Integrated Retiming and Simultaneous Vdd/Vth Scaling for Total Power Minimization

Mongkol Ekpanyapong

Advisor: Prof. Sung Kyu Lim

School of Electrical and Computer Engineering
Georgia Institute of Technology

Outline

- Introduction and Motivation
- Related Work
- Methodology
- Experimental Results
- Conclusions

Introduction

- Both static and dynamic power are the important issue in deep submicron design
- Performance is important issue
- The objective of this work is to minimize total power consumption while maintain the target clock period

Retiming Algorithm

- Linear Programming
  - Can easily be modified to handle any linear objective
- Bellman-Ford Algorithm
  - Can handle large circuits

Power Minimization

- Minimize total number of Flip-flop to reduce flip-flop power
- Using dual Vdd and Vth to minimize static and dynamic power

Outline

- Introduction and Motivation
- Related Work
- Methodology
- Experimental Results
- Conclusions

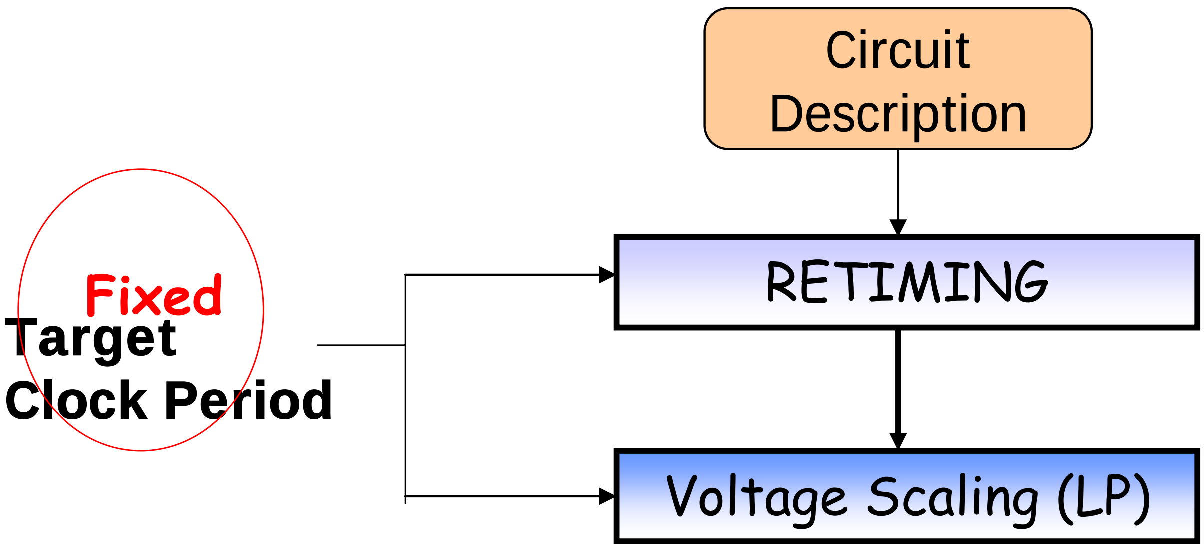
Retiming and Voltage Scaling

- C. E. Leiserson and J. B. Saxe, “Retiming synchronous circuitry,” *Algorithmica 1991*
- K. Usami and M. Horowitz, “Clustered Voltage Scaling Technique for Low-Power Design“ , *ISLPED 1995*
- N. Chabini and W. Wolf, “Reducing Dynamic Power Consumption in Synchronous Sequential Digital Designs Using Retiming and Supply Voltage Scaling,” *TVLSI 2004*

Outline

- Introduction and Motivation
- Related Work
- Methodology
- Experimental Results
- Conclusions

Power Minimization with Retiming



Retiming Formulation

Objective:

Minimize the number of flip-flops (FF.)

Constraints:

- Num. FF. has to be satisfied
  $r(u) \leq w(e_{u,v}) + r(v)$
- Num. FF. on critical paths has to be greater than zero

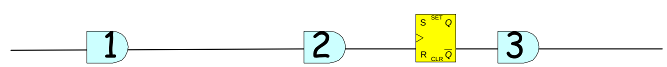
$$\text{Minimize } \{FI(v) - FO(v)\} \cdot r(v) \qquad (53)$$

Subject to:

$$r(u) - r(v) \leq w(e_{u,v}), \ \forall e_{u,v} \in E \qquad (54)$$

$$r(u) - r(v) \leq W(u,v) - 1, \ \forall D(u,v) > L, \ \forall u,v \in V \qquad (55)$$

w(e) 0 0 0 0



r(v) 0 0 0

V is the set of gates and E is the set of edges. $v \in V$ and $e \in E$

r(v) is the number of FF. moved from fanout of node v to fanin of node v

$w(e_{u,v})$ is the FF. count on edge u,v,

D(u,v) is the maximum delay on path u,v

W(u,v) is minimum number of FF. on path u,v

Retiming Formulation

Objective:
Minimize the number of flip-flops (FF.)

$$\text{Minimize } \{FI(v) - FO(v)\} \cdot r(v) \tag{53}$$

Subject to:

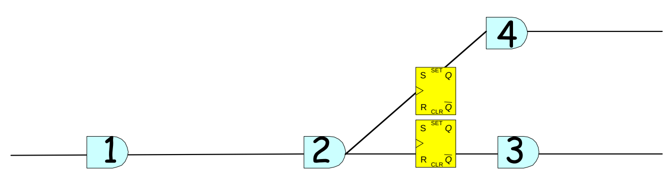
$$r(u) - r(v) \leq w(e_{u,v}),\ \forall e_{u,v} \in E \tag{54}$$

$$r(u) - r(v) \leq W(u,v) - 1,\ \forall D(u,v) > L,\ \forall u,v \in V \tag{55}$$

Constraints:

- Num. FF. has to be satisfied
  $r(u) \leq w(e_{u,v}) + r(v)$
- Num. FF. on critical paths has to be greater than zero

0



r(v) 0 0 0

V is the set of gates and E is the set of edges. $v \in V$ and $e \in E$
r(v) is the number of FF. moved from fanout of node v to fanin of node v
$w(e_{u,v})$ is the FF. count on edge u,v,
D(u,v) is the maximum delay on path u,v
W(u,v) is minimum number of FF. on path u,v

Retiming Formulation

Objective:

Minimize the number of flip-flops (FF.)

Constraints:

- Num. FF. has to be satisfied
  $r(u) \le w(e_{u,v}) + r(v)$
- Num. FF. on critical paths has to be greater than zero

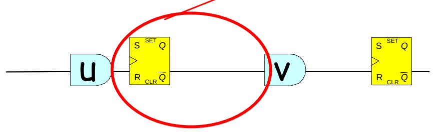
$$\text{Minimize } \{FI(v) - FO(v)\} \cdot r(v) \quad (53)$$

Subject to:

$$r(u) - r(v) \le w(e_{u,v}),\ \forall e_{u,v} \in E \quad (54)$$

$$r(u) - r(v) \le W(u,v) - 1,\ \forall D(u,v) > L,\ \forall u,v \in V \quad (55)$$

Only these 2 FF. can move out of u



V is the set of gates and E is the set of edges. $v \in V$ and $e \in E$

r(v) is the number of FF. moved from fanout of node v to fanin of node v

$w(e_{u,v})$ is the FF. count on edge u,v,

D(u,v) is the maximum delay on path u,v

W(u,v) is minimum number of FF. on path u,v

Retiming Formulation

Objective:
Minimize the number of flip-flops (FF.)

Constraints:
- Num. FF. has to be satisfied
  $r(u) \leq w(e_{u,v}) + r(v)$
- Num. FF. on critical paths has to be greater than zero

$$\text{Minimize } \{FI(v) - FO(v)\} \cdot r(v) \quad (53)$$

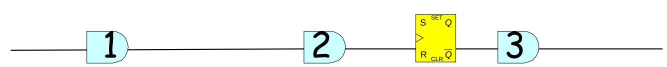
Subject to:

$$r(u) - r(v) \leq w(e_{u,v}), \ \forall e_{u,v} \in E \quad (54)$$

$$r(u) - r(v) \leq W(u,v) - 1, \ \forall D(u,v) > L, \ \forall u,v \in V \quad (55)$$

$r(1)-r(3) \leq 0 \Rightarrow r(1) \leq r(3)$

Cycle Time (L) =2



D(1,2) = 2
D(1,3) = 3
D(2,3) = 2

W(1,2) = 0
W(1,3) = 1
W(2,3) = 1

V is the set of gates and E is the set of edges. $v \in V$ and $e \in E$
r(v) is the number of FF. moved from fanout of node v to fanin of node v
$w(e_{u,v})$ is the FF. count on edge u,v,
D(u,v) is the maximum delay on path u,v
W(u,v) is minimum number of FF. on path u,v

Retiming Formulation

Objective:

Minimize the number of flip-flops (FF.)

$$\text{Minimize } \{FI(v) - FO(v)\} \cdot r(v) \qquad (53)$$

Subject to:

$$r(u) - r(v) \le w(e_{u,v}),\ \forall e_{u,v} \in E \qquad (54)$$

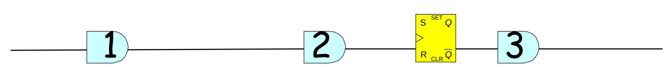
$$r(u) - r(v) \le W(u,v) - 1,\ \forall D(u,v) > L,\ \forall u,v \in V \qquad (55)$$

Constraints:

- Num. FF. has to be satisfied
  $r(u) \le w(e_{u,v}) + r(v)$
- Num. FF. on critical paths has to be greater than zero

$r(1)-r(3) \le 0 \Rightarrow r(1) \le r(3)$

Cycle Time (L) =2

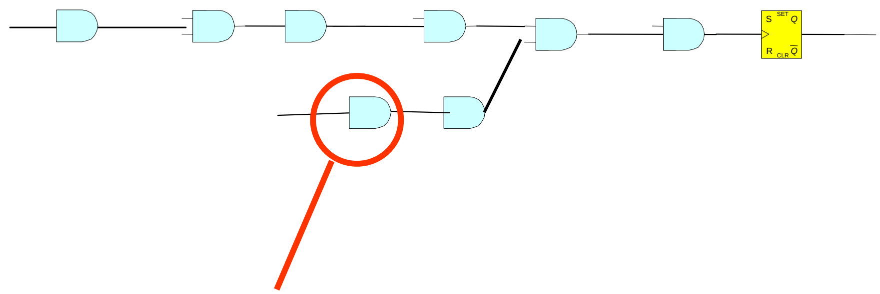


D(1,2) = 2
D(1,3) = 3
D(2,3) = 2

W(1,2) = 0
W(1,3) = 1
W(2,3) = 1

V is the set of gates and E is the set of edges. $v \in V$ and $e \in E$
r(v) is the number of FF. moved from fanout of node v to fanin of node v
$w(e_{u,v})$ is the FF. count on edge u,v,
D(u,v) is the maximum delay on path u,v
W(u,v) is minimum number of FF. on path u,v

Non-critical Gates for Power Minimization

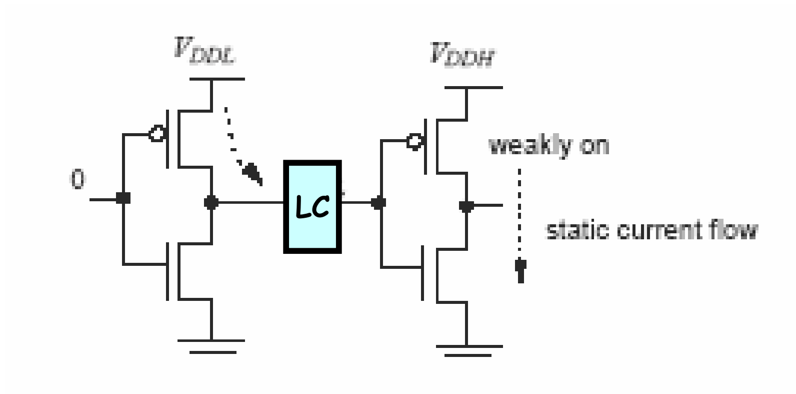


Non-critical gates:
What should we do?

We can use the voltage scaling for non-critical gates after retiming to minimize total power consumption

Low-to-High V_{dd} Conversion

- Level Converter (LC) requirement



Voltage Scaling Formulation

Objective:

Minimize gate power + LC power

Constraints:

Each gate has to be assigned to only one voltage state

Arrival time + gate delay of each node ≤ target clock period

Level converter inserted if low V_{dd} node drives high V_{dd} node

MILP Formulation

$$\text{Minimize } \left(\sum_{v\in V}\sum_{k=1}^{4} p_{v,k}\cdot x_{v,k}\right) + \left(\sum_{e\in E} p_{lc}\cdot m_e\right) \tag{58}$$

Subject to:

$$\sum_{k=1}^{4} x_{v,k} = 1,\ \forall v \in V \tag{59}$$

Timing constraints:

$$\sum_{k=1}^{4} d_{v,k}\cdot x_{v,k} + s(v) \le L,\ \forall v \in V \tag{60}$$

$$\sum_{k=1}^{4} d_{u,k}\cdot x_{u,k} + d_{lc}\cdot m(e) + s(u) \le s(v),\ \forall e_{u,v} \in E \tag{61}$$

$$s(v) \ge 0,\ \forall v \in V \tag{62}$$

Level converter (LC) constraints:

$$\sum_{i=1}^{4} z_{u,i}\cdot x_{u,i} - \sum_{j=1}^{4} z_{v,j}\cdot x_{v,j} + D\cdot m(e) \ge 0,\ \forall e \in E \tag{63}$$

Integer constraints:

$$x_{v,k} \in \{0,1\},\ \forall v \in V \tag{64}$$

$$m(e) \in \{0,1\},\ \forall e \in E \tag{65}$$

Voltage Scaling Formulation

MILP Formulation

$$\text{Minimize } \left(\sum_{v \in V} \sum_{k=1}^{4} p_{v,k} \cdot x_{v,k}\right) + \left(\sum_{e \in E} p_{lc} \cdot m_e\right) \tag{58}$$

Subject to:

$$\sum_{k=1}^{4} x_{v,k} = 1, \ \forall v \in V \tag{59}$$

Timing constraints:

$$\sum_{k=1}^{4} d_{v,k} \cdot x_{v,k} + s(v) \leq L, \ \forall v \in V \tag{60}$$

$$\sum_{k=1}^{4} d_{u,k} \cdot x_{u,k} + d_{lc} \cdot m(e) + s(u) \leq s(v), \ \forall e_{u,v} \in E \tag{61}$$

$$s(v) \geq 0, \ \forall v \in V \tag{62}$$

Level converter (LC) constraints:

$$\sum_{i=1}^{4} z_{u,i} \cdot x_{u,i} - \sum_{j=1}^{4} z_{v,j} \cdot x_{v,j} + D \cdot m(e) \geq 0, \ \forall e \in E \tag{63}$$

Integer constraints:

$$x_{v,k} \in \{0,1\}, \ \forall v \in V \tag{64}$$

$$m(e) \in \{0,1\}, \ \forall e \in E \tag{65}$$

v

- v — V_{dd} High V_{th} Low ($x_{v,4}$=1)
- v — V_{dd} High V_{th} High ($x_{v,3}$=1)
- v — V_{dd} Low V_{th} Low ($x_{v,2}$=1)
- v — V_{dd} Low V_{th} High ($x_{v,1}$=1)

Voltage Scaling Formulation

MILP Formulation

$$\text{Minimize } \left(\sum_{v \in V} \sum_{k=1}^{4} p_{v,k} \cdot x_{v,k}\right) + \left(\sum_{e \in E} p_{lc} \cdot m_e\right) \quad (58)$$

Subject to:

$$\sum_{k=1}^{4} x_{v,k} = 1, \ \forall v \in V \quad (59)$$

Timing constraints:

$$\sum_{k=1}^{4} d_{v,k} \cdot x_{v,k} + s(v) \le L, \ \forall v \in V \quad (60)$$

$$\sum_{k=1}^{4} d_{u,k} \cdot x_{u,k} + d_{lc} \cdot m(e) + s(u) \le s(v), \ \forall e_{u,v} \in E \quad (61)$$

$$s(v) \ge 0, \ \forall v \in V \quad (62)$$

Level converter (LC) constraints:

$$\sum_{i=1}^{4} z_{u,i} \cdot x_{u,i} - \sum_{j=1}^{4} z_{v,j} \cdot x_{v,j} + D \cdot m(e) \ge 0, \ \forall e \in E \quad (63)$$

Integer constraints:

$$x_{v,k} \in \{0,1\}, \ \forall v \in V \quad (64)$$

$$m(e) \in \{0,1\}, \ \forall e \in E \quad (65)$$

s(u) = 0
d(u) = 1

s(v) = 1

Voltage Scaling Formulation

MILP Formulation

$$\text{Minimize } \left(\sum_{v \in V} \sum_{k=1}^{4} p_{v,k} \cdot x_{v,k}\right) + \left(\sum_{e \in E} p_{lc} \cdot m_e\right) \tag{58}$$

Subject to:

$$\sum_{k=1}^{4} x_{v,k} = 1, \ \forall v \in V \tag{59}$$

Timing constraints:

$$\sum_{k=1}^{4} d_{v,k} \cdot x_{v,k} + s(v) \leq L, \ \forall v \in V \tag{60}$$

$$\sum_{k=1}^{4} d_{u,k} \cdot x_{u,k} + d_{lc} \cdot m(e) + s(u) \leq s(v), \ \forall e_{u,v} \in E \tag{61}$$

$$s(v) \geq 0, \ \forall v \in V \tag{62}$$

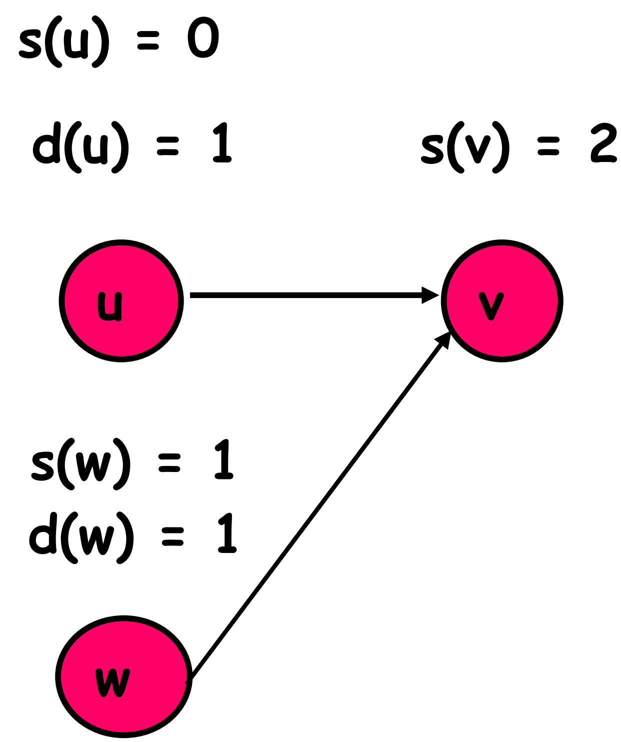
Level converter (LC) constraints:

$$\sum_{i=1}^{4} z_{u,i} \cdot x_{u,i} - \sum_{j=1}^{4} z_{v,j} \cdot x_{v,j} + D \cdot m(e) \geq 0, \ \forall e \in E \tag{63}$$

Integer constraints:

$$x_{v,k} \in \{0,1\}, \ \forall v \in V \tag{64}$$

$$m(e) \in \{0,1\}, \ \forall e \in E \tag{65}$$



Voltage Scaling Formulation

MILP Formulation

$$\text{Minimize } \left(\sum_{v\in V}\sum_{k=1}^{4} p_{v,k}\cdot x_{v,k}\right) + \left(\sum_{e\in E} p_{lc}\cdot m_e\right) \tag{58}$$

Subject to:

$$\sum_{k=1}^{4} x_{v,k} = 1,\ \forall v \in V \tag{59}$$

Timing constraints:

$$\sum_{k=1}^{4} d_{v,k}\cdot x_{v,k} + s(v) \le L,\ \forall v\in V \tag{60}$$

$$\sum_{k=1}^{4} d_{u,k}\cdot x_{u,k} + d_{lc}\cdot m(e) + s(u) \le s(v),\ \forall e_{u,v}\in E \tag{61}$$

$$s(v) \ge 0,\ \forall v \in V \tag{62}$$

Level converter (LC) constraints:

$$\sum_{i=1}^{4} z_{u,i}\cdot x_{u,i} - \sum_{j=1}^{4} z_{v,j}\cdot x_{v,j} + D\cdot m(e) \ge 0,\ \forall e \in E \tag{63}$$

Integer constraints:

$$x_{v,k} \in \{0,1\},\ \forall v \in V \tag{64}$$

$$m(e) \in \{0,1\},\ \forall e \in E \tag{65}$$

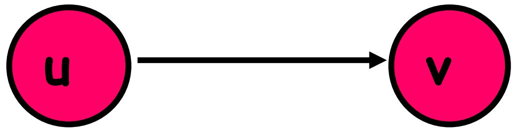
Cycle time (L) = 2

s(u) = 0

d(u) = 1

s(v) = 1

d(v) = 1



$s(u) + d(u) \le 2$

$s(v) + d(v) \le 2$

Voltage Scaling Formulation

MILP Formulation

$$\text{Minimize } \left(\sum_{v \in V} \sum_{k=1}^{4} p_{v,k} \cdot x_{v,k}\right) + \left(\sum_{e \in E} p_{lc} \cdot m_e\right) \tag{58}$$

Subject to:

$$\sum_{k=1}^{4} x_{v,k} = 1, \ \forall v \in V \tag{59}$$

Timing constraints:

$$\sum_{k=1}^{4} d_{v,k} \cdot x_{v,k} + s(v) \leq L, \ \forall v \in V \tag{60}$$

$$\sum_{k=1}^{4} d_{u,k} \cdot x_{u,k} + d_{lc} \cdot m(e) + s(u) \leq s(v), \ \forall e_{u,v} \in E \tag{61}$$

$$s(v) \geq 0, \ \forall v \in V \tag{62}$$

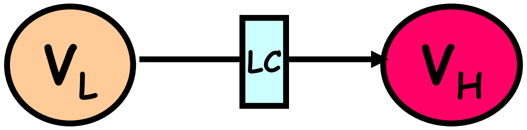
Level converter (LC) constraints:

$$\sum_{i=1}^{4} z_{u,i} \cdot x_{u,i} - \sum_{j=1}^{4} z_{v,j} \cdot x_{v,j} + D \cdot m(e) \geq 0, \ \forall e \in E \tag{63}$$

Integer constraints:

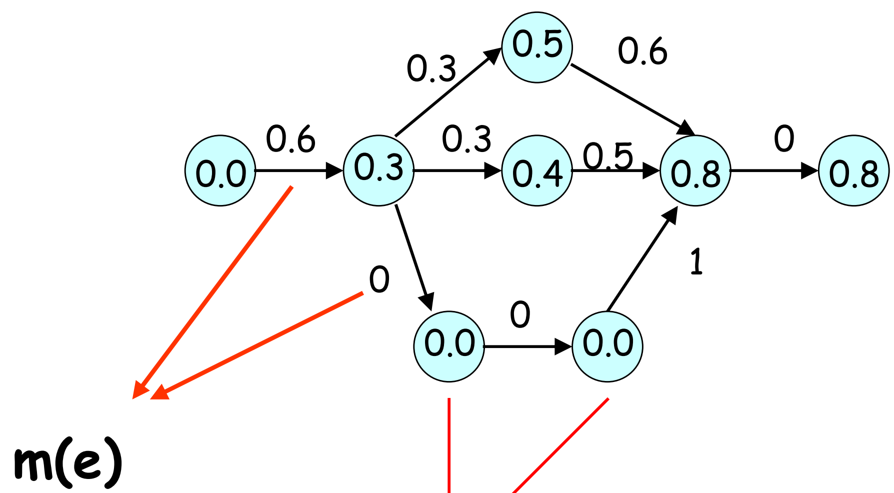
$$x_{v,k} \in \{0, 1\}, \ \forall v \in V \tag{64}$$

$$m(e) \in \{0, 1\}, \ \forall e \in E \tag{65}$$



m(e) = 1

Convert from ILP to LP



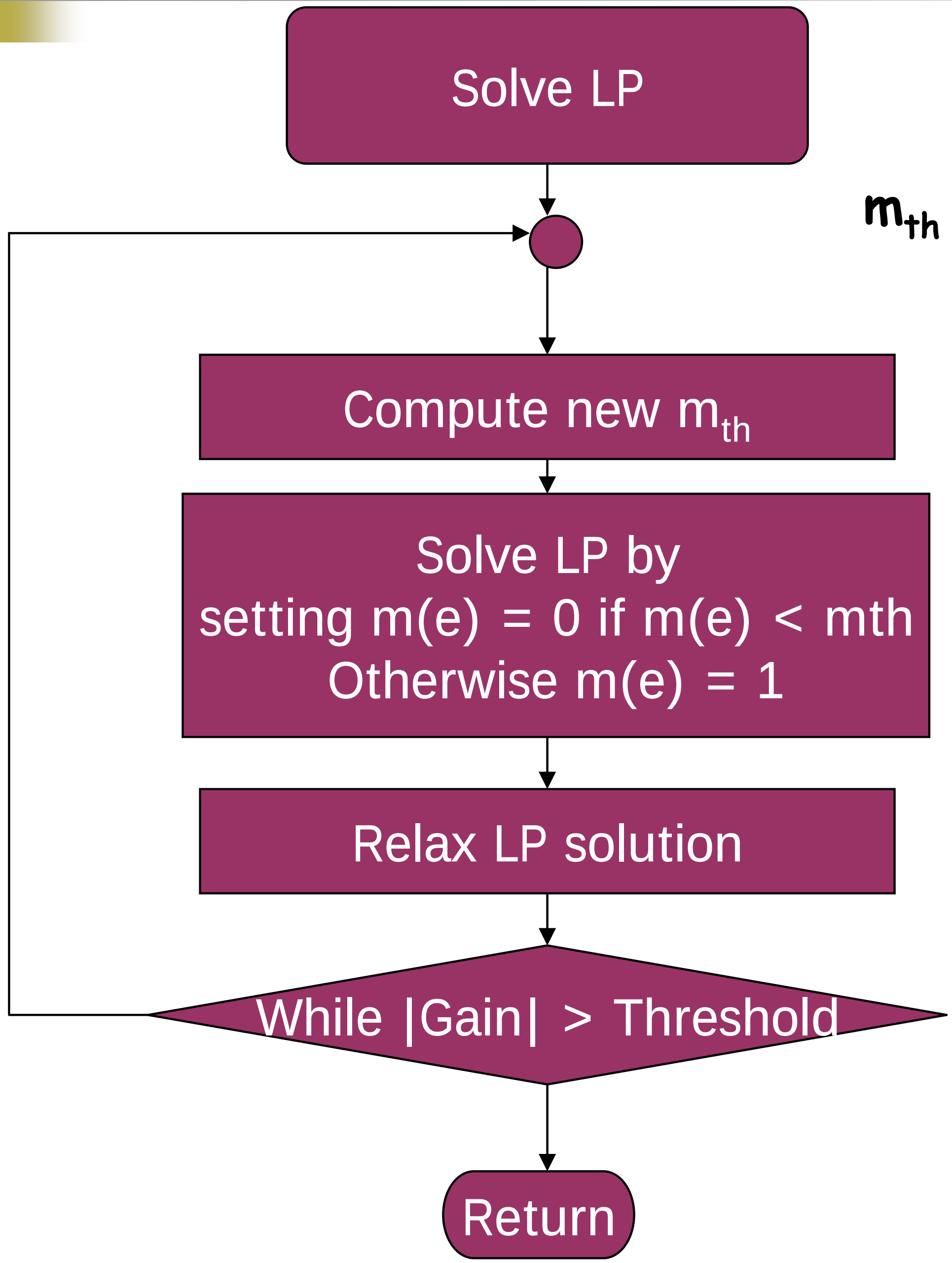
$$x_{u,2} = x(u) \begin{cases} 0 = \text{low Vdd} \\ 1 = \text{high Vdd} \end{cases}$$

V_L $x_{u,1}=1$ $x_{u,2}=0$

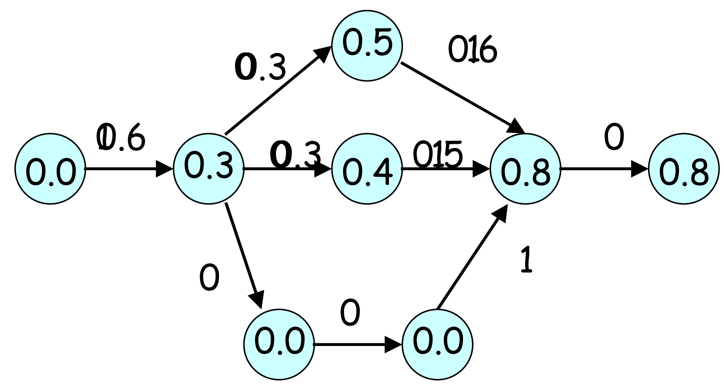
V_H $x_{u,1}=0$ $x_{u,2}=1$

Assume only two states for illustration purpose

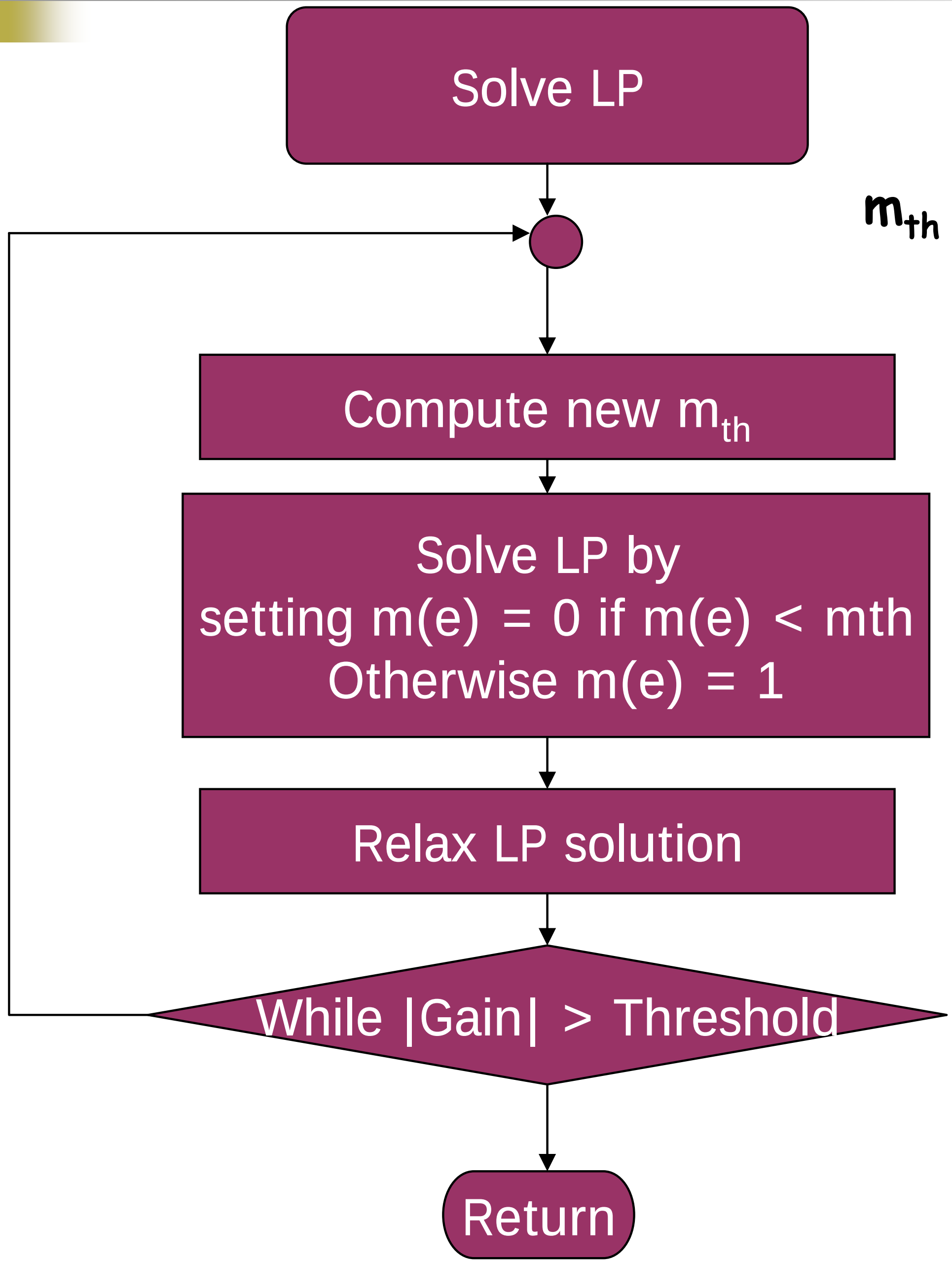
Gradient Search Algorithm for LC Relaxation



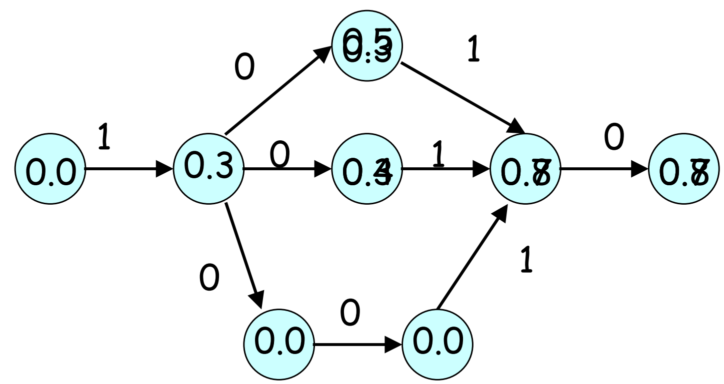
$m_{th} = 0.5$



Gradient Search Algorithm for LC Relaxation



$m_{th} = 0.5$



Voltage Assignment Relaxation

Voltage Assignment

Four possible voltage assignment:

- High V_{dd}, low V_{th} node

  Fastest gate, high dynamic power, high leakage power

- High V_{dd}, high V_{th} node

  High dynamic power, low leakage power

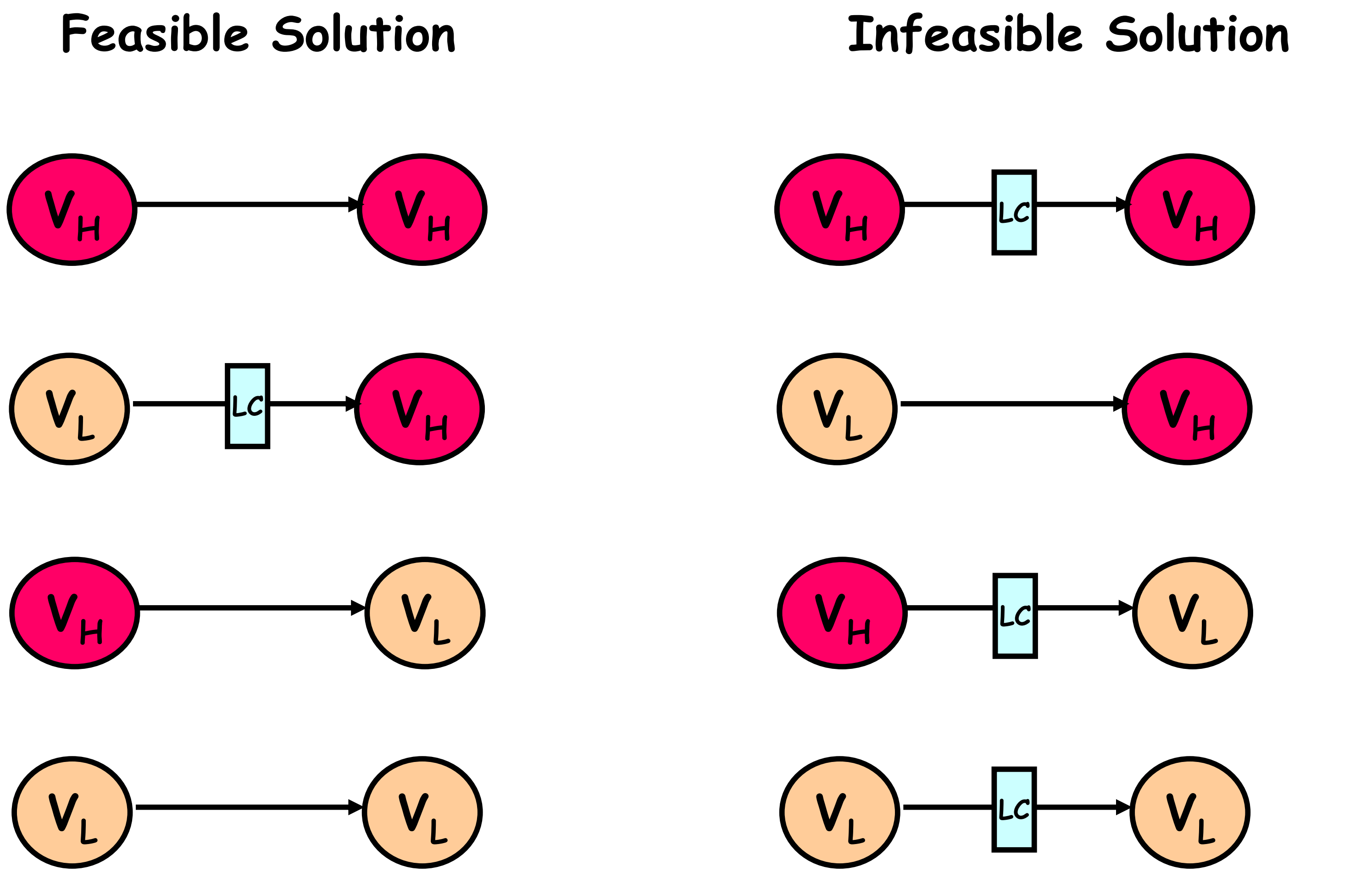
- Low V_{dd}, low V_{th} node

  Low dynamic power, high leakage power

- Low V_{dd}, high V_{th} node

  Slowest gate, low dynamic power, low leakage power

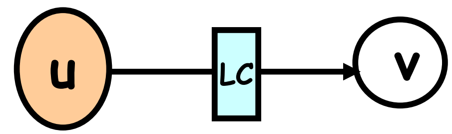
Possible Supply Voltage Assignment

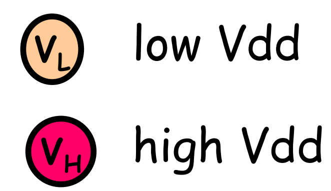


LP Relaxation for Voltage State Assignment

Voltage Mapping Algorithm
input: LP-based voltage scaling with LC inserted
output: ILP-based voltage scaling with reduced LC set

T = topological ordering of gates;
assign low-V_{dd}+high-V_{th} to all PIs;
while (T is not empty)
  $v = T$.pop;
  $dly(v) = \sum_{k=1}^{4} x_{v,k} \cdot d_{v,k}$;
  $vdd(v) = x_{v,1} + x_{v,3}$;
  $v \leftarrow V_{dd}$-L+V_{th}-H;
// V_{dd} mapping
  if ($\exists u \in FI(v) | m(e_{u,v}) = 1$)
    $v \leftarrow V_{dd}$-H;
  if ($vdd(v) > 0$)
    $v \leftarrow V_{dd}$-H;
// LC removal
  if ($\exists u \in FI(v) | u = V_{dd}$-H & $m(e_{u,v}) = 1$ or
  $u = V_{dd}$-L & $m(e_{u,v}) = 1$ and $v = V_{dd}$-L)
    $m(e_{u,v}) \leftarrow 0$
// V_{th} mapping
  if ($v = V_{dd}$-H & $dly(v) <$ delay(V_{dd}-H+V_{th}-H))
    $v \leftarrow V_{th}$-L;
  if ($v = V_{dd}$-L & $dly(v) <$ delay(V_{dd}-L+V_{th}-H))
    $v \leftarrow V_{th}$-L;

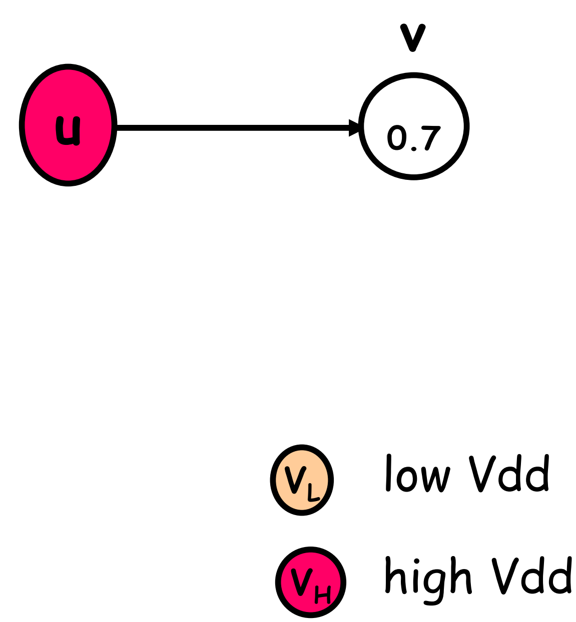




LP Relaxation for Voltage State Assignment

Voltage Mapping Algorithm
input: LP-based voltage scaling with LC inserted
output: ILP-based voltage scaling with reduced LC set

T = topological ordering of gates;
assign low-V_{dd}+high-V_{th} to all PIs;
while (T is not empty)
    $v = T$.pop;
    $dly(v) = \sum_{k=1}^{4} x_{v,k} \cdot d_{v,k}$;
    $vdd(v) = x_{v,1} + x_{v,3}$;
    $v \leftarrow V_{dd}$-L+V_{th}-H;
// V_{dd} mapping
    if ($\exists u \in FI(v) | m(e_{u,v}) = 1$)
        $v \leftarrow V_{dd}$-H;
    if ($vdd(v) > 0$)
        $v \leftarrow V_{dd}$-H;
// LC removal
    if ($\exists u \in FI(v) | u = V_{dd}$-H & $m(e_{u,v}) = 1$ or
    $u = V_{dd}$-L & $m(e_{u,v}) = 1$ and $v = V_{dd}$-L)
        $m(e_{u,v}) \leftarrow 0$
// V_{th} mapping
    if ($v = V_{dd}$-H & $dly(v) <$ delay(V_{dd}-H+V_{th}-H))
        $v \leftarrow V_{th}$-L;
    if ($v = V_{dd}$-L & $dly(v) <$ delay(V_{dd}-L+V_{th}-H))
        $v \leftarrow V_{th}$-L;



LP Relaxation for Voltage State Assignment

Voltage Mapping Algorithm
input: LP-based voltage scaling with LC inserted
output: ILP-based voltage scaling with reduced LC set

1. T = topological ordering of gates;
2. assign low-V_{dd}+high-V_{th} to all PIs;
3. while (T is not empty)
4. $\quad v = T$.pop;
5. $\quad dly(v) = \sum_{k=1}^{4} x_{v,k} \cdot d_{v,k}$;
6. $\quad vdd(v) = x_{v,1} + x_{v,3}$;
7. $\quad v \leftarrow V_{dd}$-L+$V_{th}$-H;

// V_{dd} mapping

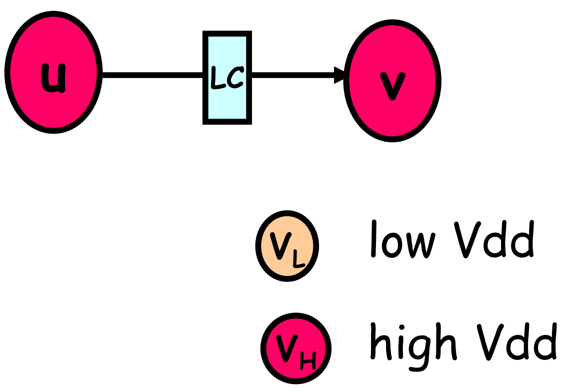
8. $\quad$ if $(\exists u \in FI(v) | m(e_{u,v}) = 1)$
9. $\quad\quad v \leftarrow V_{dd}$-H;
10. $\quad$ if $(vdd(v) > 0)$
11. $\quad\quad v \leftarrow V_{dd}$-H;

// LC removal

12. $\quad$ if ($\exists u \in FI(v) | u = V_{dd}$-H & $m(e_{u,v}) = 1$ or $u = V_{dd}$-L & $m(e_{u,v}) = 1$ and $v = V_{dd}$-L)
13. $\quad\quad m(e_{u,v}) \leftarrow 0$

// V_{th} mapping

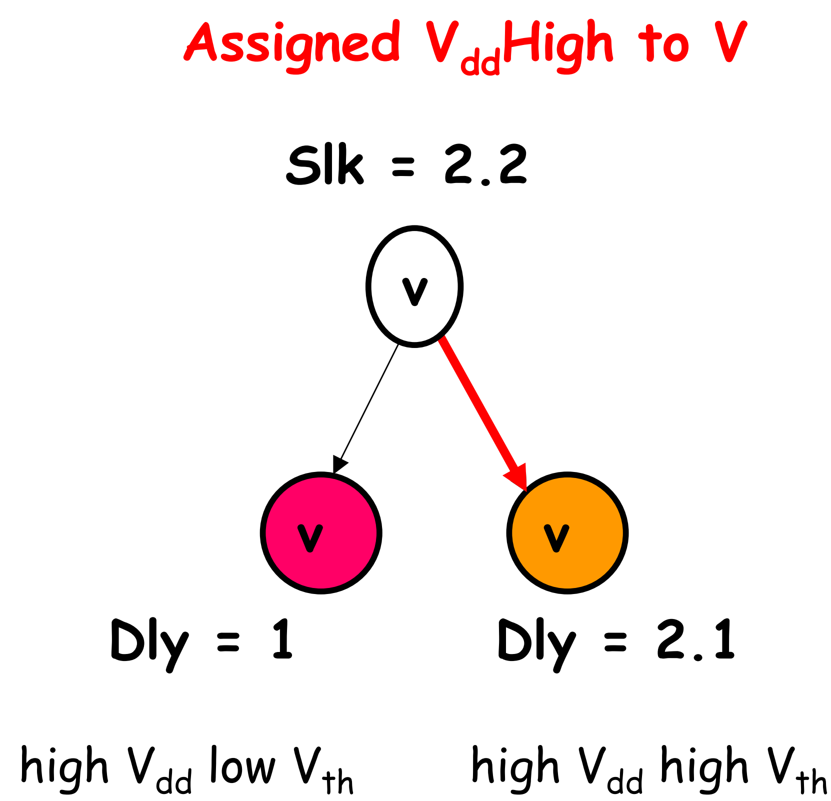
14. $\quad$ if ($v = V_{dd}$-H & $dly(v) <$ delay(V_{dd}-H+V_{th}-H))
15. $\quad\quad v \leftarrow V_{th}$-L;
16. $\quad$ if ($v = V_{dd}$-L & $dly(v) <$ delay(V_{dd}-L+V_{th}-H))
17. $\quad\quad v \leftarrow V_{th}$-L;



LP Relaxation for Voltage State Assignment

Voltage Mapping Algorithm
input: LP-based voltage scaling with LC inserted
output: ILP-based voltage scaling with reduced LC set

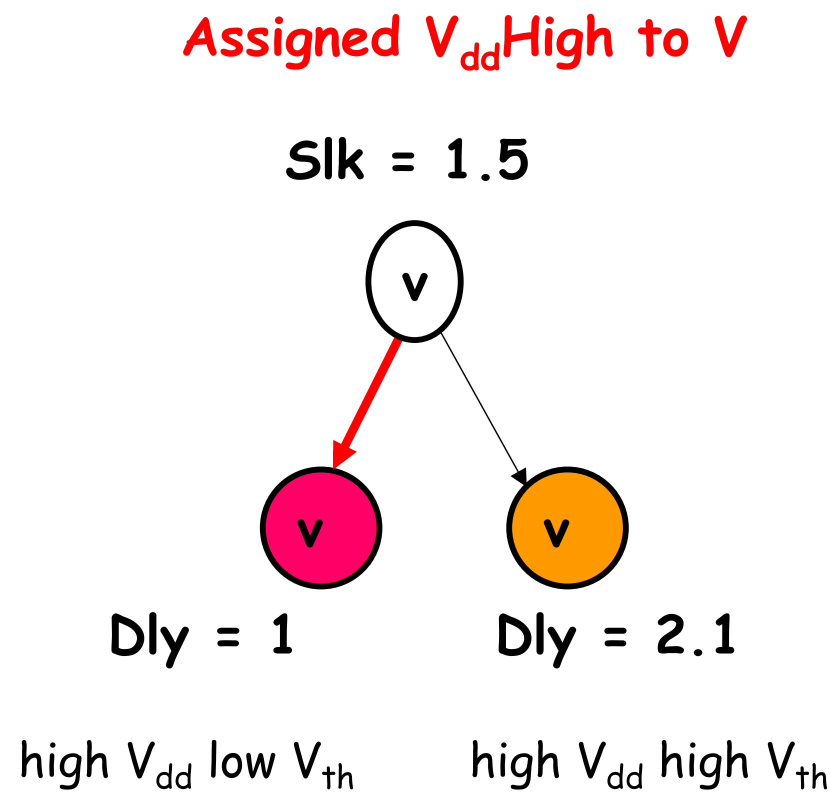
T = topological ordering of gates;
assign low-V_{dd}+high-V_{th} to all PIs;
while (T is not empty)
    $v = T$.pop;
    $dly(v) = \sum_{k=1}^{4} x_{v,k} \cdot d_{v,k}$;
    $vdd(v) = x_{v,1} + x_{v,3}$;
    $v \leftarrow V_{dd}$-L+V_{th}-H;
// V_{dd} mapping
    if ($\exists u \in FI(v) | m(e_{u,v}) = 1$)
        $v \leftarrow V_{dd}$-H;
    if ($vdd(v) > 0$)
        $v \leftarrow V_{dd}$-H;
// LC removal
    if ($\exists u \in FI(v) | u = V_{dd}$-H & $m(e_{u,v}) = 1$ or
    $u = V_{dd}$-L & $m(e_{u,v}) = 1$ and $v = V_{dd}$-L)
        $m(e_{u,v}) \leftarrow 0$
// V_{th} mapping
    if ($v = V_{dd}$-H & $dly(v) <$ delay(V_{dd}-H+V_{th}-H))
        $v \leftarrow V_{th}$-L;
    if ($v = V_{dd}$-L & $dly(v) <$ delay(V_{dd}-L+V_{th}-H))
        $v \leftarrow V_{th}$-L;



LP Relaxation for Voltage State Assignment

Voltage Mapping Algorithm
input: LP-based voltage scaling with LC inserted
output: ILP-based voltage scaling with reduced LC set

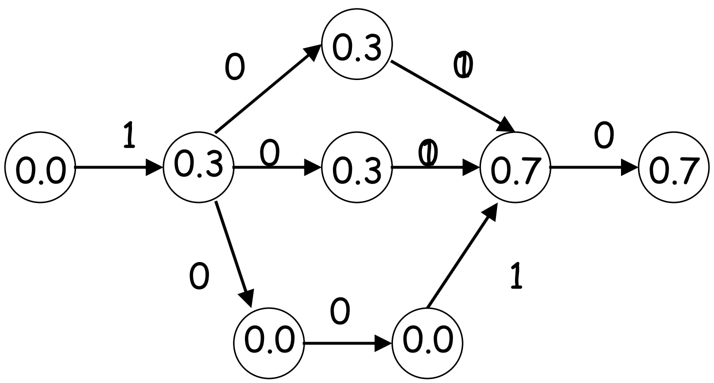
T = topological ordering of gates;
assign low-V_{dd}+high-V_{th} to all PIs;
while (T is not empty)
  $v = T$.pop;
  $dly(v) = \sum_{k=1}^{4} x_{v,k} \cdot d_{v,k}$;
  $vdd(v) = x_{v,1} + x_{v,3}$;
  $v \leftarrow V_{dd}$-L+V_{th}-H;
// V_{dd} mapping
  if ($\exists u \in FI(v) | m(e_{u,v}) = 1$)
    $v \leftarrow V_{dd}$-H;
  if ($vdd(v) > 0$)
    $v \leftarrow V_{dd}$-H;
// LC removal
  if ($\exists u \in FI(v) | u = V_{dd}$-H & $m(e_{u,v}) = 1$ or
  $u = V_{dd}$-L & $m(e_{u,v}) = 1$ and $v = V_{dd}$-L)
    $m(e_{u,v}) \leftarrow 0$
// V_{th} mapping
  if ($v = V_{dd}$-H & $dly(v) <$ delay(V_{dd}-H+V_{th}-H))
    $v \leftarrow V_{th}$-L;
  if ($v = V_{dd}$-L & $dly(v) <$ delay(V_{dd}-L+V_{th}-H))
    $v \leftarrow V_{th}$-L;



LP Relaxation for Voltage State Assignment

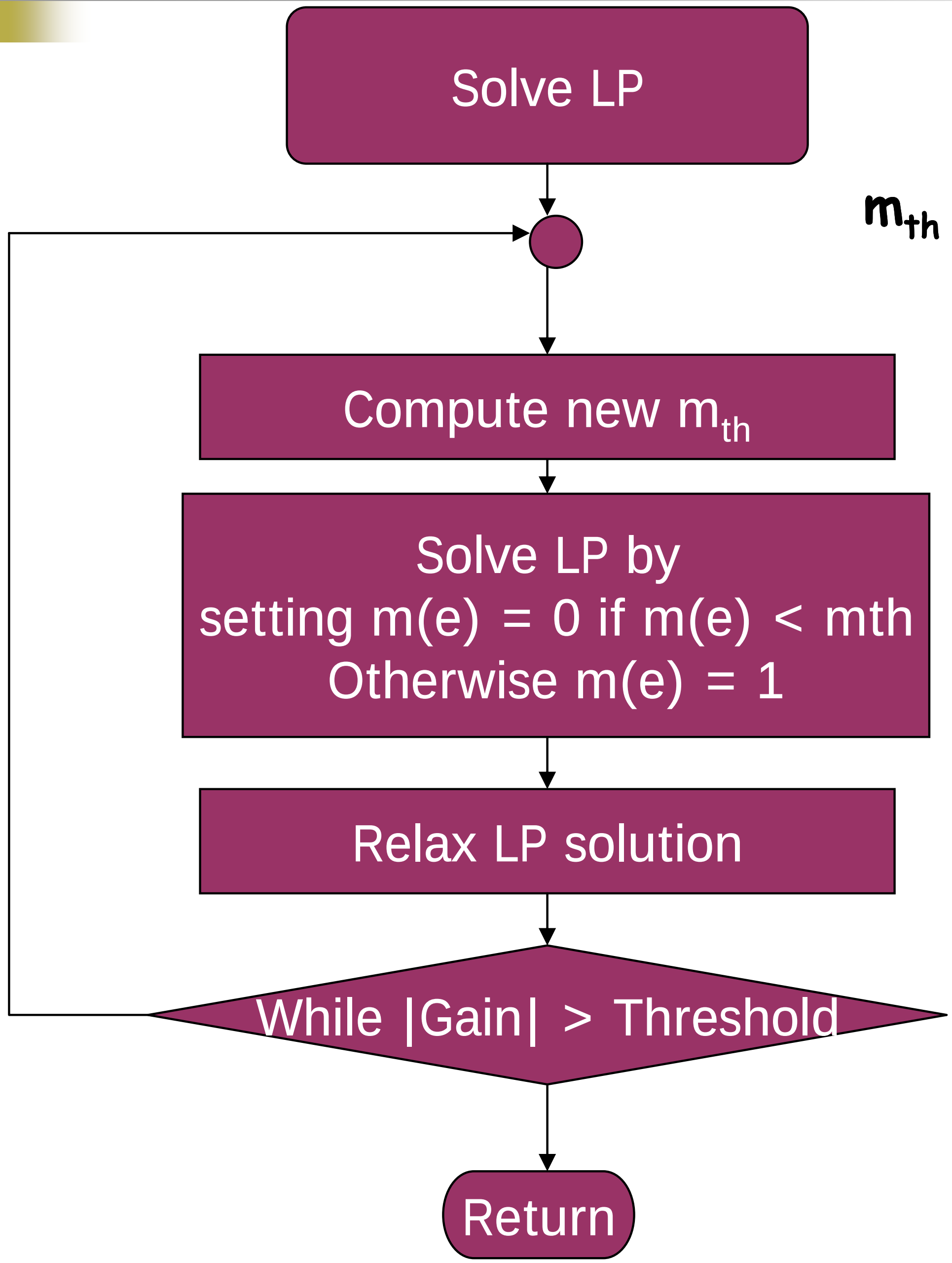
Voltage Mapping Algorithm
input: LP-based voltage scaling with LC inserted
output: ILP-based voltage scaling with reduced LC set

T = topological ordering of gates;
assign low-V_{dd}+high-V_{th} to all PIs;
while (T is not empty)
  $v = T$.pop;
  $dly(v) = \sum_{k=1}^{4} x_{v,k} \cdot d_{v,k}$;
  $vdd(v) = x_{v,1} + x_{v,3}$;
  $v \leftarrow V_{dd}$-L+V_{th}-H;
// V_{dd} mapping
  if ($\exists u \in FI(v) | m(e_{u,v}) = 1$)
    $v \leftarrow V_{dd}$-H;
  if ($vdd(v) > 0$)
    $v \leftarrow V_{dd}$-H;
// LC removal
  if ($\exists u \in FI(v) | u = V_{dd}$-H & $m(e_{u,v}) = 1$ or
  $u = V_{dd}$-L & $m(e_{u,v}) = 1$ and $v = V_{dd}$-L)
    $m(e_{u,v}) \leftarrow 0$
// V_{th} mapping
  if ($v = V_{dd}$-H & $dly(v) <$ delay(V_{dd}-H+V_{th}-H))
    $v \leftarrow V_{th}$-L;
  if ($v = V_{dd}$-L & $dly(v) <$ delay(V_{dd}-L+V_{th}-H))
    $v \leftarrow V_{th}$-L;



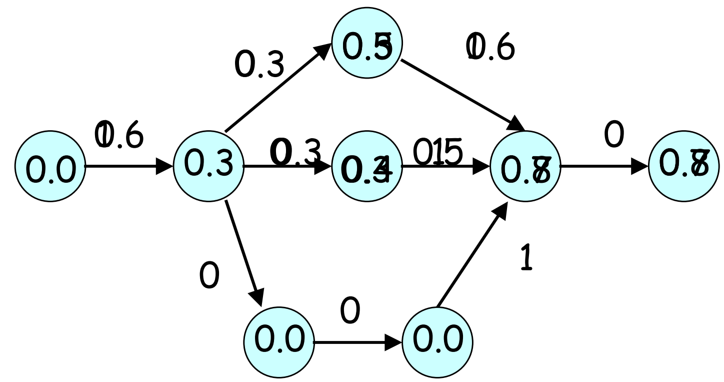
Assume only two states

Gradient Search Algorithm for LC Relaxation



Compute for next m_{th}

m_{th} = 0.5



Post Refinement

```
Post Refinement
input: retimed and voltage scaled solution
output: refined voltage scaling solution
--------------------------------------------
// clustering
1.  perform static timing analysis;
2.  mark all nodes with positive timing slack;
3.  form clusters among marked nodes;
4.  sort clusters based on its size;
// main loop
5.  for (each cluster C)
6.      while (there is power reduction)
7.          for (each node v ∈ C)
8.              power_gain(v, slk(v), C);
9.          z = max power gain node;
10.         commit voltage change for z;
11.         update slack for downstream nodes of z;
```

Outline

- Introduction and Motivation
- Related Work
- Methodology
- Experimental Results
- Conclusions

Impact of Retiming on Power

ckt	Retiming + Scaling [Chabini04]				min FF. retiming			
	V_{dd} (uW)		V_{dd} + V_{th} (uW)		V_{dd} (uW)		V_{dd} + V_{th} (uW)	
	GL	GLF	GL	GLF	GL	GLF	GL	GLF
s641	295.9	372.1	170	246.1	316.6	361.8	187.4	232.6
s713	311.1	399.2	181.9	269.9	330.5	375.8	199.4	244.6
s820	381	404.8	299	322.8	400.8	415	296	310.3
s832	384	407.8	307.4	331.2	403.9	418.2	299.9	314.2
s838	383.5	543	247.6	407	436.9	586.9	283.6	433.5
s1196	567.9	758.3	389.3	579.7	538.7	591	381.8	434.1
s1238	567	764.6	404.8	602.3	546.7	599.1	395.5	447.9
s1488	761.8	821.3	568.1	627.6	765	781.7	535.7	552.4
s1494	761.4	835.2	569.8	643.5	765	781.7	536.2	552.8
Ratio	-	**1**	-	**0.76**	-	**0.93**	-	**0.66**

GL = Gate Power + LC Power
GLF = Gate Power + LC Power + FF Power

Power Comparison on Different Voltage Scaling Techniques (in uW)

ckt	INIT	CVS[Usami95]	LP	ILP
s641	434.3	374.5	232.6	230.6
s713	458	392.3	244.6	243.3
s820	425.7	412.4	310.3	309.7
s832	428.9	415.6	314.2	312.5
s838	627.3	579.6	433.5	428.6
s1196	646.8	616.2	434.1	434.1
s1238	648.4	619.1	447.9	446.6
s1488	796.1	773.8	552.4	551.4
s1494	795.5	773.2	552.8	550.5
ratio	**1**	**0.94**	**0.66**	**0.66**
time	28 sec	29 sec	44 sec	1 day

INIT = all nodes V_{dd}-H + V_{th}-L
LP = Linear Programming

CVS= clustered Voltage Scaling
ILP = Integer Linear Programming

Outline

- Introduction and Motivation
- Related Work
- Methodology
- Experimental Results
- Conclusions

Conclusions

- Power minimization is an important VLSI design issue: both static and dynamic power
- We propose a mathematical model to solve power optimization issue while maintain the target clock period
- The experiment results show up to 30% power reduction

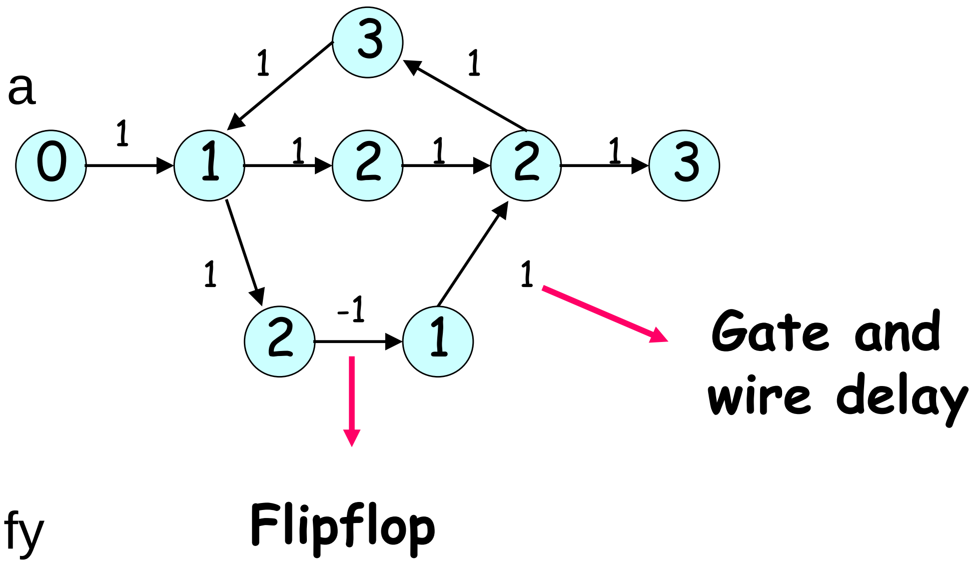
QUESTIONS?

Delay and Power for Voltage Scaling

config	delay	dynamic	leakage
High-Vdd/Low-Vth gate	1.00	1.69	0.44
Low-Vdd/Low-Vth gate	2.53	0.36	0.44
High-Vdd/High-Vth gate	1.24	1.69	0.058
Low-Vdd/High-Vth gate	4.26	0.36	0.058
level conversion FF	11.39	5.07	1.34
level converter	1.77	1.03	0.25

Retiming Algorithm

- FF. edge has weight =
  clock period * number of FF.
- If Bellman-Ford algorithm has a feasible solution, the target clock period is feasible
- Binary search is used to identify smallest feasible clock period (cycle time)



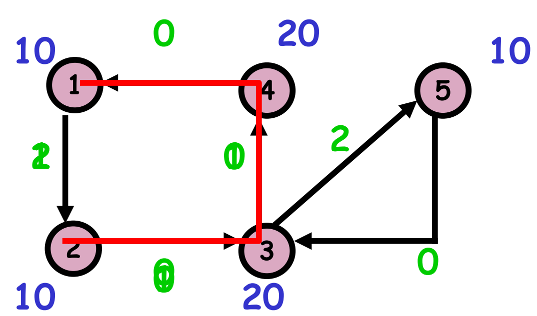
Retiming LP Formulation

$$\text{Minimize } \{FI(v) - FO(v)\} \cdot r(v) \qquad (55)$$

Subject to:

$$r(u) - r(v) \leq w(e_{u,v}),\ \forall e_{u,v} \in E \qquad (56)$$

$$r(u) - r(v) \leq W(u,v),\ \forall D(u,v) > L,\ \forall u,v \in V \qquad (57)$$



Retiming Formulation

Objective:

Minimize the number of flip-flops (FF.)

Constraints:

- Num. FF. has to be satisfied

  $r(u) \leq w(e_{u,v}) + r(v)$
- Num. FF. on critical paths has to be greater than zero

$$\text{Minimize } \{FI(v) - FO(v)\} \cdot r(v) \tag{53}$$

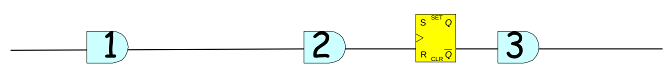
Subject to:

$$r(u) - r(v) \leq w(e_{u,v}), \ \forall e_{u,v} \in E \tag{54}$$

$$r(u) - r(v) \leq W(u,v) - 1, \ \forall D(u,v) > L, \ \forall u,v \in V \tag{55}$$

w(e) 0 0 0 0

Cycle Time (L) =2



D(1,2) = 2
D(1,3) = 3
D(2,3) = 2

r(v) 0 0 0

W(1,2) = 0
W(1,3) = 1
W(2,3) = 1

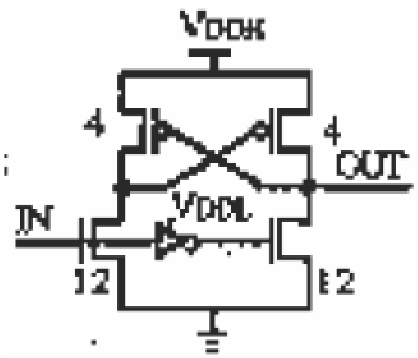
r(v) is the number of FF. moved from fanout of node v to fanin of node v

$w(e_{u,v})$ is the FF. count on edge u,v,

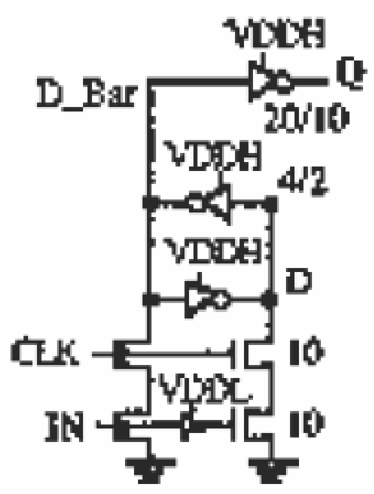
D(u,v) is the maximum delay on path u,v

W(u,v) is minimum number of FF. on path u,v

LC



LCFF



Power Comparison on Different Voltage Scaling Techniques (in uW)

ckt	INIT	CVS[Usami95]	MVVS [Srivastava04]	LP	ILP
s641	434.3	374.5	253.9	232.6	230.6
s713	458	392.3	268.7	244.6	243.3
s820	425.7	412.4	327.7	310.3	309.7
s832	428.9	415.6	331.6	314.2	312.5
s838	627.3	579.6	440.1	433.5	428.6
s1196	646.8	616.2	439.0	434.1	434.1
s1238	648.4	619.1	454.1	447.9	446.6
s1488	796.1	773.8	577.1	552.4	551.4
s1494	795.5	773.2	575.0	552.8	550.5
ratio	**1**	**0.94**	**0.69**	**0.66**	**0.66**
time	28 sec	29 sec	35 sec	44 sec	1 day

INIT = all nodes V_{dd}-H + V_{th}-L
MVS = modified Vdd/Vth and Sizing
CVS= clustered Voltage Scaling
LP = Linear Programming
ILP = Integer Linear Programming