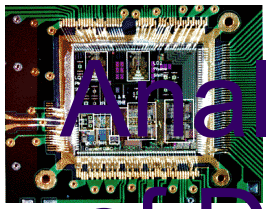


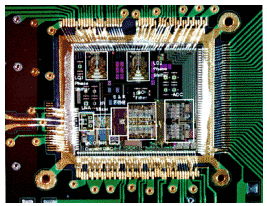
Analytical Modeling and Reduction of Direct Tunneling Current during Behavioral Synthesis of Nanometer CMOS Circuits

Saraju P. Mohanty, Valmiki Mukherjee and Ramakrishna Velagapudi

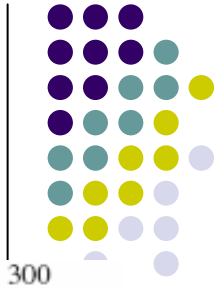
Dept of Computer Science and Engineering

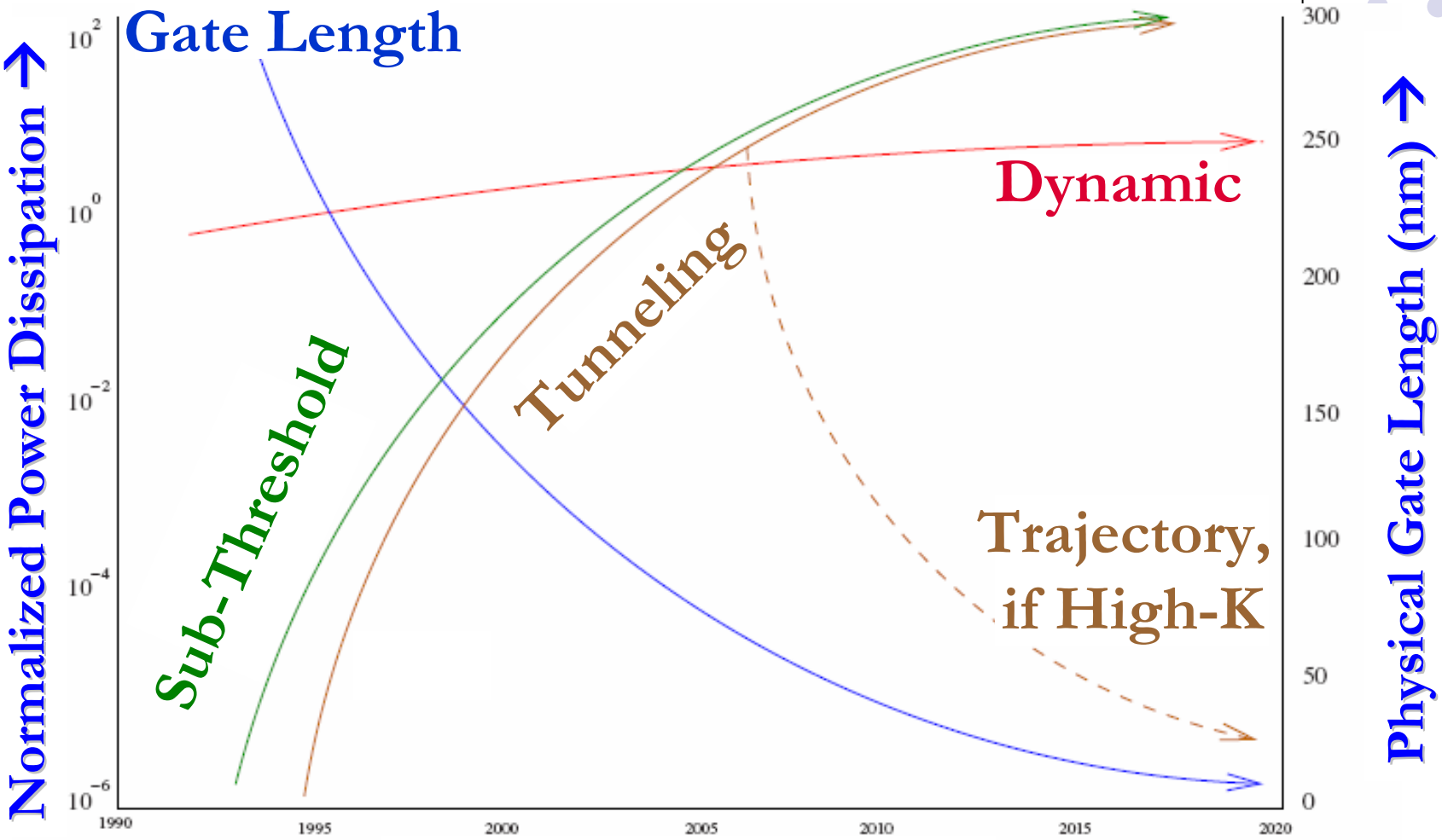
University of North Texas

Email: smohanty@cs.unt.edu

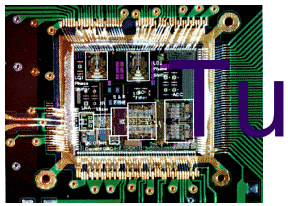


Power Dissipation Trend



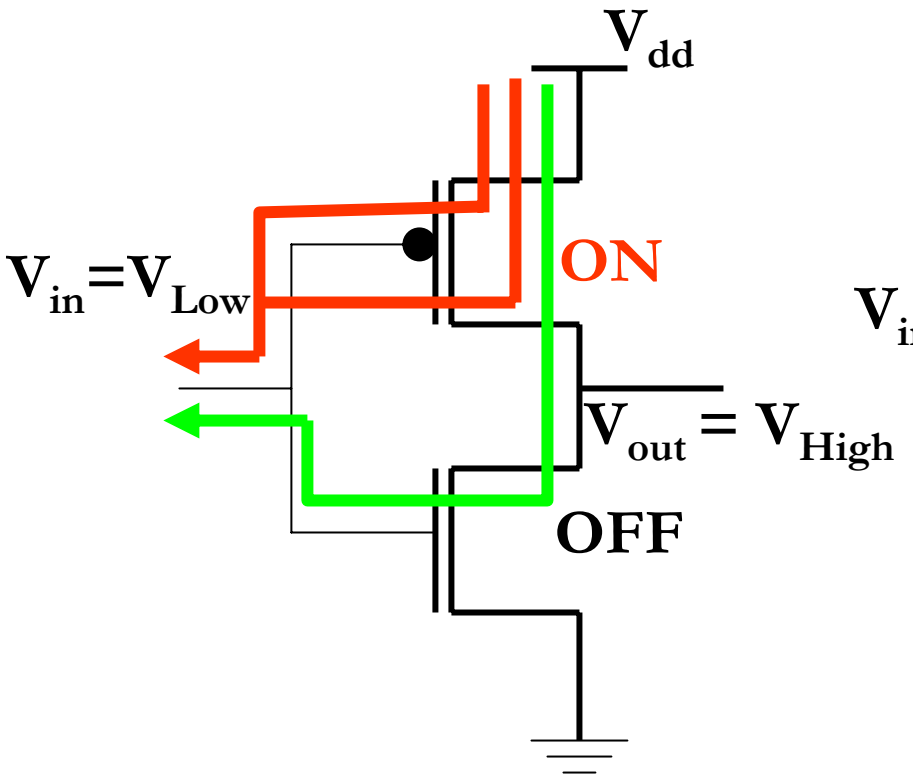


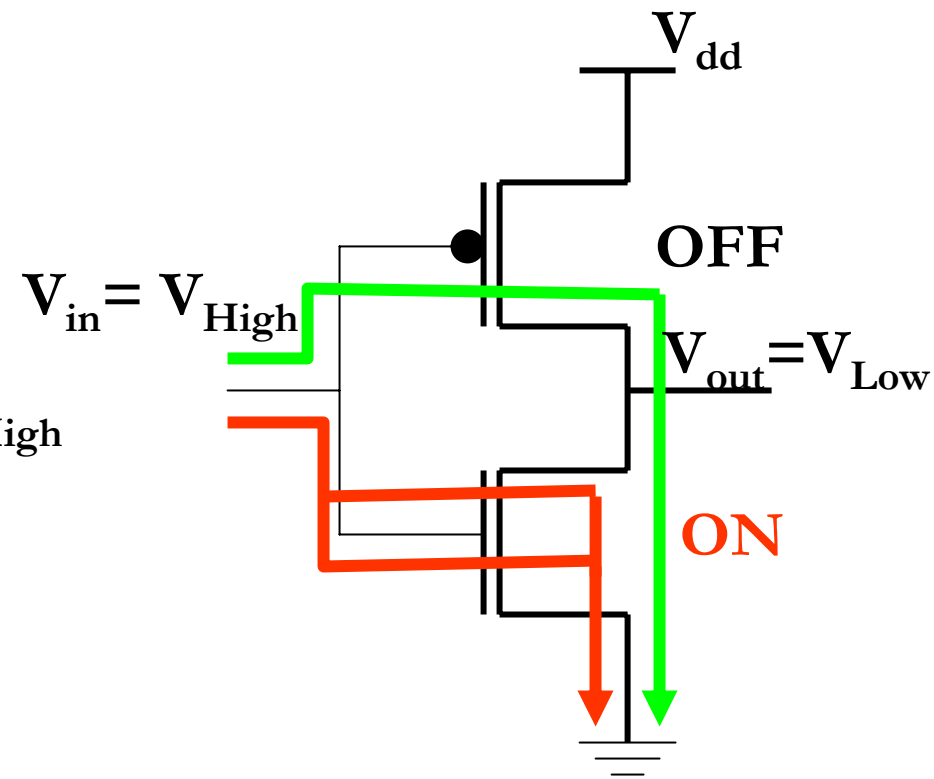
Source: Hansen 2004

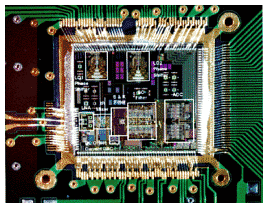


Tunneling Paths in an Inverter

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





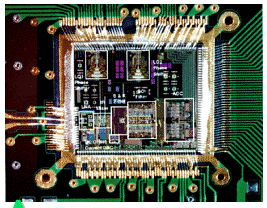


Dual-T_{OX} ?

- Gate oxide tunneling current I_{ox} [Kim2003, Chandrakasan2001] (α is an experimentally derived factor) :

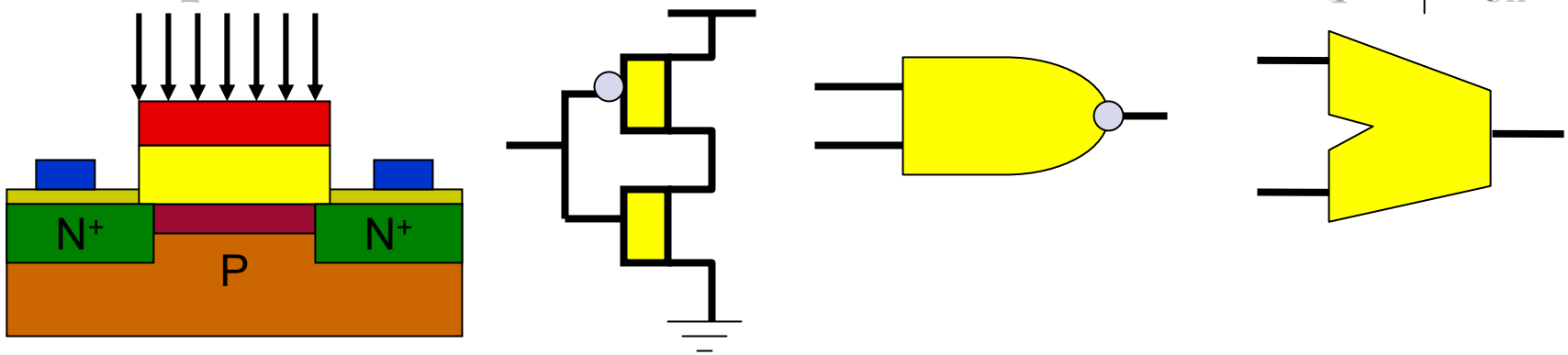
$$I_{ox} \propto (V_{dd} / T_{ox})^2 \exp(-\alpha T_{ox}/V_{dd})$$

- **Options for reduction of tunneling current** :
  - Decreasing of supply voltage V_{dd} (will play its role)
  - Increasing gate SiO_2 thickness T_{ox} (delay increases)
- We believe that combined use of high-T_{ox} resources and low-T_{ox} resources can reduce the gate oxide tunneling current of a datapath with little compromise in circuit performance.

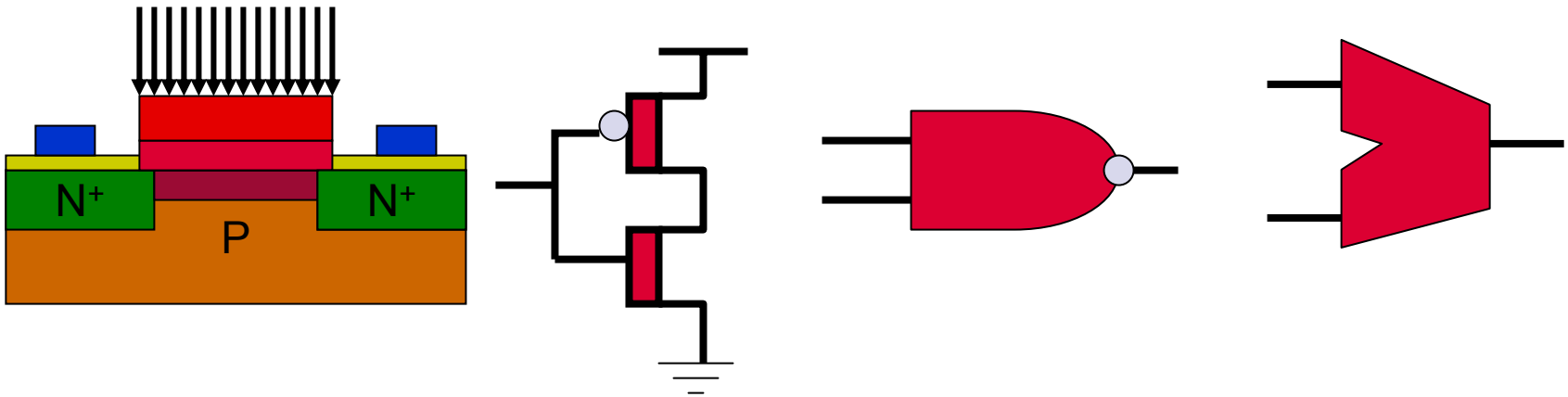


Dual-T_{OX} ?

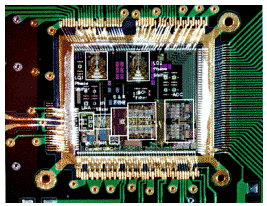
Assumption: All transistors of a resource are of equal T_{ox}.



High T_{ox} Resource : Smaller I_{ox}, but Larger delay



Low T_{ox} Resource : Larger I_{ox}, but Smaller delay



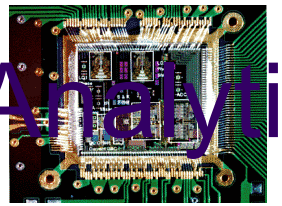
Related Works

Behavioral Level Subthreshold :

- **Khouri, TVLSI 2002 :** Algorithms for subthreshold leakage power analysis and reduction using dual threshold voltage.
- **Gopalakrishnan, ICCD2003 :** MTCMOS approach for reduction of subthreshold current

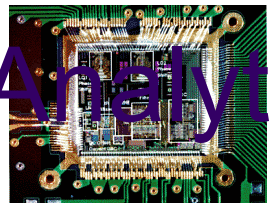
Logic / Transistor Level Tunneling :

- **Lee, TVLSI2004 :** Pin reordering to minimize gate leakage during standby positions of NOR and NAND gates.
- **Sultania, DAC2004 :** Heuristic for dual T_{ox} assignment for tunneling current and delay tradeoff.



Analytical Model for Tunneling Current

- We assumed that resources such as adders, subtractors, multipliers, dividers, are constructed using 2-input NAND.
- There are total n_{total} NAND gates in the network of NAND gates constituting a *n*-bit functional unit.
- n_{cp} number of NAND gates are in the critical path.



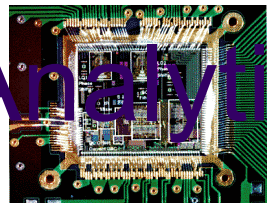
Analytical Model for Tunneling Current

- The tunneling current of a *n*-bit functional unit :

$$I_{DT\text{FU}} = \sum_{j=(1 \rightarrow ntotal)} Pr_j \sum_{\text{MOSi } \varepsilon \text{ NAND}j} Pr_i I_{DTi}$$

Pr_j is the probability that input of the NAND gate is at logic “0”, and Pr_i is the probability that inputs of the transistors that are connected in the parallel i.e. PMOS are at logic “0”.

- The average tunneling current for a NAND is calculated as $I_{DT\text{NAND}} = \sum_{\text{MOS}i \ \varepsilon \text{ NAND}} Pr_i I_{DTi}$



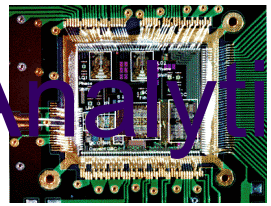
Analytical Model for Tunneling Current

- The direct tunneling current of a MOS :

$$I_{DT} = \frac{WLq^3V_{ox}^2}{16\pi^2\phi_B T_{ox}^2}\exp\left[-\frac{4\sqrt{2m_{eff}}\phi_B^{1.5}T_{ox}}{3\hbar qV_{ox}}\left\{1-\left(1-\frac{V_{ox}}{\phi_B}\right)^{1.5}\right\}\right]$$

- The voltage across the MOSFET gate dielectric V_{ox} is expressed as follows:

$$V_{ox} = V_{gs} - V_{fb} - \Psi_S - V_{poly}$$

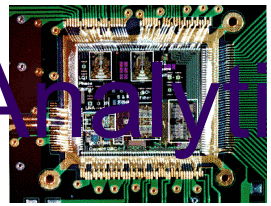


Analytical Model for Tunneling Current

- By solving a quadratic equation we obtain an expression for V_{ox}.

$$V_{OX} = \frac{\sqrt{1 - 2(V_{fb} + \psi_S - V_{gs})\left(\frac{\varepsilon_{OX}^2}{q\varepsilon_{Si} N_{poly} T_{OX}^2}\right)} - 1}{\left(\frac{\varepsilon_{OX}^2}{q\varepsilon_{Si} N_{poly} T_{OX}^2}\right)}$$

- The flat-band voltage V_{fb} can be obtained using the expression $(qN_{channel}T^2_{ox} / 2\varepsilon_{Si})$.
- Ψ_S = 2 * Fermi-Level, for strong inversion.



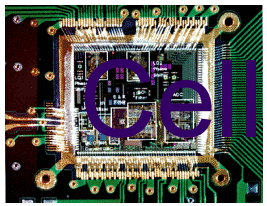
Analytical Model for Propagation Delay

- The critical path delay of a *n* bit functional unit using the NAND gates as building blocks :

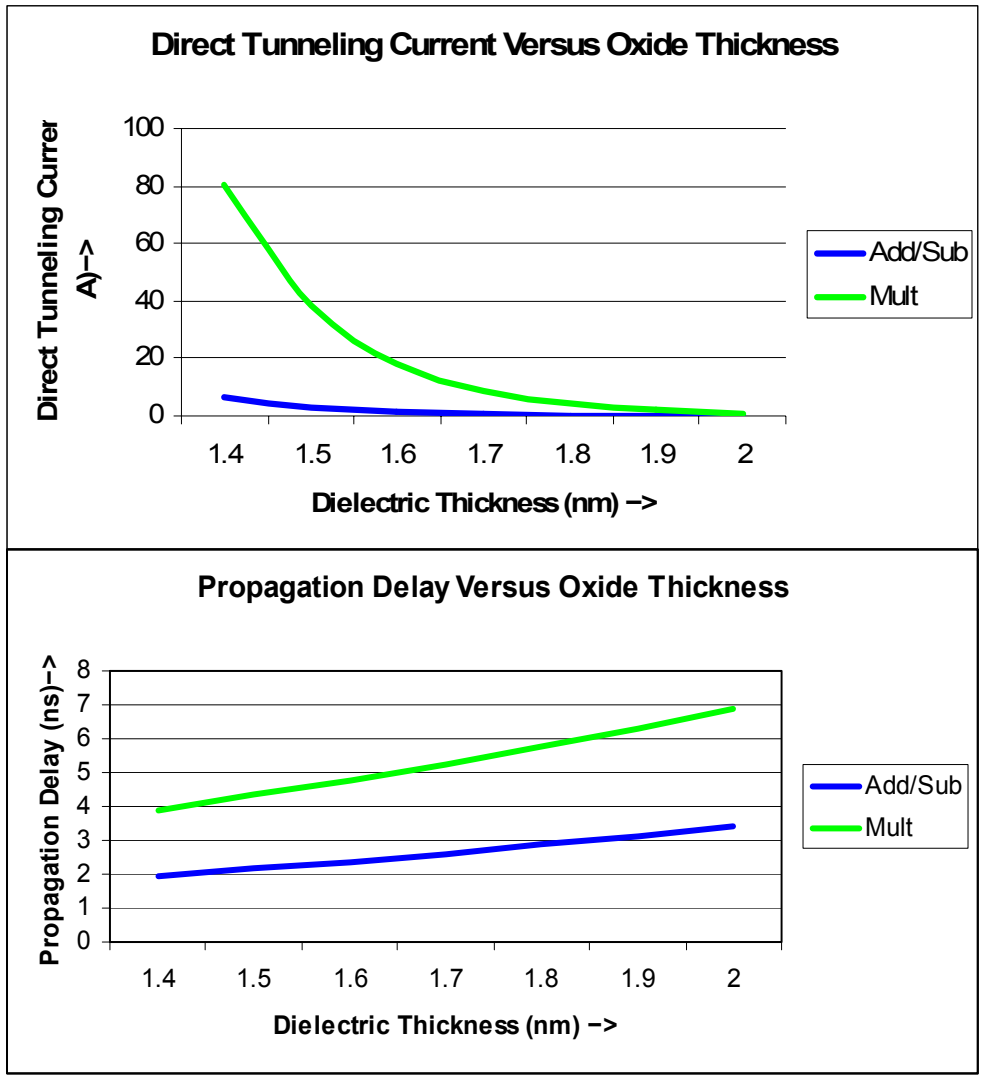
$$Tpd_{FU} = \sum_{I=(1 \rightarrow ncp)} 0.5(n_{fan\text{-}in} Tpd_{NMOS} + Tpd_{PMOS})$$

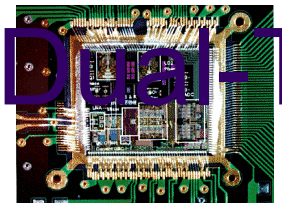
- $n_{fan\text{-}in}$ is the effective fan-in factor.
- Using physical-alpha-power model the delay of a MOS, where I_{DSat0} is the saturation drain current of the MOS for $V_{gs} = V_{dd}$.

$$T_{pd} = \frac{0.5 C_L V_{dd}}{I_{DSat\ 0}} + T_T \left\{ \frac{0.5 - \left(\frac{V_{dd} - V_{Th}}{V_{dd}}\right)}{\alpha + 1} \right\}$$

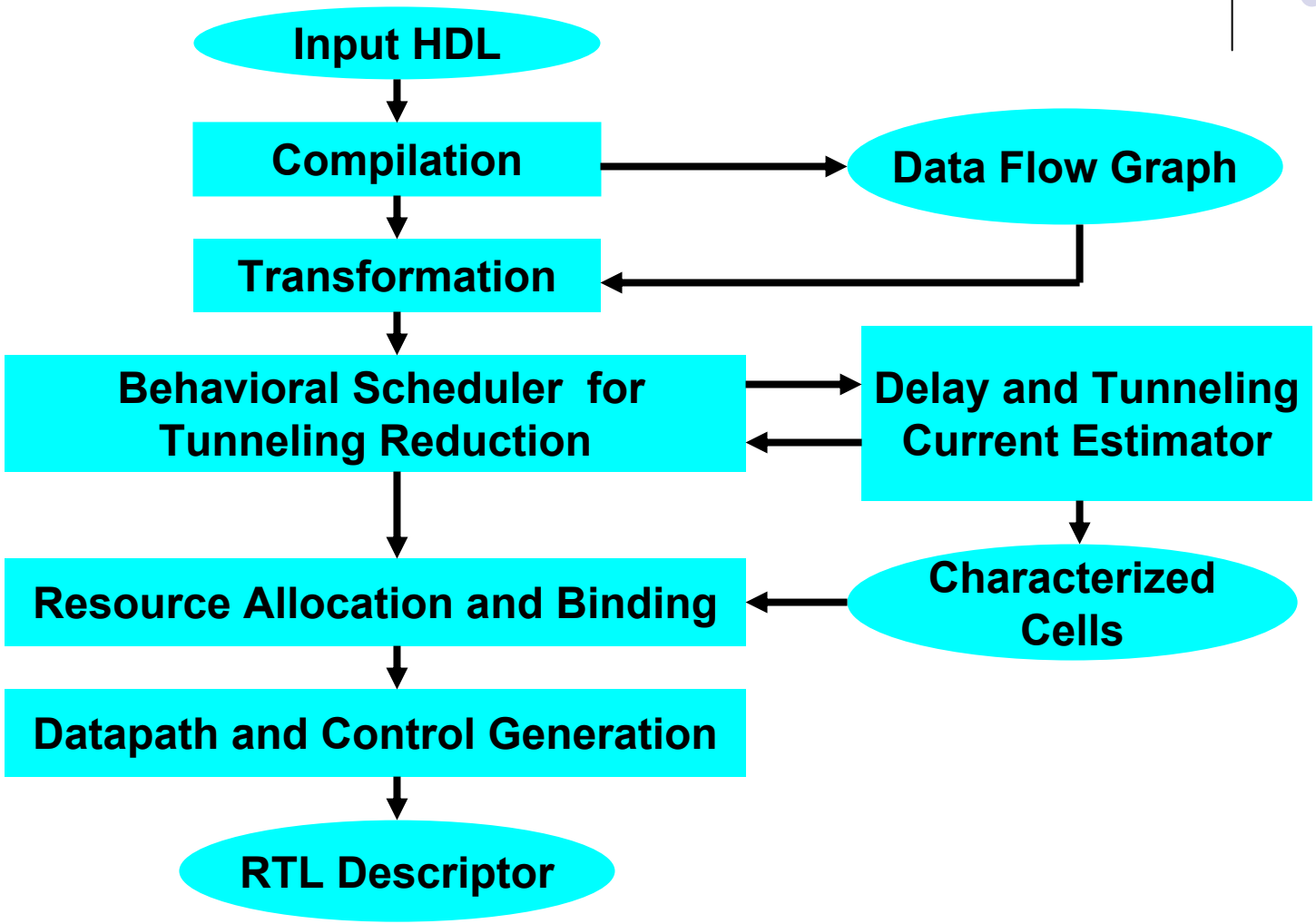


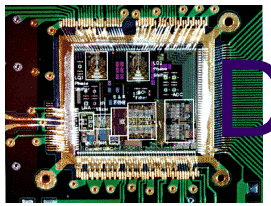
Characterization : 45nm Tech





Dual-T_{ox} Based Behavioral Synthesis

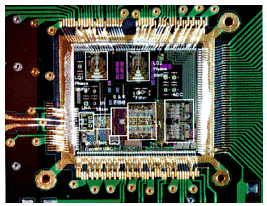




Dual-T_{OX} Assignment : Basis

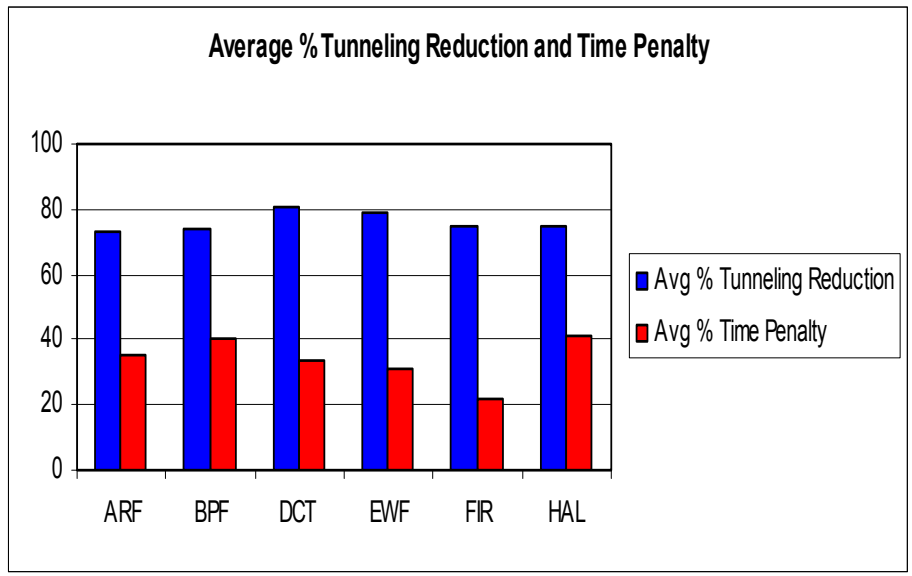
- **Observation**: Tunneling current of Functional Units increases and propagation delay decreases as oxide thickness deceases.

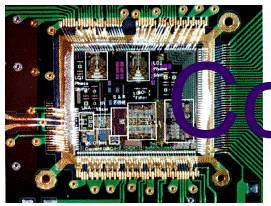
- **Strategy**: Maximize utilization of high-T_{OX} high leaky resources (e. g. multipliers) and low-T_{ox} low leaky resources (e.g. adder-subtractor) to improve chances of tunneling current reduction with minimal performance degradation.



Experimental Results

- The library consists of Multipliers and Adder-Subractor units, characterized for the 45 nm technology.
- We used T_1 = 1.4 nm, and T_2 = 1.7 nm to perform our experiments.
- The value of T_1 is chosen as the default value from the BSIM4.4.0 model card and value of T_2 is intuitively chosen.





Conclusions and Future Works

- Tunneling current is a major component of total power consumption of a low-end CMOS nanometer circuit.
- Dual-T_{OX} approach results significant reductions in tunneling current with minimal performance penalty.
- Development of optimal assignment algorithm is under progress.
- Tradeoff of tunneling, area and performance needs to be explored.
- Dual-T_{OX} based design may need more masks for the lithographic process during fabrication.