ON RETIMING OF MULTIRATE DSP ALGORITHMS

Vojin Živojnović *Rainer Schoenen*

Integrated Systems for Signal Processing, IS2
Aachen University of Technology
Templergraben 55, 52056-Aachen, Germany

ABSTRACT

In the paper retiming of DSP algorithms exhibiting multirate behavior is treated. Using the non-ordinary marked graph model and the reachability theory, we provide a new condition for valid retiming of multirate graphs. We show that for a graph with n nodes the reachability condition can be split into the reachability condition for the topologically-equivalent unit-rate graph (all rates set to one), and $(n^2-n)/2$ rate-dependent conditions. Using this property a class of equivalent graphs of reduced complexity is introduced which are equivalent in terms of retiming. Additionally, the circuit-based necessary condition for valid retiming of multirate graphs is extended for the sufficient part.

1. INTRODUCTION

Retiming was introduced as a technique to optimize hardware circuits by redistributing registers without affecting functionality [1]. Retiming is also useful for DSP software design. It changes precedence constraints among instructions or tasks, and can improve single-processor [2] and multiprocessor [3,4] schedules. In both cases, hardware and software design, marked graph can be used as an appropriate model of computation, and retiming is a transformation changing the distribution of tokens on arcs.

This paper extends retiming principles to non-ordinary marked graphs, characterized by nodes consuming and producing one or more tokens at each firing. Non-ordinary marked graphs [5] are equivalent to synchronous data-flow graphs [6] and a special case of computation graphs [7]. Necessity for this work emerged during development of a commercial DSP code synthesis tool with support for multirate processing.

Introduction of rates is not a straightforward extension of the ordinary case. The main difference lies in the incidence matrix which is not totally unimodular in the non-ordinary case. Whereas in the ordinary case the solution for the retiming vector can be found using linear programming (LP), for non-ordinary marked graphs integer linear programming (ILP) has to be used. Additionally, the ordinary token conservation theorem is not valid anymore, limiting the applicability of numerous useful results developed for the ordinary case.

In the past retiming was treated mostly as ordinary (unit-rate) retiming. Only marginal treatment of non-ordinary (multirate) retiming can be found (e.g. in [3]).

The focus of this paper is on reachability of non-ordinary marked graphs. It continues along the work of Teruel et al. [8], and provides new reachability results useful for retiming of multirate DSP algorithms.

After the introduction, we revise the background and introduce the notation. In Section 3 the relation between retiming and reachability is discussed and the existing reachability results are reviewed. Section 4 introduces the rate equalization transformation which simplifies the presentation of the reachability results. Main contribution of the paper is the new reachability theorem presented in Section 5. Using the theorem a new equivalence transformation for retiming is presented in Section 6. Optimization using retiming is discussed in Section 7. Finally, Section 8 presents the conclusions.

2. BACKGROUND AND NOTATION

We assume that the underlying DSP algorithm can be represented as a double-weighted directed graph (marked graph) $G = \{V, E, W, M\}$, with n nodes $v \in V$ and m arcs $e \in E$ which model the processing functions and their connections, respectively. We assign a pair of nonnegative integers $w(e) = (w_+(e), w_-(e)) \in W$, and a nonnegative integer number $m(e) \in M$ of tokens to each arc e. Vector W is called the rate vector, and M is the marking of the graph. The incidence matrix $B = [b_{ij}]$ is an $n \times m$ matrix with entry $b_{ij} = w_+(j)$ ($b_{ij} = -w_-(j)$) if arc j leaves (meets) node i. The entry is zero otherwise. Matrix B_+ (B_-) will denote the matrix $b_{ij} = w_+(j)$ ($b_{ij} = w_-(j)$) where arc j leaves (meets) node i. The left annulator of matrix B^T is the q-vector, i.e. $B^T q = 0$ [6].

If all rates of the marked graph are equal to one, the graph is called ordinary. Graph $\bar{G} = (V, E, 1, M)$ is the ordinary topological equivalent of graph $G = (V, E, W, M)$, and is obtained by setting all rates of the original graph to value one. Its incidence matrix is denoted $\bar{B}$.

Circuit matrix C of a graph is a matrix having as rows all solutions of $Bx = 0$. The fundamental circuit matrix C_f is an $(m - n + 1) \times m$ matrix of linearly independent rows of C.

Firing of node v is legal if $m(e) \geq w_-(e)$, for all arcs e meeting node v. The activation vector U_k is an $n \times 1$

vector of $n-1$ zeros and a single one at the position of the activated node. The change of the marking at activation U_k is described by the state equation $M_k = M_{k-1} + B^T U_k$, $k = 1, 2, \ldots$. In the sequel we shall assume that only live graphs are treated, i.e. at each moment at least one node of the graph can be fired.

3. REACHABILITY AND RETIMING

The theoretical base for the retiming transformation is determinacy of computation graphs [7]. Determinacy guarantees that the sequence of tokens appearing on each arc is independent of the firing sequence, as long as the firing is valid. As a consequence, any marking obtained by valid firing of the nodes can be used as an initial marking, without affecting functionality. Finding a new, advantageous initial marking is in the essence of the retiming transformation.

In the theory of marked graphs the set of functionally equivalent initial markings is denoted as the reachable set.

Definition 1: Marking M is reachable from M_0, if there exists a legal firing sequence $\{U_1, U_2, \ldots\}$ such that $M = M_0 + B^T \sum_k U_k$. The reachable space of M_0 is denoted by $\mathcal{R}(M_0)$.

Figure 1 presents a simple non-ordinary marked graph with initial marking $M_0 = (3, 1, 0)$ and the reachability space $M = \mathcal{R}(M_0)$ denoted as points in the 3D space spanned by the marking vector (m_1, m_2, m_3). Solid directed lines denote marking transitions initiated by firing of a node.

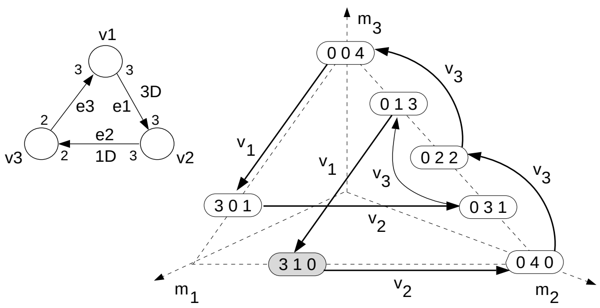

Figure 1. Reachability Space.

The following two theorems are well known results which provide the necessary and sufficient condition for reachability of ordinary graphs [5].

Theorem 1 [State equation based reachability condition for ordinary marked graphs]: Let $G = (V, E, M_0)$ be an ordinary graph with incidence matrix B. Marking M_r is reachable from M_0 iff $B^T r = M_r - M_0$ has a solution.

Remark: Existence of a solution of $B^T r = M_r - M_0$ is already necessary and sufficient for the existence of a legal firing sequence $\{U_1, U_2, \ldots\}$ converting marking M_0 to M_r.

Theorem 2 [Circuit based reachability condition for ordinary marked graphs]: Let $G = (V, E, M_0)$ be an ordinary graph with fundamental circuit matrix C_f. Marking M_r is reachable from M_0 iff $C_f(M_r - M_0) = 0$.

In the case of non-ordinary marked graphs only the state equation based reachability condition is known [8].

Theorem 3 [State equation based reachability condition for non-ordinary marked graphs]: Let (G, M_0) be a non-ordinary marked graph with incidence matrix B. Marking M_r is reachable from M_0 iff $B^T r = M_r - M_0$ has an integral solution.

Main reason for the lack of an equivalent circuit-based reachability result for non-ordinary marked graphs is the existence and use of techniques transforming non-ordinary to ordinary graphs. In the field of marked graphs the transformation proposed by Hack [9] converts the non-ordinary marked graph into an ordinary Petri net, thereby loosing the advantages of the marked graph theory. An alternative equivalence transformation was proposed by Lee [10]. This transformation is closed in the sense that the transformed graph is a marked graph again. Main drawbacks of this procedure are high complexity of the transformed graph, as well as the loss of visual correspondence to the original graph.

4. TOKEN REDEFINITION AND GRAPH EQUALIZATION

In order to proceed, we have to introduce some additional definitions and transformations.

Definition 2: Token redefinition of arc e of marked graph $G = (V, E, W, M)$ is a transformation in which the arc input rate $w_+(e)$, arc output rate $w_-(e)$, and marking $m(e)$ are multiplied by a positive value $\mu(e)$, such that $\mu(e)w_+(e)$, $\mu(e)w_-(e)$, and $\mu(e)m(e)$ are integers.

Remarks: It is easy to show that token redefinition preserves liveness, consistency, and the value of the q vector. Also, from Theorem 1 it is obvious that token redefinition preserves reachability.

Definitions 3: An equalized node is a node with all rates equal. An equalized marked graph is a graph of equalized nodes. The n-dimensional vector of node rates $s(v)$ of an equalized graph will be denoted by S.

Theorem 4: Let $G = (V, E, W, M)$ be a consistent marked graph. There exists a diagonal $m \times m$ matrix of token redefinitions $\mu = \mathrm{lcm}(B_+^T q)\mathrm{diag}(B_+^T q)^{-1}$, such that $G_\mu = (V, E, \mu W, \mu M)$ is an equalized marked graph.

Proof: If G is consistent, then $B_+^T q > 0$ follows. In this case $\mu = \mathrm{lcm}(B_+^T q)\mathrm{diag}(B_+^T q)^{-1}$ can be computed for every graph. For every node v in $G_\mu = (V, E, \mu W, \mu M)$, the number of tokens processed on each arc of node v during one iteration is equal, i.e. $\mu(e_1)w(e_1)q(v) = \mu(e_2)w(e_2)q(v)$ for all pairs (e_1, e_2) of arcs of node v, and the equalization property of G_μ follows.

Graph equalization is similar to the normalization proposed in [3], and was used in the context of hardware design in [11].

5. REACHABILITY OF NON-ORDINARY MARKED GRAPHS

The following reachability theorems are the main contribution of the paper. For the sake of brevity and simplicity, the theorems are provided for equalized graphs. Nevertheless, it is easy to apply them to other marked graphs, too. Equalization does not change the reachability condition of the graph, and the inverse transformation of an equalized graph always exists. So, all the results which are valid for equalized graphs can be applied to other marked graphs equally well.

Theorem 5 [State equation based reachability condition for equalized non-ordinary graphs]: Let $G = (V, E, S, M_0)$ be an equalized non-ordinary graph. Marking M_r is reachable from M_0 iff

$$\bar{B}^T \bar{r} = M_r - M_0 \qquad (1)$$

has a solution such that

$$\bar{r}_i \equiv \bar{r}_j \pmod{\gcd(s_i, s_j)}, \quad 1 \le i < j \le n. \qquad (2)$$

Proof: See Appendix A.

Remark: Theorem 5 decouples the reachability condition into two conditions. Equation (1) is the retiming validity condition under assumption that all rates are set to one. Equation (2) introduces $(n^2 - n)/2$ congruental conditions which take the rates into account.

The following theorem is the extension of Theorem 2 to the non-ordinary case.

Theorem 6 [Circuit based reachability condition for equalized non-ordinary graphs]: Let $G = (V, E, S, M_0)$ be an equalized non-ordinary graph, and C_f its fundamental circuit matrix. Marking M_r is reachable from M_0 iff

$$C_f(M_r - M_0) = 0 \qquad (3)$$

and

$$m_r(i,j) \equiv m_0(i,j) \pmod{\gcd(s_i, s_j)}, \quad 1 \le i < j \le n \qquad (4)$$

where $m(i,j)$ is the number of tokens on a path connecting nodes i and j.

Remarks: We define the remainder marking (or R-marking) of a path connecting nodes i and j as:

$$m_R(i,j) = m(i,j) \bmod \gcd(s_i, s_j). \qquad (5)$$

Condition (4) states simply that the R-marking on any path connecting nodes i and j has to be invariant under retiming. Figure 2 provides an example. The R-marking on the path connecting nodes A and C is 1, and on the path connecting nodes B and D is 2.

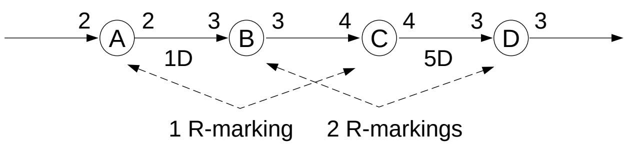

Figure 2. Remainder Marking.

6. RETIMING-EQUIVALENT UNIT-RATE GRAPH

For the purpose of retiming we can modify the rates of the multirate graph in such a way that condition (2) is not changed and thereafter transform the graph to its unit-rate equivalent. Figures 3, 4, and 5 provide an illustrative example. Figure 3 presents the original multirate graph, Figure 4 its unit-rate equivalent, and Figure 5 the retiming-equivalent unit-rate graph.

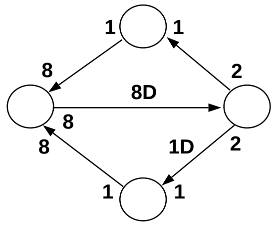

Figure 3. Original Graph.

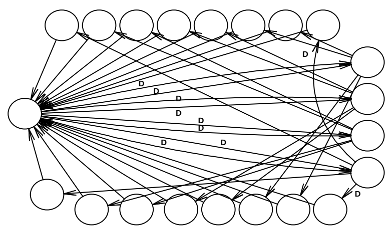

Figure 4. Equivalent Unit-Rate Graph.

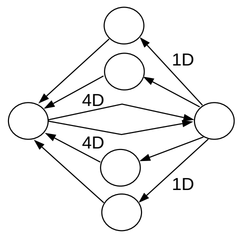

Figure 5. Retiming-Equivalent Unit-Rate Graph.

The greatest common divisor of all rate pairs in Eq. (2) will not change if the rate of the leftmost node from Fig. 3 is changed from 8 to 2. The resulting graph can be further transformed to a unit-rate equivalent with a reduced number of nodes. Whereas the graph on Figure 4 has 21 nodes, the graph from Figure 5 has only 6.

7. OPTIMUM TOKEN DISTRIBUTION

Redistributing the tokens of a marked graph according to some criterion is a well known problem in the theory of marked graphs [12]. If the criterion is a linear function, we have to minimize the weighted sum of tokens $c^T M$ under the condition that the solution M_r belongs to the reachability set of M_0, i.e. $M_r \in \mathcal{R}(M_0)$. In the general case we get an instance of an integer linear programming (ILP) problem $\min\{c^T M | B^T r = M - M_0; M, r \in Z \; M \ge 0\}$.

For ordinary graphs the incidence matrix is totally unimodular which guarantees that even without the constraint $r \in Z$ the solution for r will be integral [12], and an algorithm of polynomial complexity is available. It is easy to see that the incidence matrix of a non-ordinary marked graph is not totally unimodular [13].

In order to show that in the non-ordinary case the optimum token distribution problem is an ILP problem, we need the following theorem:

Theorem 7 [13]: Let A be an integral matrix of full row rank. Then the polyhedron $\{x|x \geq 0; Ax = b\}$ is integral for each integral vector b, if and only if A is unimodular.

If we take $A = [I; -B^T]$, $x = [M; r]^T$, and $b = M_0$, the resulting polyhedron determines the reachability space. Matrix A is a full row rank matrix, but obviously not unimodular. As a consequence, the polyhedron has also non-integer vertices and in the general case the problem cannot be reduced to a linear programming (LP) problem.

One theoretically interesting exception is the case of an equalized graph with rates co-prime in pairs. In this case, according to Theorem 5 the reachability condition reduces to $\bar{B}^T\bar{r} = M_r - M_0$, and the minimization problem reduces to $\min\{c^T M|\bar{B}^T\bar{r} = M - M_0\}$. Matrix $\bar{B}$ is the incidence matrix of the topologically equivalent unit-rate graph, and is totally unimodular. In this special case the optimum marking can be found using the LP algorithm. It has to be stressed that in the case of co-prime rates the alternative way of conversion to an equivalent unit-rate graph is of exponential complexity.

8. CONCLUSIONS

The presented results are the first step in exploring the theoretical foundations of the non-ordinary retiming transformation. The new reachability theorems are part of the effort to close the gap between the theory of ordinary and non-ordinary marked graphs. One possible application is the retiming-equivalent unit-rate transformation. We believe that additional useful applications will be found in the future.

Our future work shall concentrate on the modification of existing retiming algorithms in order to cover the non-ordinary case. We have already observed that for acyclic graphs Bellman's equations for the shortest path can be used for non-ordinary retiming, too. We shall continue to work along this result.

REFERENCES

[1] C. E. Leiserson, F. M. Rose, and J. B. Saxe, "Optimizing synchronous circuitry by retiming," in *Proc. CalTech Conf. VLSI*, 1983.

[2] M. Lam, "Software pipelining: An effective scheduling technique for VLIW machines," *Proc. SIGPLAN'88, Atlanta, Georgia*, 1988.

[3] K. Parhi, "Algorithm transformation techniques for concurrent processors," *Proceedings of the IEEE*, vol. 77, pp. 1879–1895, Dec. 1989.

[4] P. D. Hoang and J. M. Rabaey, "Scheduling of DSP programs onto multiprocessors for maximum throughput," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 41, pp. 2225–2235, June 1993.

[5] T. Murata, "Circuit theoretic analysis and synthesis of marked graphs," *IEEE Trans. on Circuits and Systems*, vol. CAS-24, pp. 400–405, July 1977.

[6] E. A. Lee and D. G. Messerschmitt, "Static scheduling of synchronous data flow programs for digital signal processing," *IEEE Trans. on Computers*, vol. C-36, No. 1, pp. 24–35, January 1987.

[7] R. M. Karp and R. E. Miller, "Properties of a model for parallel computations: Determinacy, termination, queueing," *SIAM J.*, vol. 14, pp. 1390–1411, Nov. 1966.

[8] E. Teruel, P. Chrzatowsky-Wachtel, J. Colom, and M. Silva, "On weighted T-systems," *Apllication and Theory of Petri Nets*, pp. 348–367, 1992.

[9] M. Hack, *Decidability Questions for Petri Nets.* PhD thesis, Massachusetts Institute of Technology, Cambridge, 1976.

[10] E. A. Lee, *A Coupled Hardware and Software Architecture for Programmable Digital Signal Processors.* PhD thesis, University of California, Berkeley, 1986.

[11] P. Zepter, T. Grötker, and H. Meyr, "Digital receiver design using VHDL generation from data flow graphs," in *Proc. 32nd Design Automation Conf.*, June 1995.

[12] T. Murata, "Petri nets: Properties, analysis and applications," *Proceedings of the IEEE*, vol. 77, pp. 541–580, Apr. 1989.

[13] A. Schrijver, *Theory of Linear and Integer Programming.* John Wiley & Sons, New York, 1987.

[14] D. Knuth, *The Art of Computer Programming - Vol. II*, ch. 4, pp. 277,585. Addison-Wesley, Reading, 1981.

APPENDIX A

Proof of Theorem 5: For equalized graphs $B^T = \bar{B}^T\mathrm{diag}(S)$, where S is the rate vector. Condition

$$B^T r = M_r - M_0 \text{ has integer solution} \quad (6)$$

is equivalent to

$$\bar{B}^T\bar{r} = M_r - M_0 \text{ has integer solution} \quad (7)$$

and

$$\mathrm{diag}(S)r = \bar{r} + k1^T \text{ has integer solution for some } k \in Z \quad (8)$$

where 1^T is a column vector of ones. From Eq. (1), unimodularity of $\bar{B}$, and from $\mathrm{rank}(\bar{B}) = n - 1$ follows that there is a set of integer solutions of form $\bar{r} + k1^T, \forall k \in Z$. Eq. (8) is equivalent to the requirement that $\{k \in Z | k \equiv \bar{r}_i \pmod{s_i}, 1 \leq i \leq n\}$ is not empty. Vector S is a positive integer vector, and $\bar{r}$ also contains only integers, so from the Generalized Chinese Remainder Theorem [14] follows that an integer k satisfying (8) with r integral can be found iff

$$\bar{r}_i \equiv \bar{r}_j \pmod{\gcd(s_i, s_j)}, \; 1 \leq i < j \leq n. \quad (9)$$

This proves the Theorem.

Proof of Theorem 6: From Theorems 1 and 2 follows that for ordinary graphs existence of a solution for $B^T r = M_r - M_0$, and condition $C_f(M_r - M_0) = 0$ are equivalent conditions. Using

$$\bar{r}_i - \bar{r}_j = m_r(i,j) - m_0(i,j), \; 1 \leq i, j \leq n \quad (10)$$

proves the Theorem.

ON RETIMING OF MULTIRATE DSP ALGORITHMS

Vojin Živojnović and Rainer Schoenen

Integrated Systems for Signal Processing, IS2
Aachen University of Technology
Templergraben 55, 52056-Aachen, Germany

In the paper retiming of DSP algorithms exhibiting multirate behavior is treated. Using the non-ordinary marked graph model and the reachability theory, we provide a new condition for valid retiming of multirate graphs. We show that for a graph with n nodes the reachability condition can be split into the reachability condition for the topologically-equivalent unit-rate graph (all rates set to one), and $(n^2-n)/2$ rate-dependent conditions. Using this property a class of equivalent graphs of reduced complexity is introduced which are equivalent in terms of retiming. Additionally, the circuit-based necessary condition for valid retiming of multirate graphs is extended for the sufficient part.