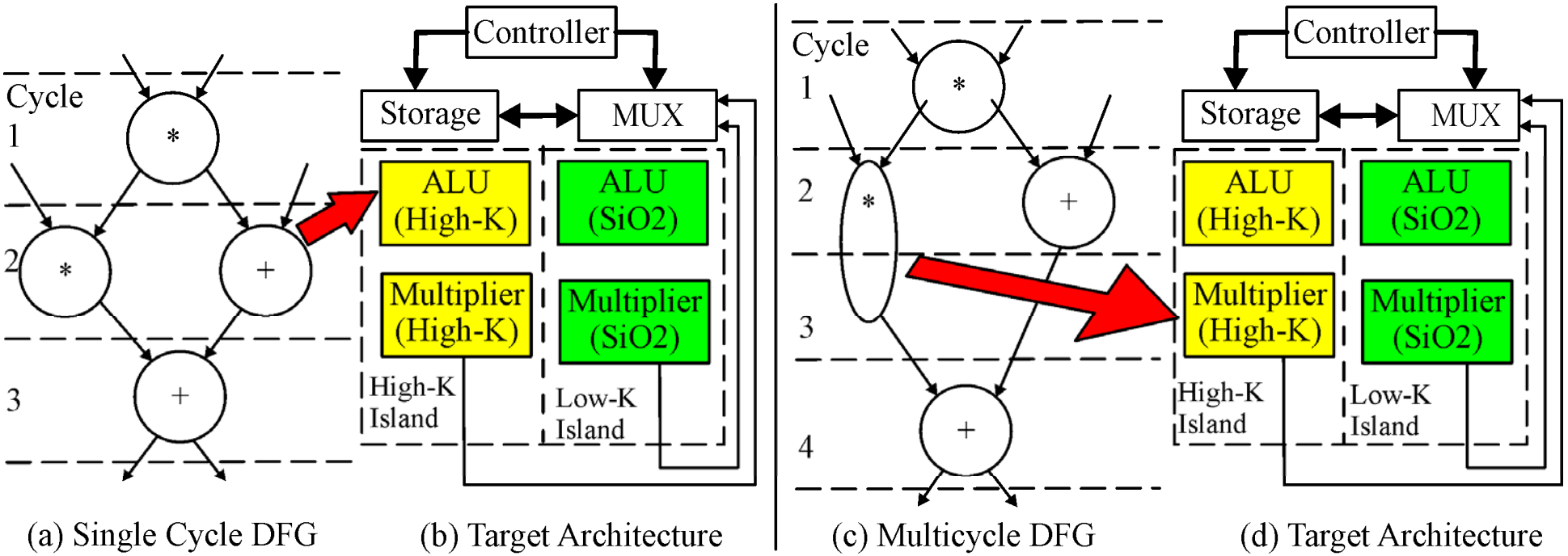
- The effect of varying dielectric material was modeled by calculating an equivalent oxide thickness (T^*_{ox}) according to the formula:

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

- 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.

Problem Formulation

Target Architecture



High-K and Low-K Islands

Problem Formulation

- Given an unscheduled data flow graph (UDFG) $G_U(V,E)$, it is required to find the scheduled data flow graph (SDFG) $G_S(V,E)$ with appropriate resource binding such that the total gate leakage and delay product (LDP) is minimized under given resource.

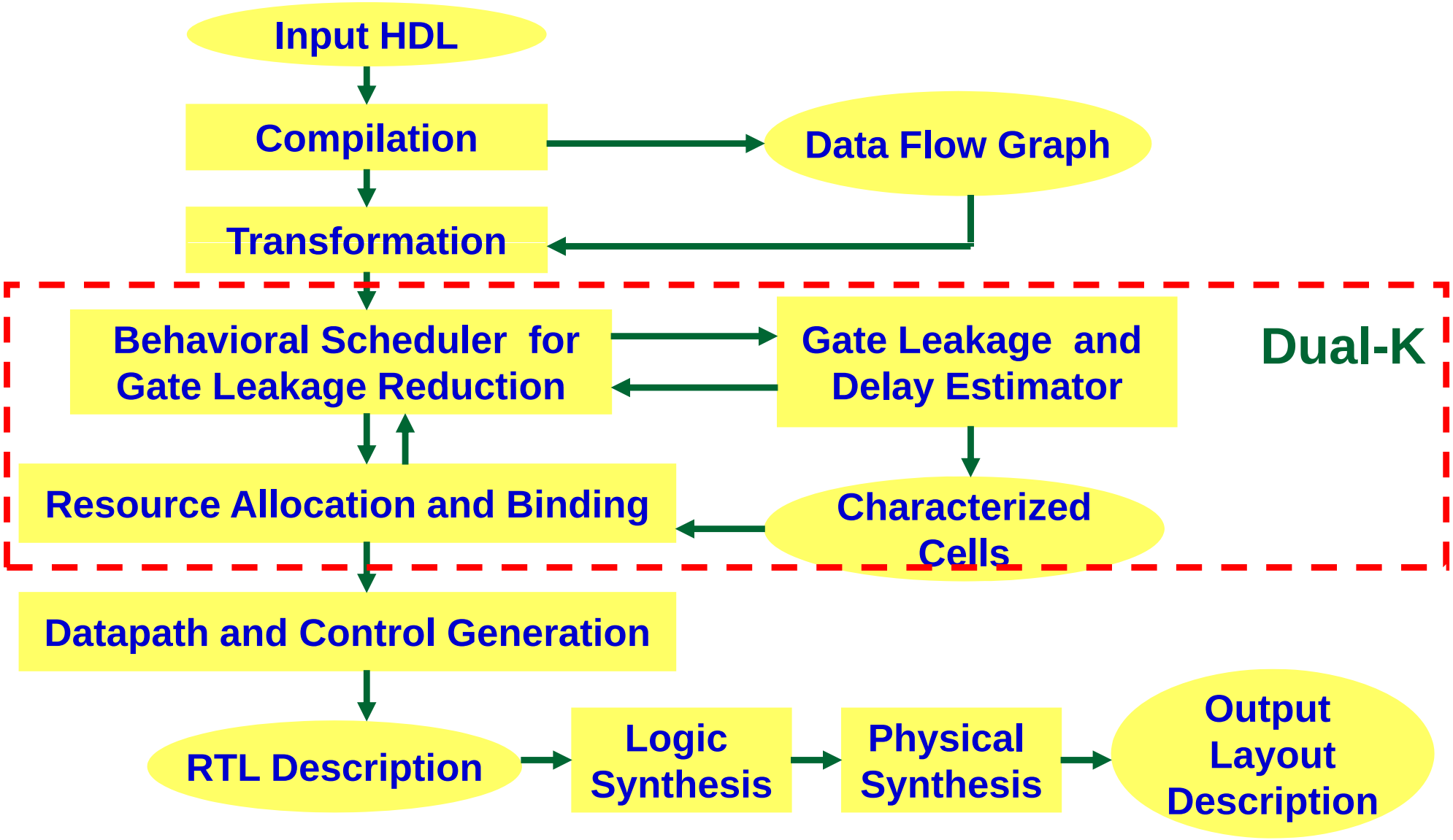
$$\text{Minimize: } LDP(DFG)$$

$$\text{Allocated}(R_{t,K}) \leq \text{Available}(R_{t,K}), \forall c \in N$$

$$LDP(DFG) = \sum_{c=1}^{N} \sum_{\forall v_{i,c}} P_{gate}(v_{i,c}) \times d_c$$

ILP based Gate Leakage Optimization

Synthesis for Low Gate Leakage



ILP Based LDP Optimization …

1. Preprocess given behavioral description to construct a sequencing DFG.
2. Perform simulations to estimate gate leakage and delay of RTL units.
3. Construct resource allocation table and available resource table based on input resource constraints.
4. Obtain ASAP and ALAP schedules of the input DFG.
5. Determine the number of different resources for each K using the resource allocation table.

ILP Based LDP Optimization

6. Modify both ASAP and ALAP schedules obtained above using the number of resources found in previous step.
7. Construct the mobility graph based on above schedules.
8. Fix the total number of clock cycles as the maximum of modified ASAP and ALAP schedules' control step.
9. Model the ILP formulations of the DFG using AMPL.
10. Obtain the final solution by solving the ILP formulations.

ILP Formulations …

- Objective Function: The objective is to minimize the LDP of the whole DFG over all control steps. This can be expressed using decision variable as:

$$\text{Minimize} : LDP(DFG)$$

$$\text{Minimize} : \sum_{l}\sum_{i}\sum_{k} X_{i,K,l,(l+L_{i,K}-1)} \times LDP(i,K)$$

ILP Formulations

Uniqueness Constraints: These constraints are represented as, *∀i, 1 ≤ i ≤ V ,*

$$\sum_{K} \sum_{l=S_i}^{S_i+E_i+1-L_{i,K}} X_{i,K,l,(l+L_{i,K}-1)} = 1$$

Precedence Constraints: These constraints should also ensure the multicycling and are modeled as, *∀i, j, $v_i \in Pred_{vj}$,*

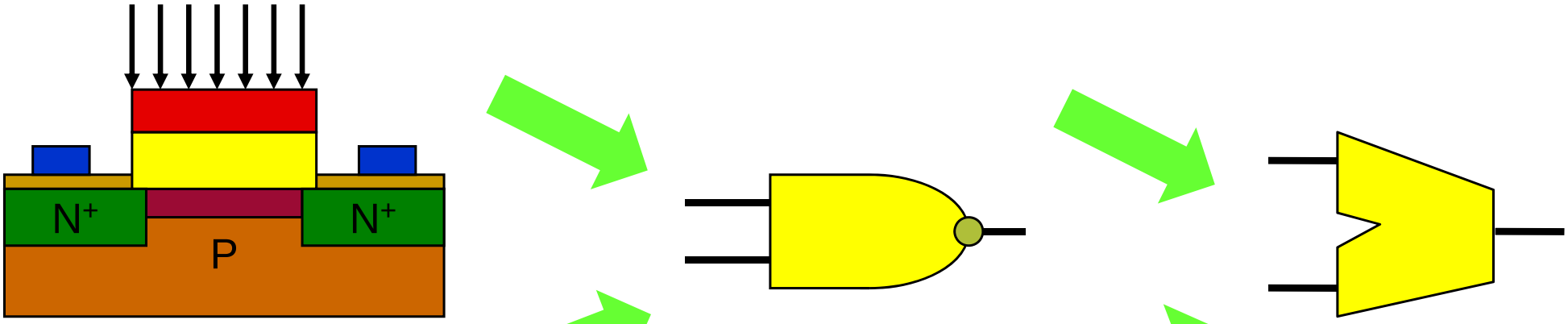
$$\sum_{K} \sum_{l=S_i}^{E_i} (l + L_{i,K} - 1) \times X_{i,K,l,(l+L_{i,K}-1)} - \sum_{K} \sum_{l=Sj}^{E_j} l \times X_{j,K,l,(l+L_{j,K}-1)} \leq -1$$

Resource Constraints: These constraints ensure that each cycle uses resources not exceeding available number of resources and are enforced as, *∀K and ∀l, 1 ≤ l ≤ N,*

$$\sum_{i \in R_{t,K}} X_{i,K,l,(l+L_{i,K}-1)} \leq M_{t,K}$$

Datapath Components Library

Datapath Component Library :
3 Level Bottom-up Hierarchical Approach



Transistor Level

Logic Level

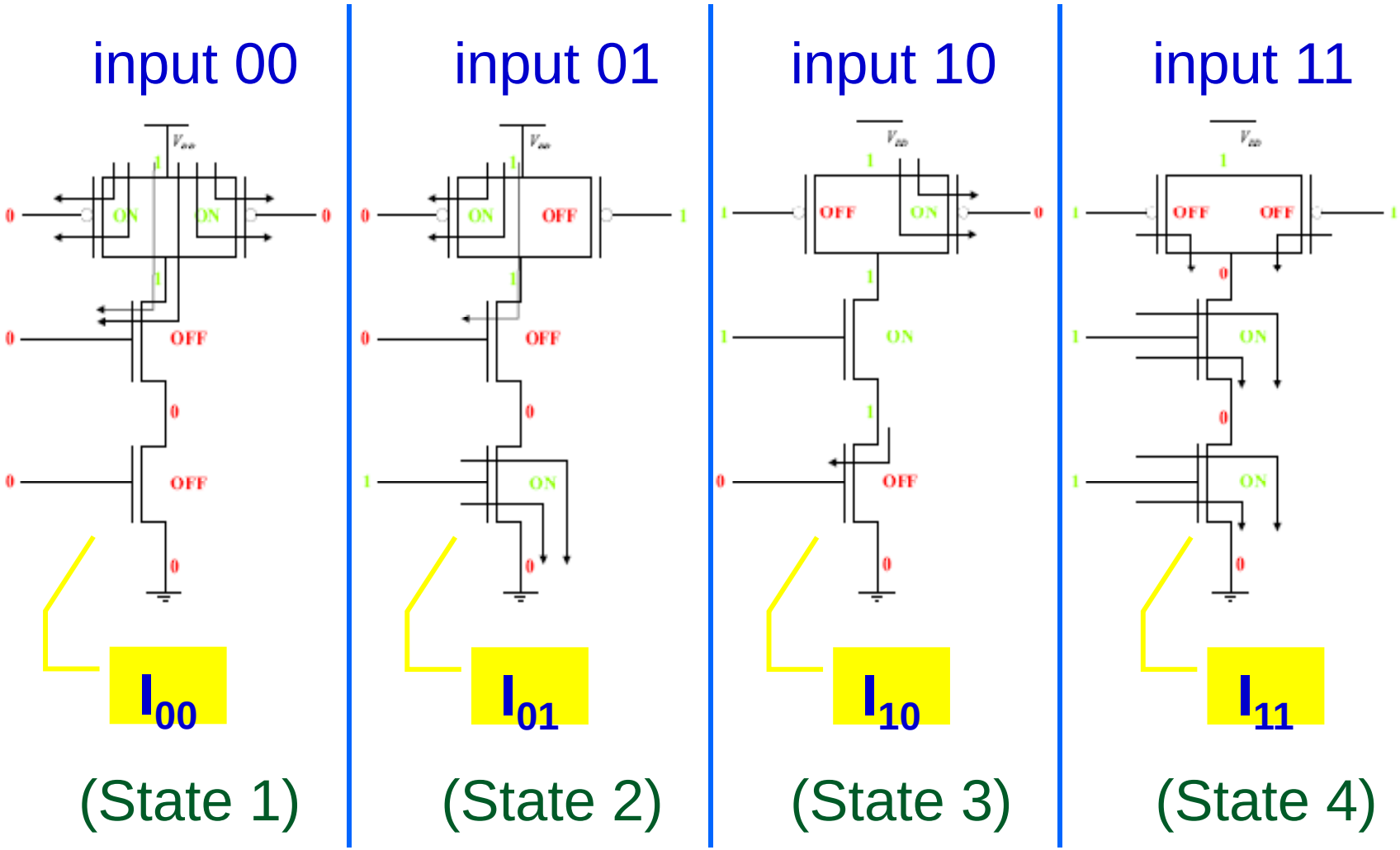
Datapath Components

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 …



Datapath Component Library

- The gate leakage current for a specific state of a logic gate is then calculated by:

$$I_{gate\,\mathrm{Logic}_{state}} = \sum_{\forall MOS_i} \left| I_{gate\,\mathrm{MOS}}[i] \right|$$

- The gate leakage of a n-bit RTL unit is calculated as:

$$I_{gate\,\mathrm{R}} = \sum_{j=1}^{n_{total}} \Pr ob(state)\, I_{gate\,\mathrm{NAND}\,j_{state}}$$

- The propagation delay of an *n*-bit functional unit is:

$$T_{pd\,\mathrm{R}} = \sum_{i=1}^{n_{cp}} T_{pd\,NANDi}$$

Experimental Results

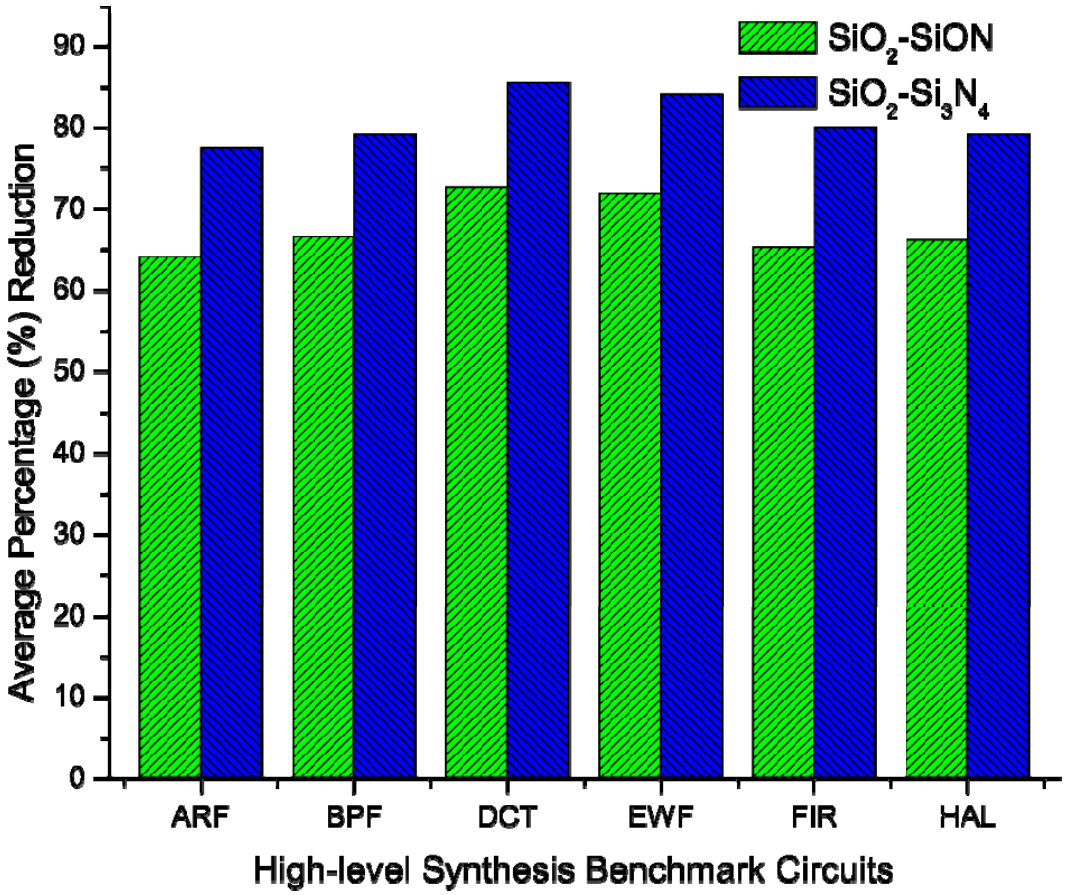
Experimental Results …

- 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.

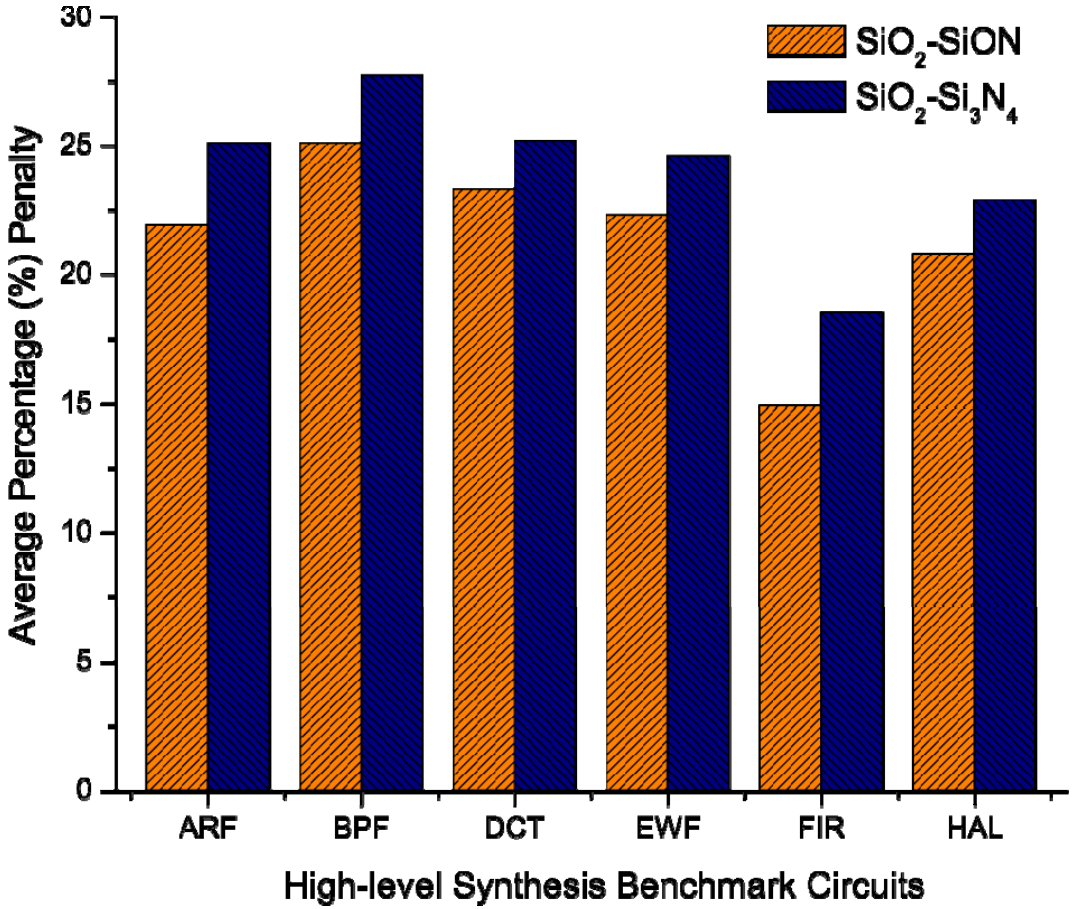
- Two pairs of dual dielectric are considered:
  (i) SiO_2(K=3.9) – SiON (K=5.7)
  (ii) SiO_2(K=3.9) – Si_3N_4(K=7)

- 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 …



Experimental Results



Conclusions and Future Works

Conclusions

- This paper presents a new process driven technique called DKCMOS for reduction of gate leakage during RTL synthesis.
- The ILP based algorithm does scheduling and assignment for gate leakage reduction for different resource constraints.
- Experimental results reveal significant reductions in gate leakage with the use of this technology, thus proving its effectiveness.

Future Works

- Further exploration of this technique is the incorporation of process variation.
- The effectiveness of DKCMOS for subthreshold leakage needs investigation.
- The ultimate objective is to extend the work on gate leakage current to provide a broader solution to the problem of power dissipation in all its forms at the behavioral level.
- The area overhead due to the use two separate islands (high-K and low-K) will also investigated.