Peak Power Minimization Through Datapath Scheduling

Saraju P. Mohanty, N. Ranganathan and Sunil K. Chappidi
Dept of Computer Science and Engineering, NNRC
University of South Florida, Tampa, FL 33620, USA
smohanty, ranganat, chappidi@csee.usf.edu
Outline of the talk

- Introduction
- Related work
- Target architecture
- Peak power model
- ILP formulations
- Scheduling algorithm
- Experimental results
The peak power is the maximum power consumption of the circuit at any instance during its execution.
Why peak power reduction?

Reduction of peak power consumption is essential:
(i) to maintain supply voltage levels
(ii) to increase reliability
(iii) to use smaller heat sinks
(iv) to make packaging cheaper
Energy Vs Peak power efficient scheduling

Fig. (a) is energy efficient schedule, whereas Fig. (b) is peak power efficient schedule for same resource constraint
Related work
(Energy efficient scheduling using voltage reduction)

- Chang and Pedram [3], 1997 – Dynamic programming
- Johnson and Roy [4], 1997 – ILP based MOVER algorithm using multiple supply voltages
- Lin, Hwang and Wu [5], 1997 – ILP and heuristic for variable voltages (VV) and multicycling (MC)
- Mohanty and Ranganathan [7], 2003 – Heuristic based using multiple supply voltage and dynamic clocking
Related work
(Peak Power efficient scheduling)

- Martin and Knight [6], 1996 – Simultaneous assignment and scheduling
- Raghunathan, Ravi and Raghunathan [10], 2001 – data monitor operations in VHDL
- Shiue [12], 2000 – ILP based and modified force direct scheduling for peak power minimization
- Shiue and Chakrabarti [13], 2000 - ILP model to minimize peak power and area for single voltage
Voltage, Frequency and Power Trade-offs

(i) **voltage** reduction $\rightarrow$ increase in delay

(ii) **frequency** reduction $\rightarrow$ reduction in power (not energy) (and increase in delay)

“Beyond of (i) and (ii) reduction of **switching capacitance** can be considered.”
What is our approach?

Adjust the frequency and reduce the supply voltage for peak power reduction during datapath scheduling.
All functional units have one register each and one multiplexor.

Each functional unit feeds one register only.

The register and the multiplexor operate at the same voltage level as that of the functional units.

Level converters are used when a low-voltage functional unit is driving a high-voltage functional unit.

Operational delay of a FU: \( (d_{FU} + d_{Mux} + d_{Reg} + d_{Conv}) \).
Peak power model

For a DFG let us assume:

c = any control step or clock cycle in DFG
N = total number of control steps in the DFG
R_c = number of resources active in step c (same as number of operations in step c)
\( f_c \) = cycle frequency for control step c
\( \alpha_{i,c} \) = switching at resource i active in step c
\( C_{i,c} \) = load capacitance of resource i active in step c
\( V_{i,c} \) = operating voltage of resource i active in step c
The power consumption for any control step $c$ is given by,

$$P_c = \sum_{i=1\rightarrow R_c} \alpha_{i,c} \ C_{i,c} \ V_{i,c}^2 \ f_c$$

The peak power consumption of the DFG is the maximum power consumption over all the control steps,

$$P_{\text{peak}} = \text{maximum} (P_c)_{c=1\rightarrow N}$$
Using the above two equations the peak power consumption of the DFG is described as,

\[ P_{\text{peak}} = \text{maximum} \left( \sum_{i=1 \rightarrow R_c} \alpha_{i,c} C_{i,c} V_{i,c}^2 f_c \right)_{c=1 \rightarrow N} \]

This would serve as an **objective function** for the scheduling algorithm.
ILP formulations for MVDFC: notations

O : total number of operations in the DFG

\( o_i \) : any operation \( i, 1 \leq i \leq O \)

\( F_{k,v} \) : functional unit of type \( k \) operating at voltage level \( v \)

\( M_{k,v} \) : maximum number of functional units of type \( k \) operating at voltage level \( v \)

\( S_i \) : as soon as possible time stamp for the operation \( o_i \)

\( E_i \) : as late as possible time stamp for the operation \( o_i \)

\( P(i,v,f) \) : power consumption of operation \( o_i \) at voltage level \( v \) and operating frequency \( f \)

\( x_{i,c,v,f} \) : decision variable which takes the value of 1 if operation \( o_i \) is scheduled in control step \( c \) using the functional unit \( F_{k,v} \) and \( c \) has frequency \( f \)
ILP formulations for MVDFC ....

(i) Objective Function
(ii) Uniqueness Constraints
(iii) Precedence Constraints
(iv) Resource Constraints
(v) Frequency Constraints
(vi) Peak Power Constraints
**ILP formulations for MVDFC ....**

**Objective Function:** Minimize $(P_{\text{peak}})$

**Uniqueness Constraints:** ensure that every operation $o_i$ is scheduled to one unique control step and represented as, $\forall i, 1 \leq i \leq O$, 

$$\sum_c \sum_v \sum_f x_{i,c,v,f} = 1$$
Precedence Constraints: guarantee that for an operation \( o_i \), all its predecessors are scheduled in an earlier control step and its successors are scheduled in an later control step and are; \( \forall i, j, o_i \) belong to \( \text{Pred}(o_j) \), 
\[
\sum_v \sum_f \sum_{d=S_i \rightarrow E_i} d \cdot x_{i,c,v,f} - \sum_v \sum_f \sum_{d=S_j \rightarrow E_j} e \cdot x_{j,c,v,f} \leq -1
\]

Resource Constraints: make sure that no control step contains more than \( F_{k,v} \) operations of type \( k \) operating at voltage \( v \) and are enforced as, \( \forall c, 1 \leq c \leq N \) and \( \forall v, \sum_{i \in F_{k,v}} \sum_f x_{i,c,v,f} \leq M_{k,v} \)
ILP formulations for MVDFC …. 

Frequency Constraints: lower operating voltage functional unit can not be scheduled in a higher frequency control step; these constraints are expressed as, $\forall i, 1 \leq i \leq O, \forall c, 1 \leq c \leq N$, if $f < v$, then $x_{i,c,v,f} = 0$.

Peak Power Constraints: ensure that the maximum power consumption of the DFG does not exceed $P_{\text{peak}}$ for any control step and we enforce these constraints as follows, $\forall c, 1 \leq c \leq N$ and $\forall v$,

$$\sum_{i \in F_k} \sum_x x_{i,c,v,f} P(i,v,f) \leq P_{\text{peak}}$$
ILP formulations for MVMC: notations

O : total number of operations in the DFG

\( o_i \) : any operation \( i, 1 \leq i \leq O \)

\( F_{k,v} \) : functional unit of type \( k \) operating at voltage level \( v \)

\( M_{k,v} \) : maximum number of functional units of type \( k \) operating at voltage \( v \)

\( S_i \) : as soon as possible time stamp for the operation \( o_i \)

\( E_i \) : as late as possible time stamp for the operation \( o_i \)

\( P(i,v,f_{\text{clk}}) \) : power consumption of operation \( o_i \) at voltage level \( v \) and operating frequency \( f_{\text{clk}} \)

\( y_{i,v,l,m} \) : decision variable which takes the value of 1 if operation \( o_i \) is using the functional unit \( F_{k,v} \) and scheduled in control steps \( l \rightarrow m \)

\( L_{i,v} \) : latency for operation \( o_i \) using resource operating at voltage \( v \) (in terms of number of clock cycles)
ILP formulations for MVMC ....

(i) Objective Function
(ii) Uniqueness Constraints
(iii) Precedence Constraints
(iv) Resource Constraints
(v) Peak Power Constraints
Objective Function: Minimize \( (P_{peak}) \)

Uniqueness Constraints: ensure that every operation \( o_i \) is scheduled to appropriate control steps within the range \((S_i, E_i)\) and represented as, \( \forall \ i, \ 1 \leq i \leq O, \)
\[
\sum_v \sum_{l=S_i \rightarrow (S_i+E_i+1-L_i,v)} y_{i,v,l,(l+L_i,v-1)} = 1
\]
ILP formulations for MVMC ....

Precedence Constraints: guarantee that for an operation \( o_i \), all its predecessors are scheduled in an earlier control step and its successors are scheduled in an later control step and are; \( \forall i,j, o_i \) belong to \( \text{Pred}(o_j) \),

\[
\sum_v \sum_{l=S_i \rightarrow E_i} (l+L_{i,v}-1) y_{i,v,l,(l+L_{i,v}-1)} - \sum_v \sum_{l=S_j \rightarrow E_j} 1 y_{j,v,l,(l+L_{j,v}-1)} \leq -1
\]
Resource Constraints: make sure that no control step contains more than $F_{k,v}$ operations of type $k$ operating at voltage $v$ and are enforced as,

$$\sum_{\{i \in F_{k,v}\}} \sum_{l} y_{i,v,l,(l+L_i,v-1)} \leq M_{k,v}$$

Peak Power Constraints: ensure that the maximum power consumption of the DFG does not exceed $P_{\text{peak}}$ for any control step and we enforce these constraints as follows, for all $c$, $1 \leq c \leq N$ and for all $v$,

$$\sum_{\{i \in F_{k,v}\}} \sum_{v} y_{i,v,l,(l+L_i,v-1)} P(i,v,f_{\text{clk}}) \leq P_{\text{peak}}$$
**Input:** (i) unscheduled DFG  
(ii) resource constraints  
(iii) number of voltage levels  
(iv) number of frequencies  
(v) delay of resources  

**Output:** scheduled DFG, $f_{\text{base}}$, N, $c_{\text{f}},$ power estimates
Scheduling algorithm ....

Step 1: Find ASAP schedule of the UDFG.
Step 2: Find ALAP schedule of the UDFG.
Step 3: Determine the mobility graphs for each node.
Step 4: Modify the mobility graph for MVMC scheme.
Step 5: Calculate operating frequency of a FU using delay model.
Step 6: Construct the ILP formulations of the DFG.
Step 7: Solve the ILP formulations using LP-Solve.
Step 8: Obtain the scheduled DFG.
Step 9: Determine $f_c$, $f_{\text{base}}$ and $c_{f_i}$ for MVDFC scheme.
Scheduling for MVDFC

(c) Mobility Graph (d) Final Schedule

Example DFG (for RC1)
Scheduling for MVMC

Example DFG (for RC1)
# Experimental results: benchmarks

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Nodes</th>
<th>Symbols</th>
<th>Edges</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Example circuit (EXP) (8 nodes, 3*, 3+, 9 edges)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>FIR filter (11 nodes, 5*, 4+, 19 edges)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>IIR filter (11 nodes, 5*, 4+, 19 edges)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>HAL differential equation solver (13 nodes, 6*, 2+, 2-, 1 &lt;, 16 edges)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Auto-Regressive filter (ARF) (15 nodes, 5*, 8+, 19 edges)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Experimental results: resource constraints

<table>
<thead>
<tr>
<th>Multipliers</th>
<th>ALUs</th>
<th>Serial No</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3V</td>
<td>5.0V</td>
<td>3.3V</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Experimental results: notations

**P_S** : the peak power consumption (in mW) for single supply voltage and single frequency operation (SVSF)

**P_{DFC}** : the peak power consumption (in mW) for MVDFC operation

**P_{MC}** : the peak power consumption (in mW) for multiple supply voltages and multicycle operation

**PDP_S** : the power delay product (in nJ) for SVSF operation

**PDP_{DFC}** : the power delay product (in nJ) for MVDFC operation

**PDP_{MC}** : the power delay product (in nJ) for MVMC operation

\[ \Delta P_{DFC} = \frac{(P_S - P_{DFC})}{P_S} \times 100 : \text{% peak power reduction for MVDFC} \]

\[ \Delta P_{MC} = \frac{(P_S - P_{MC})}{P_S} \times 100 : \text{% peak power reduction for MVMC} \]

\[ \Delta PDP_{DFC} = \frac{(PDP_S - PDP_{DFC})}{PDP_S} \times 100 : \text{% PDP reduction for MVDFC} \]

\[ \Delta PDP_{MC} = \frac{(PDP_S - PDP_{MC})}{PDP_S} \times 100 : \text{% PDP reduction for MVMC} \]
## Experimental results: (% reduction)

<table>
<thead>
<tr>
<th>RCs</th>
<th>$\Delta P_{DFC}$</th>
<th>$\Delta P_{MC}$</th>
<th>$\Delta P_{DFC}$</th>
<th>$\Delta P_{MC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>1</td>
<td>78</td>
<td>55</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>78</td>
<td>35</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>78</td>
<td>56</td>
<td>63</td>
</tr>
<tr>
<td>EXP</td>
<td>1</td>
<td>78</td>
<td>49</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>78</td>
<td>35</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>78</td>
<td>55</td>
<td>61</td>
</tr>
<tr>
<td>FIR</td>
<td>1</td>
<td>69</td>
<td>38</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>78</td>
<td>56</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>78</td>
<td>35</td>
<td>66</td>
</tr>
<tr>
<td>IIR</td>
<td>1</td>
<td>78</td>
<td>29</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>78</td>
<td>36</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>79</td>
<td>56</td>
<td>62</td>
</tr>
</tbody>
</table>
Percentage average reduction

Different benchmark circuits -> Average peak power reduction (%) -> MVDFC

Different benchmark circuits -> Average PDP reduction (%) -> MVDFC

Different benchmark circuits -> Average peak power reduction (%) -> MVMC

Different benchmark circuits -> Average PDP reduction (%) -> MVMC
# Reductions using different schedulers

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(2) FIR</td>
<td>72</td>
<td>49</td>
<td>63</td>
<td>40</td>
<td>23</td>
</tr>
<tr>
<td>(4) HAL</td>
<td>75</td>
<td>41</td>
<td>28</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(5) ARF</td>
<td>78</td>
<td>50</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Conclusions

- Reduction of peak power is essential.
- This paper describes peak power reduction schemes at behavioral level through datapath scheduling.
- The scheduling schemes use ILP based minimization for MVDFC and MVMC mode of circuit design.
- For both the modes the scheduler could achieve significant peak power reduction.
- For some resource constraints there is increase in PDP for MVMC mode design.
- The scheduling schemes are useful for data intensive applications.
- The applicability of the scheduling schemes for pipelining is to be investigated.
- The effect of switching activity is to be taken into account.
- The detail design of controller is to be done.
- The effect on clocking network is to be studied.
Thank you