Mathematica
Balkanica

New Series Vol. 21, 2007, Fasc. 1-2

Criterions for the Valid Transformation Choice in Planar Systolic Arrays Synthesis

Dragan M. Randjelović, Igor $\hat{Z}$. Milovanović, Emina I. Milovanović

Presented by P. Boyvalenkov

The objective of this paper is to introduce some additional constraints for the transformation matrix T in order to reduce the number of valid transformations only to those that will give systolic arrays with optimal space parameters, among them number of processing elements, geometric, chip area and input-output elements. To show the importance of considering the conditions for the valid transformation T determination, at the end of this paper an example for two rectangular matrix multiplication with three different valid transformation for projection direction $\vec{\mu} = [111]^T$ is given.

Key Words: systolic arrays, criterions for the valid transformation, objective functions

1. Introduction

Consider a regular three nested loop alghorithm A. To this algorithm we can uniquely add an arranged pair (D,P_{int}), where D is a matrix of data dependence and P_{int} is the index space of interior calculation. For every allowed projection direction $\vec{\mu} = [\mu_1 \mu_2 \mu_3]^T$ it is defined a valid transformation

$$T = \begin{bmatrix} \vec{\Pi} \\ S \end{bmatrix} = \begin{bmatrix} \vec{\Pi} \\ \vec{S}_1 \\ \vec{S}_2 \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix}. \tag{1}$$

Using this transformation, due to the mapping T : $(D, P_{int}) -> (\triangle, \bar{P}_{int})$ given in notation

$$T(D, P_{int}) = (\triangle, \bar{P}_{int}) \tag{2}$$

we get an appropriate systolic array (SA).

In [1] vector $\vec{\Pi}$ is a time and matrix S space component of the mapping T.

In mathematical studies we find a great number of conditions which enable determination of this mapping for a given projection direction $\vec{\mu} = [\mu_1 \mu_2 \mu_3]^T$(see for example [1],[7] and [8]). However, these conditions do not enable the unique determination of the transformation T. So we get a whole set of transformations that can be named the set of valid transformations; but this is of no use. For different transformations from the same set of the valid transformations, topologically the same SA are synthesized. But they can be very different according to space objective functions (characteristics), as a number of processing elements- Ω_p , the surface of SA - g_a, and surface of the chip on which the SA is realized - area. This raises additional difficulties to the planning engineer who has to choose from the set of valid transformations those that he can get the best (usually the smallest) space characteristics for. Therefore for the conditions used to determine the valid transformation T, it is very important the given projection direction $\vec{\mu} = [\mu_1 \mu_2 \mu_3]^T$ to be intensified and expanded, so that the set of valid transformations to be as small as possible. For this purpose it is desirable to include only those transformations that synthesized the SA with the best space characteristics. This problem connected with planar SA is examined in this paper.

2. Determination of valid transformation T

Each regular three nested loop algorithm can be characterized (presented in space) by a pair (D;P_{int}), where $D = [\vec{e}_1^{\,3} \vec{e}_2^{\,3} \vec{e}_3^{\,3}]$ is a dependency matrix of dates,

$$P_{int} = \{(i,j,k) \mid 1 \le i \le N_1,\ 1 \le j \le N_2,\ 1 \le k \le N_3\} \ , \tag{3}$$

is the set of points where data are calculated and N_1, N_2, N_3 are the boundaries of the loop. $\vec{\Pi} = [t_{11} t_{12} t_{13}]$ in [1], like the time component of T is uniquely determined as:

$$\vec{\Pi} = \pm \left((\vec{e}_2^{\,3})^T - (\vec{e}_1^{\,3})^T \right) \times \left((\vec{e}_3^{\,3})^T - (\vec{e}_1^{\,3})^T \right) \ ,$$

the sign is determined so the one of two alternative conditions is valid:

$$\vec{\Pi}\vec{e}_i^{\,3} > 0 \ \text{ or } \ \vec{\Pi}\vec{e}_i^{\,3} < 0 \ \text{ for all } i = 1,\, 2,\, 3\,.$$

Thus the vector $\vec{\Pi}$ can be uniquely determined without many difficulties.
The procedure for the determining of matrix S, like the space component of T in [1], is more complex:

$$S = \begin{bmatrix} \vec{S}_1 \\ \vec{S}_2 \end{bmatrix} = \begin{bmatrix} t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \ .$$

First, we will mention conditions that can be found in the references (see for example [2]-[5]) due to which these matrices for given projection direction $\vec{\mu} = [\mu_1\mu_2\mu_3]^T$ can be determined.

1.) The valid transformation T, because of the unique mapping must be non singular, that means the following inequality must hold:

$$detT \neq 0 \ . \tag{4}$$

2.) The given projection direction $\vec{\mu} = [\mu_1\mu_2\mu_3]^T$ must be orthogonal to the plane of projection, that means equality must be present

$$\vec{S_1}\vec{\mu} = 0 \ \text{ and } \vec{S_2}\vec{\mu} = 0 \ . \tag{5}$$

This condition comes from the fact that the vector of projection direction $\vec{\mu} = [\mu_1\mu_2\mu_3]^T$ cannot be parallel to this plane or this vector cannot lie in this plane. Namely, this case brings a conflict of irregularity. Then we can have a situation that different values in same tact demand at the same time the access to same processing elements.

3.) It is demanded the connections between the processing elements in the SA to be realized as the closest neighbours. The requirement is equivalent to the condition that the elements of the matrix $\triangle_S$

$$\triangle_S = S \cdot D$$

belong to the set -1,0,1.

4.) To avoid conflicts during the realization of the given algorithm on the synthesized SA it is forbidden to have simultaneous equalities for both different vectors of the position $\vec{P_1}$ and $\vec{P_2}$, points from P_{int}

$$\vec{\Pi} \cdot \vec{P_1} = \vec{\Pi} \cdot \vec{P_2} \ \text{ and } \ S \cdot \vec{P_1} = S \cdot \vec{P_2} \ .$$

If these planar SA are contaversal, to this set of conditions we have to add the following conditions that are necessary so that the projection direction $\vec{\mu} = [\mu_1\mu_2\mu_3]^T$ can be allowed.

5.) The elements

$$\mu_i, i = 1, 2, 3 \text{ must belong to the set -1,0,1 .}$$

6.) The projection directions $\vec{\mu} = [\mu_1 \mu_2 \mu_3]^T$ with the characteristic

$$\vec{\Pi} \cdot \vec{\mu} = 0 \text{ and } \vec{\mu} = [0,0,0]^T$$

are excluded. The difference between the projection directions $\vec{\mu}$ and $-\vec{\mu}$ is also not made bacause we have the same SA for both of them.

As we have already mentioned these conditions do not enable to uniquely determine the valid transformation T for the given projection direction $\vec{\mu} = [\mu_1 \mu_2 \mu_3]^T$ that fulfill the conditions 5.) and 6.). We get the set of these transformations. We will mention now some new conditions which either take place of already mentioned conditions or they are added to them, in order to reduce the mentioned set as much as possible.

Condition 2.) is replaced by the following one

7.) The vectors must satisfy the equality

$$\vec{\mu} = \vec{S}_1 \times \vec{S}_2 \; ; \tag{6}$$

It is not difficult to notice that the vectors $\vec{S}_1$ and $\vec{S}_2$ satisfy both 6.) and 5.), but it needs not to be the same in the opposite way. The condition 7.) enables us, for the cofactors of the matrix T,$|T_{11}|, |T_{12}|$ and $|T_{13}|$ on the basic of 5.), to have the characteristic where they take values from the set 0,1 and that is very important for the minimization of the size g_a of the synthesized SA.

If we want to minimize the surface g_a that synthesize two dimensional SA and the surface of the chip (area) with the elements of the matrix S we have to make them satisfy one of these conditions:

8.) If $\mu_1 = 1$ the elements of the matrix S must satisfy the equality

$$t_{22}t_{32} + t_{23}t_{33} = 0 \; , \tag{7}$$

9.) If $\mu_2 = \pm 1$ the elements of the matrix S must satisfy the equality

$$t_{21}t_{31} + t_{23}t_{33} = 0 \; . \tag{8}$$

It is necessary to mention that for the elements of the matrix S there is no need to require to simultaneously satisfy 7.) and 8.). It could be considered as a dilemma if the projection direction would be $\vec{\mu} = [1 \pm 1\mu_3]^T$. In such situations we should choose either 7.) or 8.) and the criteria for that should be: If $N_1 < N_2$ we should choose 8.) and if $N_1 > N_2$ we should choose 7.). If $N_1 = N_2$ it is totally unimportant if we choose 7.) or 8.).

To show the importance of considering the conditions for determination of the valid transformation T for the given permitted projection direction $\vec{\mu} = [\mu_1 \mu_2 \mu_3]^T$ our suggestion in this paper is to examine a concrete algorithm for the two rectangular matrix multiplication.

Algorithm 1 (Multiplication of the matices C=AxB)

for k:=1 to N_3 do

for j:=1 to N_2 do

for i:=1 to N_1 do

a(i,j,k)=a(i,j-1,k)

b(i,j,k)=b(i-1,j,k)

c(i,j,k)=c(i,j,k-1)+a(i,j,k)b(i,j,k);

where a(i,o,k)$\equiv a_{ik},\ b(0,j,k) \equiv b_{kj}, c(i,j,0) \equiv 0$ for every i,j,k .

From the set of the permitted projection direction without reduction of generality we notice $\vec{\mu} = [111]^T$. As suitable transformations T take the matrices:

$$T_1 = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}, \quad T_2 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}, \quad T_3 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ 2 & -1 & -1 \end{bmatrix}.$$

The transformation T_1 satisfies the conditions 1.),3.),4.),7.),8.). The transformation T_2 (see for example [7]) satisfies the conditions 1.),3.),4.),7.) but it does not satisfy 8.) and 9.).

The transformation T_3 (see for example [6]) satisfies the conditions 1.),3.),4.) but it does not satisfy the condition 7.) and none of the alternate two: 8.) and 9.). We will mark with SA1,SA2 and SA3 the suitable synthesized SA. In Table 1 we give the space characteristics of these SA. The values of the parameters for $N_1 = N_2 = N_3 = N$ are introduced in the brackets. For the formation of Table 1 we have used the procedure for the determination of the space characteristics developed in [2] that takes into consideration the uniqueness of the algorithm 1. If we, for example, had used the procedure from the paper [5], the differences of the space parameters of the SA1 in regard to the SA2 and SA3 would have been even bigger. They would have appeared considering also the number of the processing elements.

SA	SA1	SA2	SA3
Ω_p	N_2N_3	N_2N_3	N_2N_3
-II-	(N^2)	(N^2)	(N^2)
g_a	$(N_2-1)(N_3-1)$	$(N_2-1)(N_3-1)$	$2(N_2-1)(N_3-1)$
-II-	$((N-1)^2)$	$((N-1)^2)$	$(2(N-1)^2)$
area	N_2N_3	$N_3(N_2+N_3-1)$	$(N_2+N_3-1)^2$
-II-	(N^2)	$((N-1)^2)$	$(2(N-1)^2)$

Table 1.

3. Conclusion

In this paper we have considered the criterions that should satisfy valid matrices of the transformation T that map the regular three nested loop algorithm in the SA. The aim was to reduce the number of valid transformations only to those that give SA with optimal space characteristics, among them number of processing elements, geometric, chip area and input-output elements. The results are illustrated by the example of the algorithm for the two rectangular matrix multiplications.

References

1 M. O. Esonu et al. Systolic arrays: How to chose them, *IEE. Proc.*, **139** (1992), 3, 179-188.

2 T. I. Tokić et al. Determining VLSI array size for one class of nested loop algorithms, *Advances in Computer and Information Sciences'98*, IOS Press, 1998, 389-396.

3 D. M. Randjelović. Object function 1D systolic arrays, *Math. Balk., New Series*, **19** (2005), 1-2, 195-202.

4 E. I. Milovanović et al. Designing Hexagonal Systolic Array by Composite Mappings, *Facta Univ.(Niš), Ser. Math. Inf.*, **12** (1997), 283-296.

5 C. N. Zhang et al. Determining objective functions in systolic array designs, *IEEE Trans. VLSI Systems.*, **31** (1994), (2), 357-360.

6 C. R. Wan et al. Nineteen Ways of Systolic Matrix Multiplication, *Intern. J. Comput. Math.*, Vol. **68** (1998), 39-69.

7 H. T. Kung et al. *Algorithms for VLSI Processor Arrays, In Introduction to VLSI Systems, Section 8.3* , Adison-Wesley, 1980.

8 D. M. Randjelović ¨ et al. Objective systolic arrays functions, *Math. Balk., New Series*, Vol. **20**, (2006), 287-298.

Agrycultural faculty University of Priština
Lešak 38219, SRBIJA
E-mail: ganacd@bankerinter.net

Received 02.01.2006