

Synthesis Scheme for Low Power Designs Under Timing Constraints^{*}

Wang Ling¹, Wen Dongxin¹, Yang Xiaozong¹, and Jiang Yingtao²

(1 College of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

(2 Department of Electrical & Computer Engineering, University of Nevada, Las Vegas, NV89154, USA)

Abstract : To minimize the power consumption with resources operating at multiple voltages a time-constrained algorithm is presented. The input to the scheme is an unscheduled data flow graph (DFG), and timing or resource constraints. Partitioning is considered with scheduling in the proposed algorithm as multiple voltage design can lead to an increase in interconnection complexity at layout level. That is, in the proposed algorithm power consumption is first reduced by the scheduling step, and then the partitioning step takes over to decrease the interconnection complexity. The time-constrained algorithm has time complexity of $O(n^2)$, where n is the number of nodes in the DFG. Experiments with a number of DSP benchmarks show that the proposed algorithm achieves the power reduction under timing constraints by an average of 46.5%.

Key words : low power; multiple supply voltages; partitioning; timing constraints; scheduling

EEACC : 2570D

CLC number : TN4

Document code : A

Article ID : 0253-4177(2005)02-0287-07

1 Introduction

Power considerations have become an increasingly dominant factor in the design of both portable and desktop systems that demand high-speed computations and complex functionalities with low power consumption. An effective way to reduce the power consumption is to lower the supply voltage level of a circuit due to the quadratic proportional relationship between the power consumption and the supply voltage. For this reason, more recently, the use of multiple supply voltages on the chip has attracted attention^[1-9]. This scheme has the advantage of allowing modules on the critical paths to use the highest voltage level (thus meeting the required timing constraints) while allowing modules on non-critical paths

to use lower voltages (thus reducing the energy consumption)^[1].

Most designs using multiple voltage supplies at the behavioral level have the similar synthesis flow, where the emphasis is placed on scheduling the operations. These algorithms can be classified into: (1) only time-constrained^[1,5,7]; (2) only resource constrained^[7]; (3) time and resource constrained^[2-4].

For time-constrained problem, Raje and Sarrafzaden^[5] proposed a variable voltage scheduling algorithm, while Chang and Pedram^[1] presented a dynamic programming technique for power minimization. In Ref. [7], Shiue and Chakrabarti presented scheduling scheme to minimize the power consumption by considering the effect of the interconnect complexity for time constrained and resource constrained problem, respectively. Johnson and Roy^[4] developed

* Project supported by Natural Science Foundation of Heilongjiang Province in China (No. F2004-17)

Wang Ling female, associate professor. Her research interests are in VLSI design and computer aided design including wireless network, hardware/software co-design, high-level synthesis, and low-power system design.

Wen Dongxin female, PhD candidate. Her research interests are in VLSI design, high-level synthesis, and low-power system design.

Yang Xiaozong male, professor. His research interests are computing architecture, fault tolerant computing, fault injection, wireless network, and dependable computing.

Received 7 May 2004, revised manuscript received 31 August 2004

© 2005 Chinese Institute of Electronics

an integer linear programming (ILP)-based formulation for time and resource constrained scheduling problem with exponential time complexity. Manzak and Chakrabarti^[3] presented a heuristic scheduling algorithm for the same problem with polynomial time complexity.

All above mentioned multiple voltage synthesis algorithms as well as others^[1~9] are mainly focused on operation scheduling. If only scheduling is considered, multiple voltage design can lead to an increase in routing complexity at layout level. Recognizing the layout problem, we realize that if we can partition a chip running at multiple voltages with fewer voltage regions and fewer signal passing between regions operating at different voltages at the behavioral synthesis level, the above mentioned layout problem can be significantly alleviated. That is, partitioning the chip into different voltage islands shall be performed along with the operation scheduling.

Realizing the importance of the partitioning in the multiply supply voltage synthesis, we have proposed a resource constrained scheme to minimize power consumption^[10]. In this paper, we present the multiple voltage scheduling and partitioning algorithm under timing constraints. The algorithm reduces power consumption by scheduling and then decreases interconnection complexity between clusters through partitioning. The partitioning information can thus be passed to the architecture netlist generation and floor planning tools to generate the final physical layout with nodes in one cluster mapped into functional units adjacent to each other.

Experiments with a number of DSP benchmarks show that the proposed algorithm achieves the power reduction under timing constraints by an average of 46.5%, respectively. The power reduction in our schemes is independent on the hardware structures.

2 Preliminaries

In this section, we first give a few notations, followed by the timing and the interconnection models.

2.1 Definitions

Definition 1 (data flow graph): A data flow graph (DFG) is a directed acyclic graph $G = (V, E)$, where V is a set of nodes, and E is a set of edges between the nodes. Here each node represents an operation, and a directed edge from node v_i to node v_j means execution of v_i must precede that of v_j .

Definition 2 (timing constraint): Timing constraint is the time available to execute the nodes in a DFG.

Definition 3 (ASAP schedule time): ASAP (as soon as possible) schedule time of any node is the soonest time at which it can be scheduled.

Definition 4 (ALAP schedule time): ALAP (as late as possible) schedule time of any node is the latest time at which it can be scheduled. That is, if the schedule time of any node surpasses its ALAP schedule time, then the timing constraints are violated.

Definition 5 (mobility): The mobility of a node is the difference between its ALAP schedule time and its ASAP schedule time.

Definition 6 (spatial voltage cluster): A spatial voltage cluster $C(V_i)$ is a set of nodes that operate at the same supply voltage V_i . The hardware units assigned to the nodes in a spatial voltage cluster shall be placed near each other in the final layout to minimize the communication costs.

Definition 7 (tightly connected nodes)^[11]: Two nodes are tightly connected if the shortest distance between them, in terms of the number of the edges traversed, is small.

Definition 8 (gain): The gain of a node v with respect to the k -th cluster C_k , denoted as g_{ik} , is the weighed sum of the edges between v and the nodes in C_k .

Definition 9 (resource): A resource refers to a hardware unit operating at a given voltage level (e. g. a 3.3V adder/multiplier).

Definition 10 (slack): The slack is the time between the ALAP schedule time and the computation time of the node.

2.2 Timing model

Let c denote a control step (clock cycle), the basic unit of time used in the DFG in behavioral level. For a given length of c -step, an operation may thus become a multi-cycle operation.

Let t_i be the starting time of any node v_i (t_i is also referred as the schedule time of v_i). For all the j parent nodes of node v_i in the DFG, denoted as $v_{i1}, v_{i2}, \dots, v_{ij}$, we have

$$t_i = \max\{ (t_{i1} + d_{i1}), (t_{i2} + d_{i2}), \dots, (t_{ij} + d_{ij}) \} \tag{1}$$

$$t_i^e = t_i + d_i \tag{2}$$

where d_i is the delay of node v_i , t_i^e is the ending time of node v_i , $t_{i1}, t_{i2}, \dots, t_{ij}$ are the corresponding starting time of $v_{i1}, v_{i2}, \dots, v_{ij}$, and $d_{i1}, d_{i2}, \dots, d_{ij}$ are the corresponding execution time (delay) of $v_{i1}, v_{i2}, \dots, v_{ij}$.

Thus, for a DFG with n nodes, its total schedule time T is derived as

$$T = \max_i\{ t_i^e \}, i \in \{1, 2, \dots, n\} \tag{3}$$

Proposition 1^[12] The delay of an operation is inversely proportional to its supply voltage.

Proposition 2^[3] Nodes with high power over delay ratio have to be assigned to lower-voltage resources to minimize the total power consumption under the time constraints in the multiple voltage synthesis.

2.3 Interconnection model

Assume a graph $G = (V, E)$ has n nodes, and all the nodes can be organized as r clusters (C_1, C_2, \dots, C_r).

We define factor y_{ik} ,

$$y_{ik} = \begin{cases} 0, & \text{node } v_i \text{ belongs to the cluster } C_k \\ 1, & \text{otherwise} \end{cases} \tag{4}$$

where $i = 1, 2, \dots, n, k = 1, 2, \dots, r$.

The total interconnect cost I thus is given as

$$I = \sum_{i=1}^n \sum_{k=1}^r y_{ik} g_{ik} \tag{5}$$

where g_{ik} is the gain of node v_i with respect to cluster C_k .

2.4 Problem formulation

The time-constrained multiple voltage scheduling and partitioning problem is defined as follows.

Given a DFG and the timing constraints, find a schedule and a partition that satisfy the given timing constraints with minimum power consumption due to the power consumed by the resources (P_{res}) and the power consumed by the interconnects (P_{int}).

For a DFG with a total number of n operations, P_{res} can be decomposed into

$$P_{res} = \sum_{i=1}^n P_{res}^i$$

Here P_{res}^i is the power consumed by i -th resource, and n is the total number of resources. As our focus is on the inter-cluster interconnections (between macroblocks, such as different datapaths), we shall have

$$P_{int} = \sum_{j=1}^N P_{int}^j$$

Here P_{int}^j is the power consumed by j -th interconnections between clusters. N is the total number of interconnections between clusters C_1, C_2, \dots, C_r , and r is the total number of clusters.

The objective function is given as

$$\text{Min}\{ \sum_{j=1}^N P_{int}^j + \sum_{i=1}^n P_{res}^i \} \tag{6}$$

Subject to

$$T \leq T_c \tag{7}$$

where T is the total schedule time, T_c is the timing constraint, N is n .

3 Time-constrained algorithm

3.1 Algorithm

In this section, we present an algorithm to deal with the time-constrained multiple voltage scheduling and partitioning problem. The inputs to the proposed algorithm consist of the DFG representation of a circuit, the timing constraints, and a design library with fully characterized resources. The proposed algorithm has the following four steps:

(1) Step 1: one resource running at the lowest possible voltage is initially selected for each node from

a design library ;

(2) Step 2 :nodes in the DFG are then scheduled to minimize the power consumption due to the resources ;

(3) Step 3 :nodes are clustered to form voltage islands to minimize the power consumption due to the interconnections ;

(4) Step 4 :nodes are finally bound to the resources.

Note that in both steps 2 and 3 ,some nodes may have to be reassigned to the resources running at different voltage levels to satisfy the timing constraints as well as to reduce the interconnections between clusters.

3.1.1 Step 1 :initial resource assignment

Power consumption is quadratically dependent on the supply voltage ,which is the highest contributing factor for power dissipation. Hence ,in this step ,we attempt to assign the nodes of the DFG to the resources operating at the lowest possible voltages. That is ,a resource R is searched for each node v_i such that it operates at the lowest possible voltage and the following timing constraint shall be satisfied :

$$d \leq m + d_i^M \quad (8)$$

where d is the delay of the resource R that is assigned to v_i , m is the mobility of node v_i ,and d_i^M is the delay of node v_i if node v_i is assigned to the highest voltage resource.

3.1.2 Step 2 :scheduling

After the initial resource is assigned to each node ,even though the timing constraint is satisfied at the node level (Eq. (8)) ,violation of the timing constraints may occur at the DFG level (i. e. ,Eq. (7) may not be satisfied). Thus ,the feasibility of the resource assignment has to be evaluated :

(1) If the schedule time of a node v_i surpasses its ALAP schedule time ,among all the nodes that have been scheduled before v_i ,the node with the lowest power over delay ratio (proposition 2) shall be reassigned to the resources running at a higher voltage (e. g. ,a 3.3V adder has a lower power over delay ratio than a 3.3V multiplier) ;

(2) The scheduling is iteratively performed until

the resource assignments for all the nodes satisfy the timing constraints.

The scheduling algorithm is listed in Fig. 1.

```

Schedule_Nodes_To_Minimize_Function_Power(G(V, E)) {
while (all nodes are not scheduled) {
put all the nodes that do not have parent nodes into the
Ready List ;
while( ! Ready List) {
find node  $v_i$  in the ready list with the least mobility ;
Schedule the node  $v_i$  ;
if (Schedule_Node( $v_i$ ) == False) timing constraints
are violated.
break ;
add all the child nodes of  $v_i$  into the ready list ;
delete  $v_i$  from the ready list ;
}
}
}
Boolean Schedule_Node( $v_i$ ) {
 $t = t_i$ ;  $t_i$  is denoted in Eq. (1)
if ( $t > l_i$ ) {  $l_i$  is the schedule time of  $v_i$  in ALAP scheduling
 $D = t - l_i$ ;  $D$  is denoted as lag time.
while ( $D > 0$ ) {
select the node  $v_j$  that has the lowest power over delay ra-
tio
among  $v_j$  and  $v_i$ 's ancestor nodes ;
increase  $v_j$ 's voltage by one level ,and subsequently ,
the delay of  $v_j$  is reduced by  $D_1$  ;
 $D = D - D_1$  ;
 $v_i = v_j$  ;
}
return False ;
}
else {
schedule node  $v_i$  at the schedule time  $t$  ;
return True ;
}
}
}

```

Fig. 1 Step 2 in the time-constrained scheme

3.1.3 Step 3 :partitioning

Whenever two resources running at two different voltage levels communicate with each other ,a level shifter is needed at the interface. To reduce the number of the level shifters ,initial partitioning is applied to assign the nodes running at the same voltage into the same spatial cluster. Furthermore ,in order to reduce the power consumption due to the interconnections between the clusters ,nodes may be moved from one cluster to another cluster so that the nodes within one cluster are tightly connected (defined in section 2.1). Here two kinds of node movements have to be considered :

(1) If a node needs to be moved from a lower voltage cluster to a higher voltage cluster, the node movement is allowed. Note that in this case, the timing constraints can still be satisfied, as stated in Lemma 1.

(2) In the case when a node is to be moved from a higher voltage cluster to a lower voltage cluster, if the timing constraints are violated, this node movement is not allowed; otherwise, the node shall be moved.

Lemma 1: Given a total schedule time T that satisfies the timing constraints T_c , if the supply voltage of any operation is increased, T_c can still be satisfied with T .

(Proof: it can be easily shown from proposition 1.)

The partitioning algorithm is listed in Fig. 2.

```

Partition _ Nodes _ To _ Minimize _ Interconnection _ Power
(  $G(V, E), N, T_c$  ) {
Initialize  $M$  clusters  $C(V_i), i = 1, \dots, M$ ;  $M$  is the number
of voltage levels;
count = 0;
while (count <  $N$ ) {  $N$  is the user-defined number of iterations.
for (each cluster  $C_j(V_j)$ ) {
pick up the node  $v_h$  with the maximum gain  $g_{hx}(v_h, C_x(V_x))$ ;
if (the number of nodes in the cluster  $C_x(V_x)$  is less than
 $S$ ) {
if ( $V_y > V_i$ ) move  $v_h$  into cluster  $C_x(V_x)$ ;
else {
schedule all the nodes;
if ( $T > T_c$ )  $T$ :total schedule time,  $T_c$ :timing constraints
move  $v_h$  into cluster  $C_x(V_x)$ ;
}
}
}
count = count + 1;
}
return the best partition;
}
    
```

Fig. 2 Step 3 in time-constrained scheme

3.1.4 Step 4: resource binding

After obtaining a satisfactory schedule through steps 1, 2, and 3, resource binding takes place. As we recognize that the interconnection costs between the clusters are significantly higher than those inside the

clusters, the binding algorithm is designed to avoid resource sharing between clusters while attempt to maximize the resource sharing among the nodes inside a cluster.

The binding algorithm is listed in Fig. 3.

```

Binding _ Nodes _ To _ Resources (  $G(V, E)$  ) {
for (each cluster  $C_i(V_j)$ ) {
put all the nodes from cluster  $C_i(V_j)$  into the list LS;
LR = null; LR is the bound resources list
while (! LS) {
In LS, pick up a node  $v_j$  whose parent nodes are not in
LS;
if (the desired resource R operating at  $V_j$  is available in
the LR)
bind node  $v_j$  to resource R;
else {
bind  $v_j$  to a resource R operating at  $V_j$  at control
steps from  $t_j$  to  $t_j^e$ ;
 $t_j, t_j^e$  are defined in Eqs. (1) and (2).
put R into LR;
}
delete node  $v_j$  from LS;
}
}
}
    
```

Fig. 3 Step 4 in time-constrained scheme

3.2 Complexity analysis

The complexity of ALAP and ASAP is $O(n)$, and the complexity of search of initial resources is $O(ln)$, where l is the number of types of resources in the design library and n is the number of nodes in DFG. In a meaningful design, such as filter design, we shall have $l \ll n$. Therefore, the complexity of step 1 is deemed to be $O(n)$. In the scheduling step, the worst case happens when the schedule time of each node exceeds its ALAP schedule time. In this way, the scheduling for each node needs to be performed n times. Hence, the time complexity of step 2 is $O(n^2)$.

In the partitioning step (step 3), the worst case occurs when each node movement triggers a check for possible time constraint violations. As the node movements are performed MN times, where $N > n$. M is the number of voltage levels, and N is the user-defined number of iterations. The complexity for one

scheduling run is $O(n)$, and the time complexity of the partitioning step is again $O(n^2)$.

As the last binding step is dominated by a linear search procedure, the complexity of step 4 is $O(n)$.

Putting everything together, we can see that the overall complexity of the proposed time-constrained multiple voltage scheduling and partitioning algorithm is $O(n^2)$.

4 Experiment results

In this section, we present the results obtained by running our time-constrained scheduling and partitioning algorithm on some high level synthesis benchmarks (discrete cosine transform, lattice filter, 7th order IIR filter, elliptical wave filter, volterra filter, differential equation filter, wavelet filter, and Wdf7 filter). A library with fully characterized resources^[10] is used. The switching activity of nodes is assumed to be 0.5. The number of interconnections among clusters is directly proportional to power consumption, and thus it serves as an indicator of the power reduction. The number of nodes of each circuit is reported in column Nn of Table 1. The timing constraint T_c (column T_c in Table 1) is set as $1.5T$, and $2T$ for each case, where T is the delay obtained by ASAP schedule when all the sources operate at 5V.

In Table 1, Dct, for instance, has the timing constraints T_c equal to 25 control steps. If all the resources in the circuit is powered by 5V supplies, the total power consumption, as reported in column E_{fun}^5 , will be 43132pJ; however, the power consumption by resources operating at the multiple voltages E_{fun} is dropped to 36112.2pJ, a merely 16.2% reduction shown in column R1. Column I shows the interconnect cost (weighted sum of all interconnects) if the nodes of the DFG are partitioned into a number of clusters. For instance, the lattice filter has a interconnect cost of 30, if the total number of clusters is 4 (shown in column Nc). Compared to the case when no partitioning is performed (i. e., only one cluster for its nodes as shown in column I_b), we can see the interconnect cost has been reduced by 38%. On

average, the interconnection reduction is 35% across all 8 benchmark circuits. In total, the average resource power reduction is 38.75% and 60% and the average interconnection reduction is 7.75% and 12.25%, when timing constraint is 1.5 times and 2 times of the ASAP total schedule time, respectively.

Table 1 Experiment results by time constrained scheme

Benchmark	Nn	T_c	E_{fun}^5 / pJ	E_{fun} / pJ	R1 %	I_b	I	Nc
Lattice	22	36	23598	17182.8	27.1	55	30	4
		48		16635.8	29.5		30	4
Diffeq	10	14	15614	8944.1	42.7	70	24	4
		28		3179.4	79.6		24	4
Ellipf	37	40	23100	22885.2	0.9	119	32	4
		60		15576.6	32.5		44	4
Iir7	36	30	39212	16953.5	56.7	107	59	4
		40		9909.1	74.7		62	4
Wdf7	50	45	49464	29021.1	41.3	120	60	4
		60		19520.9	60.3		69	4
Dct	42	25	43132	36112.2	16.2	132	48	4
		34		16712.6	53.2		54	4
Volterra	32	30	43748	33747.5	22.8	81	47	4
		40		19797.9	54.7		48	4
Wavelet	67	42	73180	33095.2	54.7	195	108	4
		56		18676.4	74.7		108	4

5 Conclusion

In this paper, we have presented the scheme to schedule and partition a data-flow graph under either timing constraints. The proposed algorithm tends to minimize the power consumption and to decrease the interconnection complexity. The time-constrained algorithm has a time complexity of $O(n^2)$, where n is the number of nodes in the DFG. Experiments with a number of DSP benchmarks show that the proposed algorithm under timing constraints achieves the power reduction by an average of 46.5%.

References

- [1] Chang J M, Pedram M. Energy minimization using multiple supply voltages. IEEE Trans VLSI, 1997, 5:157
- [2] Lin Y R, Hwang C T. Scheduling techniques for variable voltage low power designs. ACM Trans Design Automat Systems, 1997, 2:81
- [3] Manzak A, Chakrabarti C. A lower power scheduling scheme

- with resources operating at multiple voltages. IEEE Trans VLSI, 2002, 10:6
- [4] Johnson M C, Roy K. Datapath scheduling with multiple supply voltages and level converters. ACM Trans Design Automat Electronic Systems, 1997, 2:227
- [5] Raje S, Sarrafzadeh M. Variable voltage scheduling. Proc Int Symp Low Power Designing, 1995:9
- [6] Raje S, Sarrafzadeh M. Scheduling with multiple voltages. Integr VLSIJ, 1997:37
- [7] Shiue W T, Chaitali C. Low power scheduling with resources operating at multiple voltages. IEEE Trans CASII, 2000, 47:536
- [8] Usami K, Horowitz M. Clustered voltage scaling technique for low-power design. Proc Int Workshop Low Power Design, 1995:3
- [9] Martin R, Knight J. Power profiler: optimizing ASIC's power consumption at the behavioral level. Proc IEEE/ACM Design Automat Conf, 1995:42
- [10] Wang L, Jiang Y, Selvaraj H. Scheduling and optimal voltage selection with multiple supply voltages under resources constraints. Proceeding of International VLSI Conference, 2003:272
- [11] Mehra R, Guerra L M, Rabaey J M. Low power architectural synthesis and the impact of exploiting locality. J VLSI Signal Processing, 1996, 13:239
- [12] Chandrakasan P, Sheng S, Brodersen R W. Low power CMOS digital design. IEEE J Solid State Circuits, 1992, 27:473

一种基于时间限制条件的低功耗高层综合设计方案*

王 玲¹ 温东新¹ 杨孝宗¹ 蒋颖涛²

(1 哈尔滨工业大学计算机学院, 哈尔滨 150001)

(2 美国内华达大学电子与计算机工程系, 拉斯维加斯 NV89154, 美国)

摘要: 提出了多电压时间限制下电路功耗最小的高层综合设计算法, 其输入为数据流图及时间限制条件. 由于多电压设计会引起低层布局时的连线复杂性提高, 所以提出的算法在进行高层调度过程同时考虑了低层分区问题, 即算法利用调度步骤降低功耗, 利用分区步骤来减小连线的复杂性. 该算法的时间复杂性为 $O(n^2)$, n 是 DFG 图中的结点个数. 大量的 DSP 基准实验表明该算法使得电路功耗平均降低 46.5%.

关键词: 低功耗; 多供应电压; 分区; 时间限制; 调度

EEACC: 2570D

中图分类号: TN4 文献标识码: A 文章编号: 0253-4177(2005)02-0287-07

*黑龙江省自然科学基金资助项目(批准号:F2004-17)

王 玲 女, 副教授, 主要研究方向是 VLSI 设计及综合设计, 包括硬件/软件设计, 高层综合设计和低功耗系统设计.

温东新 女, 博士研究生, 主要研究方向是 VLSI 设计及计算机辅助设计, 包括硬件/软件设计、高层综合设计和低功耗系统设计.

杨孝宗 男, 教授, 博士生导师, 主要研究方向是容错技术、移动计算等.

2004-05-07 收到, 2004-08-31 定稿