# Technology Mapping for Low Leakage Power and High Speed with Hot-Carrier Effect Consideration

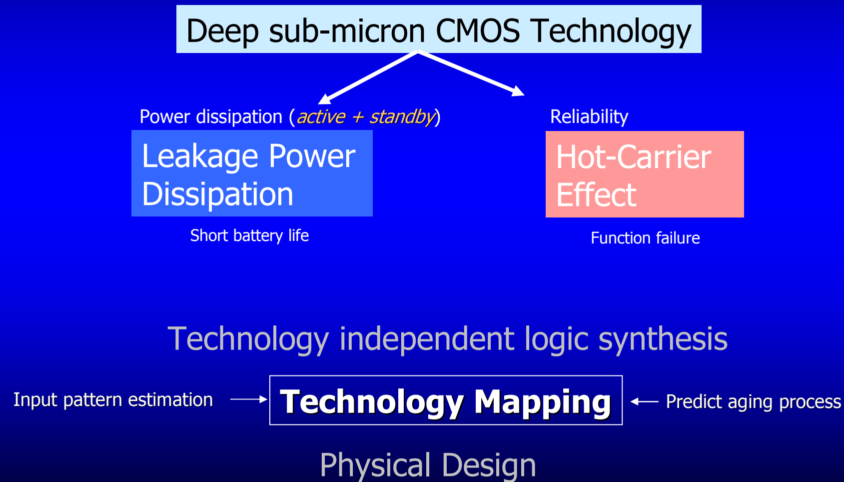
Chang Woo Kang and Massoud Pedram
University of Southern California/EE-systems
ckang@usc.edu

2D-3



## Outline

- Introduction
- Background
- Gate Modeling
- Technology Mapping
- High-Speed Heuristics
- Simulation Results
- Conclusion

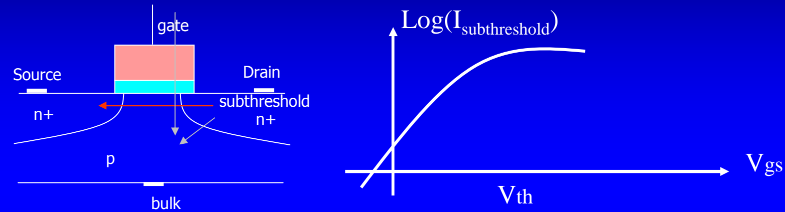



# Outline

- Introduction
- Background
- Gate Modeling
- Technology Mapping
- High-Speed Heuristics
- Simulation Results
- Conclusion

ASP-DAC 2003 4/36

# Subthreshold Current




$$I_{subthreshold} = A \cdot \exp\left(\frac{1}{kT/q}(V_{gs} - V_{th0} - \gamma \cdot V_{sb})\right)$$

- Becomes quite large as the threshold voltage and the channel length are reduced
- Increases exponentially as the threshold voltage is decreased linearly

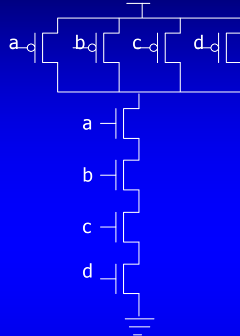


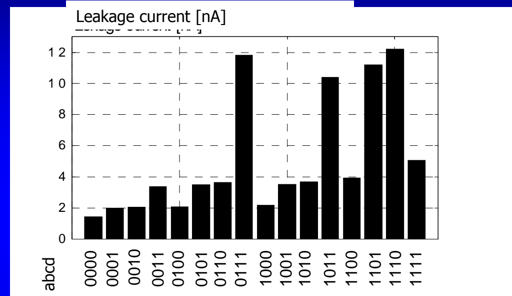
# Threshold Voltage Scaling

$$\tau = \frac{CV_{dd}}{(V_{dd} - V_{TH})^{\chi}}$$

- $V_{dd}/V_{th}$ = 3~4
  - Satisfy the DC noise margin
  - Maintain circuit performance
- Supply voltage is scaled down with each technology generation

## Input Vector Dependency

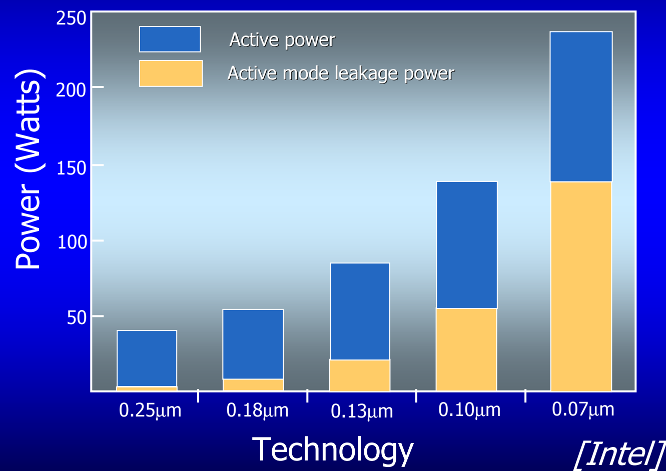



- The amount of subthreshold current is strongly dependent on the applied binary input pattern
- Techniques for leakage power reduction
  - Transistor stacking
  - Input vector assignment
  - Dual-threshold voltage devices

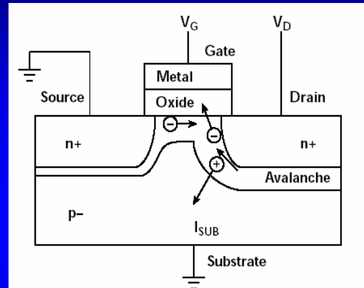


## Active Mode Leakage Power




*[Intel]*

# Hot-Carrier Effect




- Injection of high-energy electrons into the gate oxide near the drain region
- Can result in a substantial device parameter degradation
  - Threshold voltage
  - Transconductance
  - Linear and saturation drain current

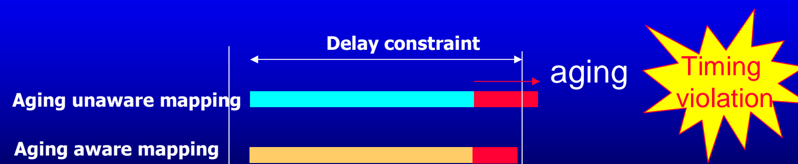


# Hot-Carrier Effects

- Change of critical timing paths due to transistor degradation
- Critical from the viewpoint of long term system reliability
- Long saturation region during transition causes device aging
- Therefore, devices are worn out quickly by
  - Slow slew rate of inputs
  - High load output capacitance
  - High switching activity

# Motivation

- Leakage power dissipation must be minimized during not only the standby mode but also the active mode
- Leakage power dissipation can be reduced during the logic synthesis phase
- Aging delay due to hot-carrier effect ought to be considered during the technology mapping step to improve the circuit reliability

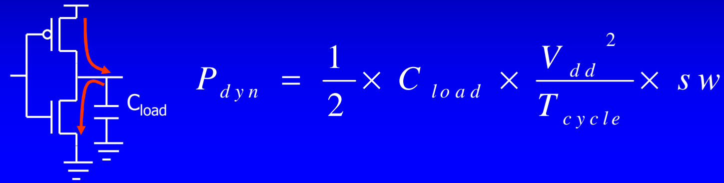





# Outline

- Introduction
- Background
- Gate Modeling
- Technology Mapping
- High-Speed Heuristics
- Simulation Results
- Conclusion

## Gate Modeling
# Dynamic Power

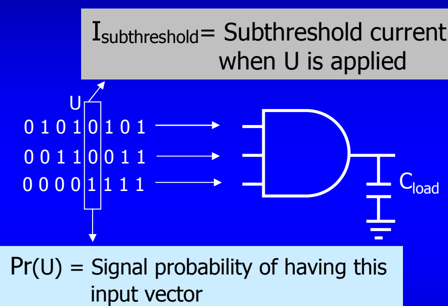



$$P_{dyn} = \frac{1}{2} \times C_{load} \times \frac{V_{dd}^{2}}{T_{cycle}} \times sw$$

- $C_{load}$ capacitive load
- $V_{dd}$ supply voltage
- $T_{cycle}$ clock cycle time
- sw switching activity



## Gate Modeling
# Leakage Power




- Input vector probability
- Lookup subthreshold current for input vectors

Gate Modeling

## Leakage Power

- Average leakage power dissipation
- 3-INPUT NAND gate

| input | subthreshold current | probability |
|---|---|---|
| [000] | 2.30E-08 | 0.1 |
| [001] | 2.96E-08 | 0.2 |
| [010] | 2.96E-08 | 0.4 |
| [011] | 6.66E-08 | 0.1 |
| [100] | 2.91E-08 | 0 |
| [101] | 6.09E-08 | 0.1 |
| [110] | 6.03E-08 | 0.05 |
| [111] | 7.92E-08 | 0.05 |

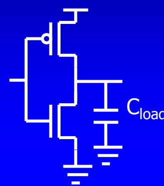
$$P_{leak} = V_{dd} \times \sum_{U} [I_{subthreshold}(U) \times pr(U)]$$



Gate Modeling

## Hot-carrier effect

$$T_{fresh} = \tau_{i,g} + R_{i,g} C_{load}$$




$$T_{aged} = T_{fresh} \times \left(1 + sw \times \alpha C_{load} \frac{\beta}{T_{avg.slew}}\right)$$

- $T_{aged}$ aged input-to-output delay
- $T_{fresh}$ fresh input-to-output delay
- $T_{avg.slew}$ average input slew rate
- $\alpha$ degradation factor due to the output load
- $\beta$ degradation factor due to the input slew rate
- sw switching probability

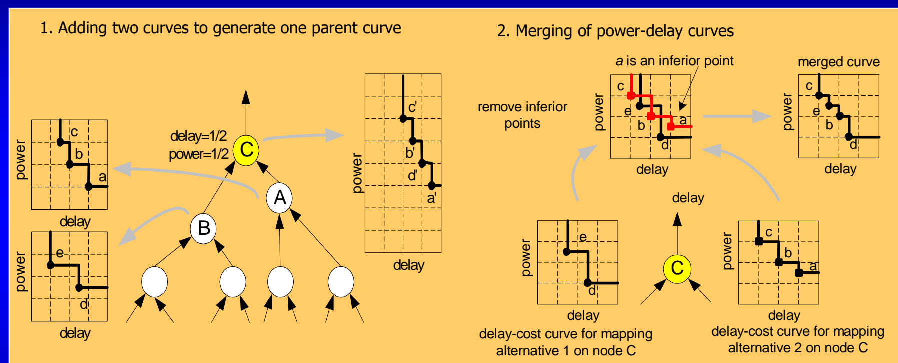
# Outline

- Introduction
- Background
- Gate Modeling
- Technology Mapping
- High-Speed Heuristics
- Simulation Results
- Conclusion



Review

# Technology Mapping




- Post-order traversal
  - Adding and lower-bound merging
- Pre-order traversal
  - Selecting the best point satisfying required time

## Leakage Power Aware Mapping

$$P(n,g)=\frac{1}{2}C_{diff}(n,g)\frac{V_{dd}^2}{T_{cycle}}sw_n+V_{dd}\sum_U I_{subth}(U)\,pr(U)$$
$$+\sum_{ni\in inputs(n,g)}\left(\frac{1}{2}C_{load}(ni)\frac{V_{dd}^2}{T_{cycle}}sw_{ni}+\frac{P(ni,gi)}{fanout(ni)}\right)$$

- Total power dissipation in a node *n* mapped by gate *g* =

*Dynamic power dissipation on diffusion capacitance*

*+ Leakage power dissipation*

*+ Dynamic power dissipation on gate capacitance*

*+ Dynamic power dissipation in transitive fanin/fanout*

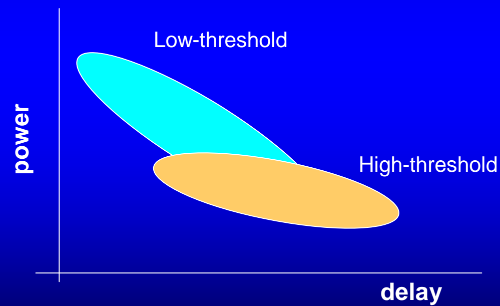


## Dual-Threshold Cell Mapping

- Two types of library cells
  - High-threshold cells on non-critical timing paths
    - Low leakage but slow
  - Low-threshold cells on critical timing path
    - High leakage but fast
- Current practice is to initially map to high-threshold gates and then selectively replace some gate with low-threshold counterparts from the library
  - Less powerful
- In our approach, we do the selection of high versus low threshold devices during the mapping step

# Dual-Threshold Cell Mapping

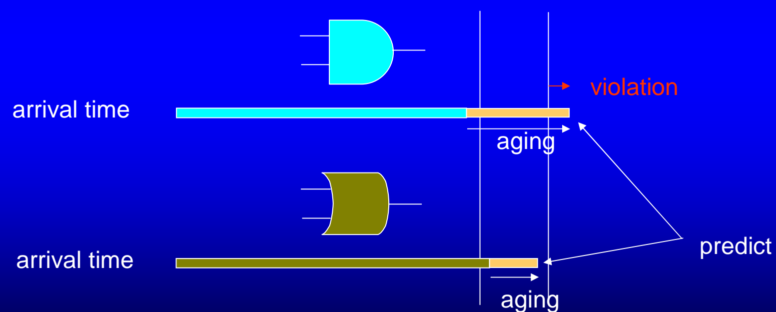
- Mapper naturally selects low-threshold cells for critical paths from the power-delay curve




ASP-DAC 2003 21/36

# Hot-carrier Effect Aware Mapping

- Predict the aging effect and apply it for arrival time computation




ASP-DAC 2003 22/36

## Hot-carrier Effect Aware Mapping

$$arrival(n, g, C_n) = \max_{ni \in inputs(n,g)} \left( T_{aged} + arrival(ni, gi, C_i) \right)$$

$$T_{aged} = T_{fresh} \times \left( 1 + sw \times \alpha C_{load} \frac{\beta}{T_{avg.slew}} \right)$$

- Arrival time is the maximum delay from primary inputs to the gate output
- Aged delay is considered as a gate delay

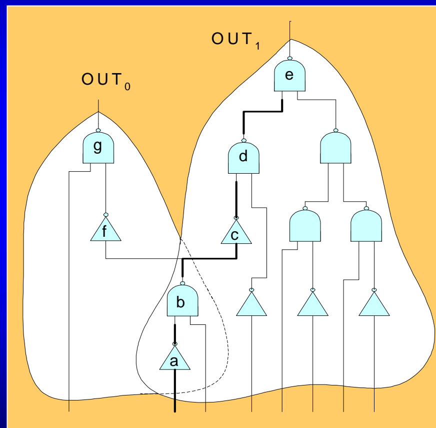


## Outline

- Introduction
- Background
- Gate Modeling
- Technology Mapping
- High-Speed Heuristics
- Simulation Results
- Conclusion

High-Speed Heuristics

## Primary Output Ordering




- Different order of mapping the logic cones generate different area, power dissipation, and speed
- Delay of a mapped cone is roughly proportional to the logic depth of the corresponding primitive cone
- Sort in descending order of their logic depths
- Simple heuristic, but effective



High-Speed Heuristics

## Pin Permutation

$$\overline{AB+C} \rightarrow \{A,B\}\{C\}$$

- Careful pin assignment for signals can result in reduced propagation delay through a CMOS gate
- Equivalent pins belong to the same set
- Pin permutation is performed during technology mapping

## Outline

- Introduction
- Background
- Gate Modeling
- Technology Mapping
- High-Speed Heuristics
- Simulation Results
- Conclusion

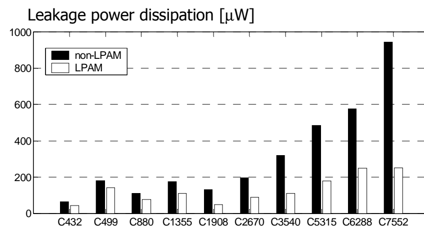


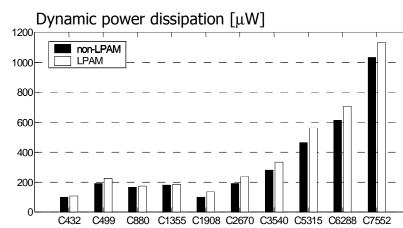
## Simulation Setup

- LP-SIS
- 0.18µm CMOS technology commercial library
- Dual-threshold library set
  - High-threshold library : 0.4V
  - Low-threshold library : 0.2V
  - Typically, ~10X leakage difference
- Calculation of signal probabilities using ordered binary decision diagram (OBDD)

Results

# Leakage Power Aware Mapping



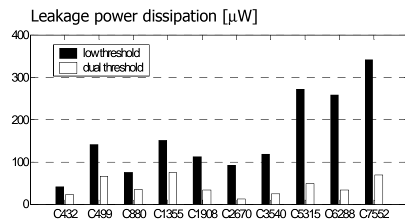



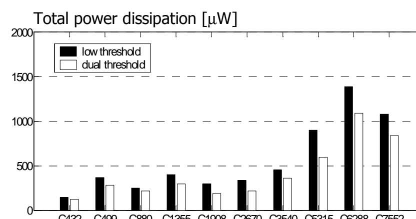

- Power dissipation reduction due to leakage power aware mapping
- Low-threshold library cells are used
- Dynamic power increases but leakage power decreases significantly



Results

# Power Dissipation

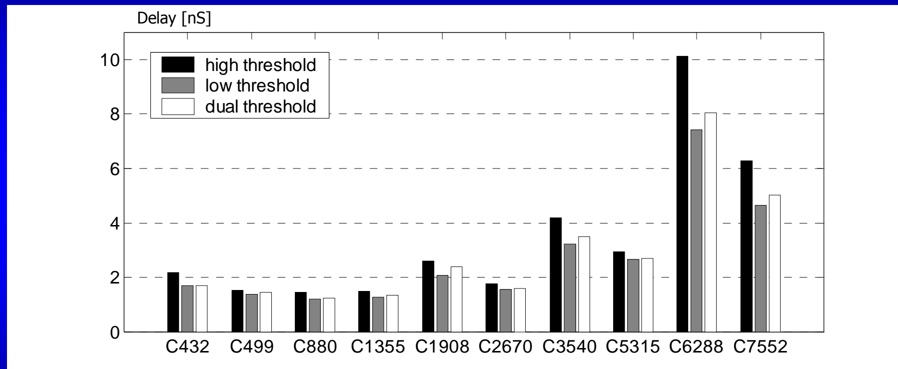



- Power dissipation reduction due to dual threshold gates
- High-threshold voltage cells on non-critical paths
- Low-threshold voltage cells on critical paths
- Low total power dissipation

Results

# Delay

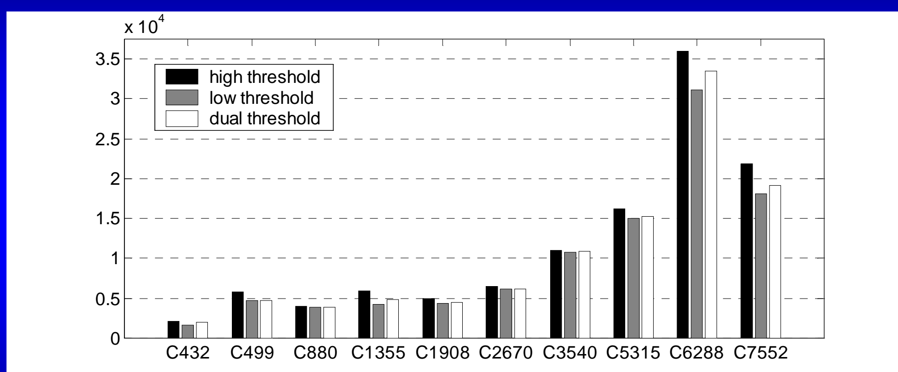



- Circuit speedup by mapping low-threshold voltage gates on critical paths



Results

# Area




- Gate area reduction by mapping low-threshold voltage gates on the critical paths

Results

# Aging-aware Mapping

| Circuit | Aging-unaware mapping Fresh [ns] | Aging-unaware mapping Aged [ns] | Aging-aware mapping Fresh [ns] | Aging-aware mapping Aged [ns] | Speed-up Fresh (%) | Speed-up Aged (%) |
|---|---|---|---|---|---|---|
| C432 | 2.21 | 2.54 | 2.08 | 2.34 | 6.3 | 7.9 |
| C499 | 1.52 | 1.74 | 1.5 | 1.66 | 1.3 | 4.6 |
| C880 | 1.39 | 1.41 | 1.33 | 1.34 | 4.5 | 5 |
| C1355 | 1.61 | 2.09 | 1.72 | 1.94 | -6.4 | 7.2 |
| C1908 | 2.65 | 2.82 | 2.45 | 2.52 | 8.2 | 10.6 |
| C2670 | 1.89 | 1.93 | 1.81 | 1.86 | 4.4 | 3.6 |
| count | 2.18 | 2.2 | 2.13 | 2.15 | 2.3 | 2.3 |
| sqrt8ml | 1.93 | 2.06 | 1.97 | 1.98 | -2 | 3.9 |
| f51m | 2.01 | 2.04 | 1.89 | 1.95 | 6.3 | 4.4 |
| alu2 | 2.1 | 2.14 | 2.03 | 2.05 | 3.4 | 4.2 |
| i2 | 0.49 | 0.5 | 0.45 | 0.46 | 8.9 | 8 |
| i7 | 1.13 | 1.14 | 1.1 | 1.11 | 2.7 | 2.6 |

- Aging-aware mapping enhances system reliability by limiting degraded delay during technology mapping



Results

# Primary Output Ordering

| Circuit | Without PO ordering Area ($\mu m^2$) | Without PO ordering Power ($\mu W$) | Without PO ordering Delay (ns) | With PO ordering Area ($\mu m^2$) | With PO ordering Power ($\mu W$) | With PO ordering Delay (ns) | Improvement (%) Area | Improvement (%) Power | Improvement (%) Delay |
|---|---|---|---|---|---|---|---|---|---|
| C432 | 2219 | 82 | 2.48 | 2128 | 78.59 | 2.36 | 4.3 | 4.2 | 4.9 |
| C499 | 3881 | 136.9 | 1.65 | 3857 | 136.3 | 1.62 | 0.6 | 0.4 | 1.8 |
| C880 | 4232 | 170.4 | 1.84 | 3986 | 155.7 | 1.67 | 6.2 | 8.6 | 9.2 |
| C1355 | 6390 | 348.5 | 2.04 | 5988 | 330.7 | 2 | 6.7 | 5.1 | 2 |
| z4ml | 385 | 14.5 | 0.84 | 365 | 14.1 | 0.79 | 5.5 | 3.2 | 6 |
| f51m | 1269 | 56 | 2.09 | 1255 | 54 | 2.05 | 1.1 | 3.6 | 1.9 |
| alu2 | 4283 | 131.1 | 2.93 | 4270 | 125.4 | 2.84 | 0.3 | 4.3 | 0.3 |
| i7 | 5271 | 142.4 | 1.51 | 5278 | 141.1 | 1.5 | -0.1 | 0.8 | 0.7 |

- In general, large deep cone can be mapped first to provide maximum mapping flexibility

Results

# Pin Permutation

| Circuit | Without PinPermutation Area ($\mu m^2$) | Power ($\mu W$) | Delay (ns) | With Pin Permutation Area ($\mu m^2$) | Power ($\mu W$) | Delay (ns) | Improvement (%) Area | Power | Delay |
|---|---|---|---|---|---|---|---|---|---|
| C432 | 3207 | 101.5 | 2.27 | 3331 | 103.5 | 2.04 | -3.9 | -2 | 10 |
| C499 | 3395 | 125.8 | 1.53 | 3380 | 126 | 1.49 | 0.4 | 0 | 2.6 |
| C880 | 2637 | 155.1 | 1.9 | 2865 | 162 | 1.7 | -8.6 | -4.5 | 10.5 |
| C1355 | 3428 | 370.8 | 1.61 | 3380 | 377.1 | 1.54 | 1.4 | -1.7 | 4.3 |
| z4ml | 402 | 45.9 | 0.57 | 336 | 37.8 | 0.52 | 16.4 | 17.7 | 8.7 |
| f51m | 1026 | 79.82 | 1.81 | 997 | 73.04 | 1.71 | 2.9 | 8.5 | 5.5 |
| alu2 | 2950 | 148.3 | 2.58 | 2958 | 147.1 | 2.43 | -0.3 | 0.8 | 5.8 |
| i2 | 1712 | 136.3 | 0.48 | 1603 | 125.8 | 0.45 | 6.3 | 7.7 | 6.3 |
| i7 | 4128 | 109.1 | 1.64 | 4134 | 107.9 | 1.43 | -0.2 | 1.1 | 12.8 |

- Delay is always reduced, but in some cases, area and power go up



# Conclusion

- A technology mapping technique considering leakage power dissipation was presented
- 52% reduction in leakage power dissipation and 27% decreases in total power dissipation was achieved
- An aging delay model capturing the hot-carrier effect was proposed and used during technology mapping
- 10.6% reduction in aged delay of the circuits was achieved
- Primary output ordering and pin permutation were effective in improving the circuit speed